

О.Ю. Реброва

Число случаев

Здоровые

# Статистический анализ медицинских данных

Применение пакета  
прикладных программ  
**STATISTICA**

О.Ю. Реброва

# **Статистический анализ медицинских данных**

## **Применение пакета прикладных программ STATISTICA**

 Издательство Медиа Сфера

Москва

2002

УДК 616:519.2  
ББК 519  
Р 31

**О.Ю. Реброва. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA**

М., МедиаСфера, 2002. 312 с.

В книге в контексте концепции доказательной медицины представлены наиболее актуальные методы статистического анализа медицинских данных, способы корректной интерпретации результатов статистического анализа, современные международные требования к представлению результатов статистического анализа в статьях и диссертациях.

Книга рассчитана на проводящих научные исследования медиков и биологов, аспирантов и студентов, заинтересованных в освоении методов статистики в научных исследованиях.

Научное и учебное издание  
Редактор М.Н. Соловова

STATISTICA® — зарегистрированный товарный знак StatSoft Inc. (USA).  
MS Windows, MS Word, MS Excel — зарегистрированные товарные знаки Microsoft (USA)

© О.Ю. Реброва, 2002

ISBN 5-89084-013-4 © Медиа Сфера, Москва, 2002 — оформление  
Отпечатано в типографии ООО «Информполиграф», Москва

# Оглавление

<b>Предисловие автора .....</b>	<b>9</b>
<b>Глава 1. Доказательная медицина и планирование научных исследований .....</b>	<b>12</b>
1.1. Основные понятия доказательной медицины .....	12
1.2. Источники научно обоснованной медицинской информации ...	16
1.3. Определение цели исследования .....	17
1.4. Планирование исследования .....	18
1.4.1. Типы исследований .....	19
1.4.2. Достоверность и обобщаемость результатов исследования .....	23
1.4.3. Систематические и случайные ошибки .....	25
1.5. Некоторые вопросы планирования клинических испытаний .....	29
1.5.1. Виды структуры клинических испытаний .....	29
1.5.2. Распределение участников по группам в рандомизированных клинических испытаниях .....	30
1.5.3. Маскирование вмешательства .....	34
<b>Глава 2. Типы данных .....</b>	<b>36</b>
<b>Глава 3. Подготовка данных к статистическому анализу .....</b>	<b>40</b>
3.1. Подготовка первичных данных .....	42
3.1.1. Оценка точности данных .....	42
3.1.2. Предварительные расчеты .....	43
3.1.3. Идентификация участников (объектов) исследования .....	44
3.1.4. Пропущенные (отсутствующие) значения данных .....	44
3.1.5. Данные динамических исследований .....	45
3.1.6. Данные для анализа времени до наступления исхода .....	46

3.1.7. Проверка данных .....	47
3.1.8. Перенос данных, подготовленных в других программах, в таблицу данных пакета прикладных программ STATISTICA .....	49
3.2. Подготовка вторичных данных .....	50
<b>Глава 4. Управление данными и сохранение результатов анализа в пакете прикладных программ STATISTICA .....</b>	<b>52</b>
4.1. Отбор строк таблицы данных по условию .....	52
4.2. Формирование подгрупп (разделение файлов) .....	53
4.3. Объединение файлов .....	54
4.4. Операции со столбцами .....	55
4.5. Операции со строками .....	57
4.6. Сохранение результатов статистического анализа .....	58
<b>Глава 5. Основные принципы и методы статистического анализа .....</b>	<b>60</b>
5.1. Описательная статистика и статистическая оценка .....	60
5.2. Проверка статистических гипотез .....	64
5.3. Статистическое моделирование .....	69
5.4. Статистическая и клиническая значимость полученных результатов .....	69
5.5. Первичный и вторичный анализ данных .....	70
5.6. Классификация статистических методов .....	71
<b>Глава 6. Методы статистического анализа, содержащиеся в пакете прикладных программ STATISTICA .....</b>	<b>75</b>
<b>Глава 7. Описание количественных признаков .....</b>	<b>77</b>
7.1. Анализ соответствия вида распределения признака закону нормального распределения .....	77
7.2. Описание количественных данных в зависимости от вида их распределения .....	83
7.3. Некоторые частные аспекты представления количественных данных .....	92
7.3.1. Точность представления описательных статистик количественных данных .....	92
7.3.2. Данные связанных групп .....	92

7.3.3. Преобразование количественных данных .....	93
7.3.4. Описание данных, полученных в малых выборках .....	94
7.3.5. Анализ выпадающих данных .....	94

## **Глава 8. Сравнение групп по количественному признаку ..... 96**

8.1. Сравнение одной группы с популяцией .....	98
8.1.1. Случай нормально распределенного признака .....	98
8.1.2. Случай любого распределения признака .....	100
8.2. Сравнение двух независимых (несвязанных) групп .....	101
8.2.1. Доверительный интервал для разности средних .....	102
8.2.2. Параметрический метод ( $t$ -критерий Стьюдента для независимых групп) .....	104
8.2.3. Непараметрические методы (критерии Манна—Уитни, Вальда—Вольфовица, Колмогорова—Смирнова) .....	109
8.3. Сравнение двух зависимых (связанных) групп .....	111
8.3.1. Доверительный интервал для средней разности .....	112
8.3.2. Параметрический метод ( $t$ -критерий Стьюдента для зависимых групп) .....	114
8.3.3. Непараметрические методы (критерий знаков, критерий Вилкоксона) .....	116
8.4. Сравнение трех независимых (несвязанных) групп и более .....	118
8.4.1. Параметрический дисперсионный анализ .....	119
8.4.1.1. Проверка гипотез о равенстве дисперсий .....	120
8.4.1.2. Собственно дисперсионный анализ .....	122
8.4.1.3. Апостериорные сравнения групп .....	125
8.4.2. Непараметрические методы сравнения независимых групп (метод Краскела—Уоллиса, медианный тест) .....	129
8.5. Сравнение трех связанных (зависимых) групп и более (непараметрический метод Фридмана) .....	132

## **Глава 9. Описание качественных признаков ..... 134**

9.1. Вычисление параметров распределения качественных признаков .....	137
9.2. Вычисление абсолютных и относительных частот (долей, процентов, вероятностей, шансов) .....	138

9.3. Описание относительной частоты бинарного признака с использованием доверительного интервала .....	142
<b>Глава 10. Сравнение групп по качественному признаку .....</b>	<b>145</b>
10.1. Сравнение наблюдаемых и ожидаемых частот (анализ одной группы) .....	146
10.2. Сравнение двух групп и более .....	148
<b>Глава 11. Сравнение групп по качественному бинарному признаку .....</b>	<b>157</b>
11.1. Сравнение выборочной относительной частоты с популяционной .....	158
11.1.1. Сравнение доверительного интервала для выборочной относительной частоты с популяционной .....	158
11.1.2. Проверка гипотез ( $z$ -критерий) .....	159
11.1.3. Проверка гипотез (критерий $\chi^2$ ) .....	161
11.2. Сравнение относительных частот внутри одной группы и в двух группах .....	162
11.3. Сравнение частот бинарного признака в двух несвязанных (независимых) группах (анализ таблиц $2 \times 2$ ) .....	166
11.3.1. Доверительный интервал для разности относительных частот .....	171
11.3.2. Доверительный интервал для относительного риска .....	173
11.3.3. Доверительный интервал для отношения шансов .....	175
11.3.4. Проверка гипотез (точный критерий Фишера, $\chi^2$ с поправкой Йетса) .....	176
11.4. Сравнение частот бинарного признака в двух связанных (зависимых) группах наблюдений (случай парных наблюдений) .....	180
11.4.1. Доверительный интервал для разности относительных частот .....	181
11.4.2. Проверка гипотез (критерий МакНемара) .....	183
11.5. Сравнение трех групп и более по бинарному признаку .....	183
11.5.1. Случай неупорядоченных групп .....	184
11.5.2. Случай упорядоченных групп .....	184

<b>Глава 12. Анализ связи (корреляции, ассоциации)</b>	
<b>двух признаков</b> .....	<b>185</b>
12.1. Параметрический метод (метод Пирсона) .....	188
12.2. Непараметрические методы (методы Спирмена, Кендалла, гамма) .....	191
12.3. Доверительный интервал для коэффициента корреляции .....	194
12.4. Сравнение двух коэффициентов корреляции .....	195
<b>Глава 13. Многофакторный анализ данных</b> .....	<b>198</b>
13.1. Многофакторный параметрический дисперсионный анализ .....	199
13.2. Дискриминантный анализ .....	202
13.3. Кластерный анализ .....	207
13.4. Анализ главных компонент и факторный анализ .....	208
13.5. Многофакторный анализ объектов исследования с пропущенными значениями .....	208
13.6. Проверка работоспособности многофакторных моделей .....	210
<b>Глава 14. Анализ вида зависимости одного признака от одного или нескольких признаков (регрессионный анализ)</b> .....	<b>211</b>
14.1. Линейный регрессионный анализ .....	213
14.2. Логистическая регрессия как один из методов нелинейного регрессионного анализа .....	221
<b>Глава 15. Некоторые общие проблемы и частные задачи статистического анализа</b> .....	<b>227</b>
15.1. Проблема множественных сравнений .....	227
15.2. Определение интервала нормы (референтного интервала) для количественного признака .....	230
15.3. Определение необходимых объемов выборок при планировании исследования. Статистическая мощность (чувствительность) исследования .....	234
15.4. Анализ времени до наступления исхода .....	238
15.4.1. Анализ вероятности наступления изучаемого исхода в определенный период времени (выживания) .....	240
15.4.2. Исследование влияния одного фактора на время до наступления изучаемого исхода .....	243



15.4.3. Анализ влияния нескольких факторов на время до наступления изучаемого исхода .....	246
15.5. Анализ точности диагностического метода .....	247
15.5.1. Операционные характеристики диагностического метода .....	247
15.5.2. Анализ согласованности независимых диагностических заключений .....	251
<b>Глава 16. Объединение результатов нескольких исследований (мета-анализ) .....</b>	<b>254</b>
<b>Глава 17. Современные требования к описанию в публикациях процедуры и результатов статистического анализа медико-биологических данных .....</b>	<b>261</b>
17.1. Рекомендации по разделу “Введение” .....	262
17.2. Рекомендации по разделу “Материалы и методы” .....	262
17.3. Рекомендации по разделу “Результаты” .....	265
17.4. Рекомендации по разделу “Обсуждение” .....	268
17.5. Описание ограничений исследования .....	269
17.6. Рекомендации по разделу “Выводы” .....	269
<b>Литература .....</b>	<b>270</b>
<b>Приложение 1. Аннотированный список рекомендуемой литературы на русском языке .....</b>	<b>272</b>
<b>Приложение 2. Список рекомендуемой литературы на английском языке .....</b>	<b>285</b>
<b>Приложение 3. Ранги, с помощью которых вычисляется доверительный интервал для медианы распределения .....</b>	<b>286</b>
<b>Приложение 4. Нижние и верхние границы 95% доверительных интервалов для относительных частот, представленных в виде простых дробей .....</b>	<b>290</b>
<b>Толковый словарь терминов .....</b>	<b>294</b>

## Предисловие автора

Тот, кто не знает математики, не может узнать никакой другой науки и даже не может обнаружить своего невежества, а поэтому не ищет от него лекарства.

*Р. Госкон*

Наука начинается тогда, когда человек подходит к явлениям природы с числом и мерою.

*В.И. Вернадский*

Предлагаемая книга является руководством по применению статистики в медицинских — клинических и эпидемиологических — исследованиях. Описываемый подход основан на использовании пакета прикладных программ (ППП) **STATISTICA** фирмы StatSoft Inc. (США) для персонального компьютера, однако изложенные принципы применения статистики универсальны и могут применяться при использовании любых пакетов статистических программ.

Для чтения книги знание основ математической статистики является желательным, но не обязательным. Тем же, кто знаком с основами статистики, книга поможет освоить технологию анализа данных с использованием ППП **STATISTICA**. Книга снабжена списками дополнительной литературы (см. Приложения 1 и 2), которой читатель может воспользоваться при желании глубже освоить те или иные методы статистики.

Довольно широко распространено мнение, что необходимость анализа данных в биологических экспериментальных исследованиях отсутствует: “если требуется анализ данных, это значит, что сам эксперимент плох” (в том смысле, что фиксируемый эффект незначителен). Однако применение статистики в медицинских и биологических исследованиях не ограничивается анализом данных. Статистические методы следует использовать также на этапе планирования биологического эксперимента или медицинского исследования. Мы придерживаемся мнения, что и в биологическом эксперименте для анализа данных необходимо приме-

ние статистики, в противном случае выводы нельзя считать научно обоснованными.

Медицинские клинические исследования намного более сложны по сравнению с биологическими экспериментами: во-первых, на человеке эксперимент возможен лишь в ограниченных рамках, поэтому задание необходимых условий исследования существенно ограничено; во-вторых, фиксируемые эффекты обычно невелики (не превышают 20% от уровня в контрольной группе); в-третьих, выборки гораздо менее однородны. Именно поэтому в клинических исследованиях статистический анализ данных значительно более сложен и трудоемок.

В настоящее время становятся все более актуальными корректное применение статистических методов, научный подход к планированию медицинских исследований. Это связано с развитием концепции и практики доказательной медицины<sup>1</sup>, постепенной интеграцией отечественной науки в мировую, развитием грантовой системы поддержки науки и, следовательно, с повышением требований к методологическому качеству исследований.

Основная цель применения статистических методов — сведение к минимуму случайных ошибок в научном исследовании. При этом могут решаться следующие основные задачи:

- планирование необходимых объемов выборок;
- выдвижение научных гипотез (в том числе изучение структуры объекта исследования и признаков, графическое представление данных и результатов их анализа, построение моделей явлений, проверка статистических гипотез);
- проверка научных гипотез путем проверки статистических гипотез.

В данной книге мы постарались осветить решение перечисленных основных задач с применением статистических методов.

В первой главе излагаются основные понятия доказательной медицины и клинической эпидемиологии, использование принципов которых при проведении исследований приобретает в мировой медицинской науке и практике все более важное значение. Показана определяющая роль статистического анализа при планировании исследования и анализе его результатов.

В последующих главах изложены основные подходы к подготовке и статистическому анализу данных, а также преставлению

---

<sup>1</sup> Evidence-based medicine (англ.).

результатов исследования. Подробно описаны методы анализа, которые наиболее часто рекомендуется применять в медицинских исследованиях, а также приводятся примеры для каждого метода. Даны пояснения по содержанию интерфейсных окон ППП **STATISTICA**.

Кратко описаны методы многофакторного статистического анализа, которые являются более сложными и должны, по нашему мнению, использоваться только специалистами в области медицинской статистики. Однако знание возможностей многофакторных методов необходимо каждому исследователю, работающему в медицинской науке.

Необходимо подчеркнуть, что, с одной стороны, целью автора книги не было освещение всего широкого спектра статистических методов, необходимых в медицинских и биологических исследованиях. С другой стороны, и возможности ППП **STATISTICA** освещены лишь частично. Мы описали лишь основные, наиболее применяемые методы анализа данных.

Выбор ППП **STATISTICA** для иллюстрации методов анализа данных обусловлен его широким распространением в России, достаточными возможностями для решения основных задач медицинских исследований, а также выпуском в 1999 г. русскоязычной версии пакета. Изложение рассчитано на читателя, обладающего навыками работы в среде MS Windows. В связи с тем что в России также широко распространена оригинальная англоязычная версия ППП **STATISTICA**, в тексте приведены также англоязычные аналоги большинства терминов и понятий.

Книга является обобщением опыта автора в области статистического анализа данных научных исследований, проводимых в клинической медицине. Автор пользуется случаем, чтобы выразить свое глубочайшее уважение и признательность И.Д. Четчиной, Б.С. Гершману, Б.В. Алешину, Е.Н. Заниной, В.В. Киликовскому за блестящие уроки математики, руководству НИИ неврологии РАМН за создание условий для написания книги, С.Е. Бащинскому, В.П. Леонову, В.В. Власову и О.И. Терещенко за помощь и ценные рекомендации, своим родным и близким за понимание и поддержку.

С благодарностью приму замечания, вопросы и предложения по содержанию книги. Адреса для корреспонденции: [olga@neurology.ru](mailto:olga@neurology.ru) или Москва 123367, Волоколамское ш., д. 80, НИИ неврологии РАМН.

# **Глава 1. Доказательная медицина и планирование научных исследований**

## **1.1. Основные понятия доказательной медицины**

Постиндустриальное информационное общество, возникающее на наших глазах, поставило перед медицинской наукой и практикой массу проблем. Лавинообразно возрастает объем медицинской информации, разрабатываются наукоемкие медицинские технологии. Растут требования пациентов к качеству медицинской помощи. Вместе с тем становится все очевиднее ограниченность материальных ресурсов здравоохранения даже в экономически развитых странах. Это вызывает необходимость поддержки принятия решений по разработке и внедрению новых эффективных медицинских технологий, которые должны приниматься руководителями здравоохранения.

С примерами использования ненадежной, ненаучной медицинской информации и пациенты, и врачи сталкиваются ежедневно. Достаточно упомянуть многочисленные рекламные ролики, в которых демонстрируется якобы высокая эффективность какого-либо нового препарата по сравнению с “обычным”, показанная в ходе научного исследования. Пациентам простительно поддаваться влиянию такой рекламы. Однако для врача недопустимо слепо доверять подобным источникам информации. Необходимо знать, каким образом были получены эти результаты — на каком контингенте больных, в ходе какого типа исследования и т.д.

Развитие идей критической оценки медицинской информации привело к возникновению в конце 80-х годов XX века концепции доказательной медицины (ДМ).

Основными постулатами ДМ являются следующие [1]:

- каждое решение врача должно основываться на научных данных;
- вес каждого факта тем больше, чем строже методика научного исследования, в ходе которого он получен.

ДМ является концепцией как для врачей, исследователей, руководителей учреждений и органов здравоохранения, так и для пациентов. Основная цель концепции ДМ состоит в том, чтобы постепенно превратить врачебную деятельность из искусства в науку.

Вследствие того, что знания об этиологии и патогенезе неполны, а учет влияния генетических, социальных, экологических факторов на исход того или иного заболевания вряд ли может быть достаточным, клинические прогнозы, построенные на основе знаний об этиологии и патогенезе, носят характер гипотез. Поэтому теоретически эффективные воздействия на практике не всегда работают (например, применение лидокаина для профилактики желудочковых аритмий не влияет на риск внезапной смерти при инфаркте миокарда [2]).

Методической основой ДМ является клиническая эпидемиология (КЭ) — наука, разрабатывающая методы клинических исследований, которые дают возможность делать научно обоснованные заключения, сводя к минимуму влияние систематических и случайных ошибок на результаты исследования. Наиболее полное представление об основных положениях этой науки можно получить из монографии Р. Флетчер, С. Флетчер, Э. Вагнер “Клиническая эпидемиология”, изданной в переводе на русский язык в 1998 г. издательством “МедиаСфера”. В данной главе мы кратко остановимся только на основах этой науки, без которых немыслимо современное научное медицинское исследование.

КЭ позволяет осуществлять прогнозирование для каждого конкретного больного на основании результатов изучения (с использованием строгих научных методов) клинического течения болезни в аналогичных случаях.

Перечислим кратко основные положения КЭ (по [3], с дополнениями):

- в большинстве случаев диагноз, прогноз и результаты лечения для конкретного больного однозначно не определены и потому должны быть выражены через вероятности;

- эти вероятности для конкретного больного лучше всего оценивать на основе предыдущего опыта, накопленного в отношении групп аналогичных больных;
- клинические исследования подвержены систематическим ошибкам, ведущим к неверным заключениям, в связи с тем, что исследования проводятся на свободных в своем поведении больных врачами, обладающими разными уровнями квалификации;
- любые исследования, в том числе клинические, подвержены влиянию случайности;
- во избежание неверных выводов врачи должны полагаться на результаты исследований, основанных на строгих научных принципах и выполненных с использованием способов сведения к минимуму систематических и случайных ошибок;
- сведение к минимуму систематических ошибок достигается правильной структурой (организацией) исследования, адекватной его задачам;



Рис. 1.1. Этапы научного исследования.

— сведение к минимуму случайных ошибок достигается корректным статистическим анализом данных.

Остановимся подробнее на структуре исследования и роли в нем статистического анализа. Следует подчеркнуть, что с точки зрения КЭ для получения надежных, научно обоснованных результатов необходимы оба компонента: правильное планирование структуры исследования (обеспечивающей возможность получения ответов на поставленные вопросы) и грамотный статистический анализ.

Ошибки в планировании исследования первичны. Если структура исследования неадекватна задачам исследования и чревата систематическими ошибками, то даже самый совершенный статистический анализ не обеспечит научно обоснованных результатов. Аналогичная ситуация возникает и в том случае, если исследование спланировано правильно, но статистический анализ проведен плохо. Ошибки в статистическом анализе ведут к неверным выводам.

В настоящее время за рубежом считается общепризнанным, что неправильно проведенное (как при планировании, так и при анализе данных) исследование является неэтичным по крайней мере по следующим причинам:

- 1) участники исследования (как люди, так и животные) в ходе такого исследования подвергаются неоправданному риску;
- 2) неэффективно используются финансовые ресурсы, время исследователей и т.п., которые могли бы быть потрачены на решение других проблем;
- 3) после публикации неверных результатов дальнейшие исследования направляются в неправильное русло;
- 4) применение неверных результатов исследования в медицинской практике может нанести вред больным.

Отметим сразу, что участие специалиста по прикладной статистике в исследовании весьма желательно не только на этапе анализа данных, но и практически на всех других этапах (рис. 1.1), которые мы рассмотрим более подробно. Так, в зарубежной практике исследования вообще не начинаются без согласования со статистиком.

В данной главе мы остановимся лишь на первых двух этапах исследования. Остальные этапы будут освещаться в последующих главах.



## 1.2. Источники научно обоснованной медицинской информации

В настоящее время в мировом медицинском сообществе сформировалось представление о том, что научно обоснованной, доказательной является лишь та информация о медицинских вмешательствах, которая получена в результате исследования, проведенного в соответствии с высокими стандартами клинического испытания.

Большинство применяемых на практике препаратов не проходили такого рода клинических испытаний. Учебники, монографии быстро устаревают, а мнения коллег часто бывают субъективными. В связи с этим в концепции ДМ утверждается, что такие знания об эффективности тех или иных препаратов не являются в достаточной мере научно обоснованными.

Таким образом, перед потребителем медицинской информации — практикующим врачом, пациентом, исследователем — встает вопрос об источниках надежной информации. Такой источник должен постоянно поддерживаться на современном уровне знаний, так как регулярно появляются новые данные о результатах исследований, которые изменяют существующие представления об эффективности того или иного лечебного вмешательства или метода диагностики.

К числу основных регулярно обновляемых источников надежной медицинской информации относятся следующие:

- Руководство “Clinical Evidence”, издаваемое British Medical Journal Publishing Group (Великобритания). В нем содержатся сведения об эффективных вмешательствах, о вмешательствах с недоказанной эффективностью и вредных вмешательствах для всех наиболее распространенных нозологических форм. Руководство обновляется один раз в 6 мес, выходит на английском языке в печатной форме (подписка стоит 140—250 долларов США в год), бесплатно доступно в Интернет ([www.clinicalevidence.org](http://www.clinicalevidence.org)). Издательством “МедиаСфера” в 2002 г. выпущен перевод данного руководства на русский язык, предполагаемая периодичность издания — один раз в год.
- Электронная библиотека систематических обзоров (мета-анализов) и реестр клинических испытаний международного сообщества врачей The Cochrane Collaboration — “The Cochrane

Library”. Выпускается на компакт-дисках (на английском языке) с периодичностью один раз в 3 мес. Подписка в России осуществляется Российским отделением Кокрановского Сотрудничества ([www.cochrane.ru](http://www.cochrane.ru)).

- Электронная (на CD-ROM) библиотека структурированных рефератов “Best Evidence”, опубликованных в журнале “American College of Physicians Journal Club”. Издается на английском языке.
- Реферативные записи, относящиеся к результатам рандомизированных клинических испытаний (РКИ), контролируемых клинических испытаний (ККИ) и мета-анализов могут быть также найдены в базе данных “Medline” следующими способами (подробнее см. [4]):
  - путем добавления к ключевым словам для поиска аббревиатур RCT, CCT или термина “meta-analysis”;
  - путем обращения к интерфейсу “Clinical Queries” на сайте National Library of Medicine (<http://www.ncbi.nlm.nih.gov/PubMed>).
  - Журналы “Evidence-Based Medicine”, “American College of Pathologists Journal Club”, “Международный журнал медицинской практики” (единственное издание на русском языке, систематически публикующее русские переводы структурированных рефератов статей из журнала “American College of Pathologists Journal Club”, т.е. информацию, соответствующую концепции ДМ, а также редакционные статьи по проблемам ДМ). Архив “Международного журнала медицинской практики” с 1996 г. доступен на CD-ROM.
- Электронные руководства на CD-ROM — “UpToDate”, “Scientific American Medicine”.
- Клинические рекомендации для врачей США и Канады, доступные бесплатно в Интернет (например, [www.guideline.gov](http://www.guideline.gov) и [www.cma.ca](http://www.cma.ca)).

Подробно о поиске научно обоснованной медицинской информации см. [4].

### 1.3. Определение цели исследования

Правильность проведения исследования определяется его структурой, т.е. таким планированием исследования, которое позволило бы ответить на поставленные вопросы.

Цель исследования рекомендуется формулировать максимально кратко и ясно. Несмотря на то что цель исследования иногда сформулирована достаточно просто, это не обязательно означает, что его структура может быть простой.

При формулировании цели необходимо дать точное определение каждого используемого понятия. По возможности должны использоваться объективные методы измерения и стандартного представления данных: все признаки, которые могут быть измерены, должны быть измерены стандартными методами, клинические признаки должны оцениваться с помощью принятых международных шкал и т.д. Там, где нет способа надежного измерения, зачастую возникают исследования того, что не существует или не обнаруживается другими исследователями.

В клинической эпидемиологии для оценки эффективности медицинских вмешательств рекомендуется стараться использовать истинные клинические исходы, например развитие заболевания, осложнения, инвалидности, смерть, а не косвенные критерии оценки (иногда их называют суррогатными исходами), такие как физиологические параметры, результаты лабораторных тестов и т.п.

## **1.4. Планирование исследования**

До настоящего времени исследования планируются довольно редко, чаще они просто “случаются”. Исключение составляют международные клинические испытания медицинских вмешательств, которые проводятся согласно весьма строгим протоколам.

Этап планирования исследования является оптимальным для начала совместной работы со статистиком.

Планирование исследования в общем виде можно разбить на 2 этапа:

1. Определение типа исследования, обеспечение достоверности и обобщаемости результатов планируемого исследования, применение способов сведения к минимуму систематических и случайных ошибок (эти вопросы рассматриваются в данном разделе).

2. Определение объемов выборок (эта проблема будет рассмотрена в разделе 15.3, после того как будет дано определение необходимых понятий).

### 1.4.1. Типы исследований

Для того чтобы выбрать тип исследования, необходимо представлять себе весь спектр существующих типов структуры (организации) исследования.

Классификация медицинских исследований может проводиться по нескольким принципам.

**По цели исследования:**

- выдвигающие гипотезу (относительно менее высокая научная ценность исследования);
- проверяющие гипотезу (относительно более высокая научная ценность исследования).

**По временным параметрам:**

- одномоментное (поперечное)<sup>1</sup> — однократное обследование участников, или объектов, исследования;
- динамическое (продольное)<sup>2</sup> — многократное обследование участников, или объектов исследования.

**По отсутствию или наличию вмешательства:**

- пассивное (наблюдение за естественным течением заболевания);
- активное (исследование медицинского вмешательства — метода лечения или профилактики).

**По соотношению времени сбора данных и формирования выборок:**

- проспективное (изучаемые группы формируют до сбора данных);
- ретроспективное (изучаемые группы формируют после сбора данных).

**Замечание.** Если группы формируют по состоянию на определенный момент времени в прошлом, а затем анализируют уже собранные данные, то такое исследование также можно считать проспективным, хотя оно чревато существенными систематическими ошибками.

Каждое исследование может быть классифицировано в соответствии с каждым из перечисленных принципов. В результате комбинации различных указанных выше характеристик исследования формируются различные типы структуры (планирования,

<sup>1</sup> Cross-sectional (англ.).

<sup>2</sup> Longitudinal (англ.).

или организации) исследований, обладающие разной степенью доказательности (перечислены в порядке возрастания доказательности):

- 1) описание отдельных случаев<sup>1</sup>;
- 2) описание серии случаев;
- 3) ретроспективное исследование случай—контроль;
- 4) аналитическое одномоментное исследование;
- 5) проспективное когортное (популяционное) исследование;
- 6) рандомизированное контролируемое испытание медицинских вмешательств (методов лечения, профилактики);
- 7) мета-анализ — обобщение результатов нескольких рандомизированных клинических испытаний.

Дадим краткую характеристику различных типов структуры (планирования, или организации) исследования.

**Описание отдельных случаев** — наиболее старый способ медицинского исследования. Он состоит в описании редкого наблюдения, “классического” случая (“классические” случаи, кстати, никогда не бывают частыми) или нового феномена. Научные гипотезы в таком исследовании не выдвигаются и не проверяются. Однако данный способ исследования также важен в медицине, так как описание редких случаев или явлений нельзя недооценивать.

**Описание серии случаев**<sup>2</sup> — исследование, включающее обычно описательную статистику группы больных, отобранных по какому-либо признаку. Описательные исследования используются, например, в эпидемиологии для изучения влияния неконтролируемых факторов на возникновение заболевания.

**Исследование случай—контроль**<sup>3</sup> — ретроспективное исследование, в котором по архивным данным или опросу его участников формируют группы из этих участников (больных) с определенным заболеванием и без него, а затем ретроспективно оценивают частоту воздействия предполагаемого фактора риска или причины заболевания. Такие исследования чаще выдвигают научные гипотезы, а не проверяют их.

Преимуществом исследования данного типа являются его относительная простота, дешевизна и быстрота выполнения. Одно-

---

<sup>1</sup> Под случаем здесь понимается объект, или участник исследования (*Примеч. ред.*).

<sup>2</sup> Case series (англ.).

<sup>3</sup> Case-control study (англ.).

ко исследования случай — контроль чреваты возникновением множества возможных систематических ошибок (смещений). Наиболее существенными из них можно считать систематические ошибки, связанные с отбором участников исследования, и систематическую ошибку, возникающую при измерении (см. также раздел 1.4.3).

**Одномоментное (поперечное) исследование**<sup>1</sup> — описательное исследование, включающее однократно обследуемые группы участников и проводимое с целью оценки распространенности того или иного исхода, течения заболевания, а также эффективности диагностики.

Такие исследования относительно просты и недороги. Основной проблемой является трудность формирования выборки, адекватно отражающей типичную ситуацию в изучаемой популяции больных (репрезентативной выборки).

**Проспективное (когортное, продольное) исследование**<sup>2</sup> — исследование, в котором выделенная когорта участников наблюдается в течение определенного времени. Сначала выделяют когорту (или две когорты, например лиц, подвергшихся фактору риска, и лиц, не подвергшихся ему), а затем проводят наблюдение за ней (ними) и сбор данных. В этом заключается отличие от ретроспективного исследования, в котором когорты выделяют после сбора данных. Такой вид исследований используют для выявления факторов риска, прогностических факторов, причин заболеваний, для определения уровня заболеваемости. Проспективные исследования весьма трудоемки, так как должны проводиться в течение длительного времени, когорты должны быть достаточно велики в связи с тем что выявляемые события (например, возникновение новых случаев заболевания) достаточно редки.

Основные проблемы, возникающие при проведении проспективного исследования, следующие:

- вероятность изучаемых событий зависит от способа формирования выборки (когорты; например, наблюдаемые участники из группы риска имеют большую вероятность заболеть, чем участники из неорганизованной популяции);

<sup>1</sup> Cross-sectional study (англ.).

<sup>2</sup> Cohort study, follow-up (англ.).

- при выбывании участников в ходе исследования необходимо выяснять, не связано ли это с изучаемым исходом или фактором;
- с течением времени могут изменяться сила и характер воздействия изучаемого фактора (например, интенсивность курения как фактора риска развития ишемической болезни сердца);
- необходимо добиваться одинакового объема обследования групп воздействия и контроля, чтобы свести к минимуму возможность более раннего выявления заболеваний (следовательно, лучшего прогноза) в более тщательно обследуемой группе.

**Рандомизированное исследование**<sup>1</sup> — это динамическое исследование какого-либо профилактического, диагностического или лечебного воздействия, в котором группы формируются путем случайного распределения объектов исследования по группам (рандомизации). Наиболее известный вариант рандомизированного исследования — *клиническое испытание*<sup>2</sup>. Клиническое испытание — это проспективное сравнительное исследование эффективности двух вмешательств и более (лечебных, профилактических) или диагностического метода, в котором группы испытуемых формируются с использованием рандомизации с учетом критериев включения и исключения. При этом обычно существует гипотеза, возникшая до проведения исследования относительно эффективности испытываемых методов, которая и проверяется в ходе испытания.

Подробнее об особенностях планирования клинических испытаний см. раздел 1.5.

**Мета-анализ**<sup>3</sup> — количественный анализ объединенных результатов нескольких клинических испытаний одного и того же вмешательства при одном и том же заболевании. Такой подход обеспечивает большую статистическую мощность (чувствительность; подробнее см. раздел 15.3), чем в каждом отдельном исследовании, за счет увеличения размера выборки. Мета-анализ используется для обобщения результатов многих испытаний, зачастую противоречащих друг другу. Подробнее о мета-анализе см. в главе 16.

---

<sup>1</sup> Randomised study (англ.).

<sup>2</sup> Clinical trial (англ.).

<sup>3</sup> Meta-analysis (англ.).

Важнейшие проблемы планирования медицинских исследований вышеперечисленных типов, в том числе основные источники возникновения систематических ошибок, подробно описаны в [3].

В биологических экспериментальных исследованиях может быть использован особый тип структуры исследования, неприменимый в медицинских исследованиях — активный эксперимент с использованием методов теории планирования эксперимента (см., например, [5—8]). Именно исследования такого типа являются классическим научным экспериментом, их результаты являются наиболее доказательными.

### 1.4.2. Достоверность и обобщаемость результатов исследования

Любое исследование в зависимости от того, насколько надежны полученные в нем результаты и насколько они применимы в клинической практике, можно охарактеризовать с двух точек зрения:

- достоверности (внутренней обоснованности)<sup>1</sup>;
- обобщаемости (внешней обоснованности, применимости)<sup>2</sup>.

**Достоверность (внутренняя обоснованность)** исследования определяется тем, в какой степени структура исследования соответствует поставленным задачам, а полученные результаты справедливы в отношении изучавшейся выборки. Достоверность зависит от структуры исследования, правильности сбора и анализа данных, наличия и выраженности систематических и случайных ошибок. Достоверность — необходимое, но не достаточное условие полезности клинического исследования.

Поскольку достоверность исследования во многом определяет его ценность, в клинической эпидемиологии разработаны представления об оптимальных вариантах структуры исследований, проводимых для решения различных задач, которые ставятся исследователями — исследований распространенности заболеваний, частоты возникновения новых случаев, риска, прогноза, факторов риска и причин заболеваний, стоимости медицинской помощи,

<sup>1</sup> Internal validity (англ.).

<sup>2</sup> External validity, generalizability (англ.).



испытания методов диагностики, методов лечения, профилактики и т.д. (рис. 1.2).

Задача исследования	Структура исследования
Исследование метода диагностики	Одномоментное
Исследование распространенности заболевания*	Одномоментное
Исследование частоты возникновения новых случаев заболеваний, исходов и т.д.**	Когортное
Исследование факторов риска	Когортное, случай—контроль
Исследование прогностических факторов	Когортное
Исследование методов лечения и профилактики	Рандомизированное клиническое испытание
Исследование причинно-следственных связей	Когортное, случай—контроль

**Рис. 1.2.** Основные варианты оптимальной структуры (планирования, или организации) исследования в зависимости от его задачи (по [3]).

*Примечание.* \* — prevalence; \*\* — incidence.

Необходимо подчеркнуть, что приведенные в таблице типы исследований оптимальны для того или иного случая, однако на практике возможны и отступления от указанной схемы. Тем не менее следует критически относиться к результатам исследования, структура которого не соответствует его задаче.

**Обобщаемость (внешняя обоснованность)** результатов исследования отражает, в какой мере результаты данного исследования применимы к другим группам больных, например к больным другого пола, другой популяции и т.п.

Обобщаемость отражает обоснованность допущения, что участники исследования (больные), включенные в него, сравнимы с другими, подобными им. В принципе, структура научного исследования подразумевает, что выборка случайна, т.е. участники исследования выбраны из исследуемой популяции случайным образом. Однако на практике это невозможно. В большинстве случаев для повышения обобщаемости результатов своего исследования авторы должны стремиться к тому, чтобы их выборка была хотя бы репрезентативна, т.е. соответствовала по основным характеристикам исследуемой популяции. При этом следует избе-

гать работы с группами, значительно отличающимися от общей популяции.

Для того чтобы читатель мог оценить репрезентативность описываемой в статье выборки, в статье должен быть подробно описан способ формирования этой выборки. Исследователь должен использовать ясные и применимые на практике критерии включения в исследование и исключения его участников (или других объектов), а также ясно описывать их в отчётах, чтобы читатели могли правильно использовать опубликованные результаты. В случае, если обобщаемость результатов исследования неясна, грамотный читатель не станет не только доверять им, но даже читать статью или отчет об этом исследовании.

Повышение обобщаемости результатов исследования достигается прежде всего сведением к минимуму систематических ошибок (см. ниже). Однако результаты даже самого безупречного исследования бессмысленны, если их перенести на группу больных, существенно отличающихся по своим характеристикам от исследованной выборки. Необходимо помнить, что чем строже критерии включения в исследование (т.е. уже спектр изучаемых форм, стадий заболевания и т.д.), тем меньше обобщаемость его результатов.

Для повышения обобщаемости применяются также многоцентровые исследования. При этом в исследование попадают больные из разных географических регионов, т.е. выборка оказывается репрезентативной по отношению к более широкой географической зоне. Результаты такого исследования могут затем более обоснованно применяться в отношении населения этой зоны.

### 1.4.3. Систематические и случайные ошибки

При планировании исследования следует стремиться избегать систематических и случайных ошибок.

*Систематическая ошибка*<sup>1</sup> — систематическое (неслучайное, однонаправленное) отклонение результатов от истинных значений. К основным видам систематических ошибок относятся следующие.

**1. Систематическая ошибка, возникающая при отборе**<sup>2</sup>. Эта систематическая ошибка возникает в тех случаях, когда сравниваемые группы участников исследования различаются не толь-

<sup>1</sup> Bias (англ.).

<sup>2</sup> Selection bias (англ.).

ко по главным изучаемым признакам, но и по другим факторам, влияющим на результат исследования, т.е. участники фактически набираются из разных популяций. Эта ошибка возникает на этапе формирования исследуемых групп.

*Пример 1:* хронологическое смещение, при котором в качестве группы контроля используют ранее набранных больных, методика обследования которых с течением времени претерпела изменения.

*Пример 2:* различия групп по полу или возрасту, т.е. факторам, которые могут влиять на оцениваемые исходы.

Следствием систематических ошибок, возникающих при отборе, является формирование контрольной группы, плохо сопоставимой с основной группой (если контрольная группа формируется из больных с другим заболеванием, то вмешиваются опосредованные факторы; если же контрольная группа формируется из общей популяции, то чаще всего она несопоставима с основной группой даже по возрасту и полу). Для предотвращения этого необходимо подбирать пары больных в контрольную и основную группы по нескольким признакам, потенциально влияющим на изучаемые явления.

*Пример:* систематическая ошибка Берксона [9], который по результатам патологоанатомических вскрытий сделал вывод о существовании отрицательной ассоциации между раком и туберкулезом.

Разновидностью систематических ошибок данного вида служит ошибка, обусловленная спецификой подбора участников в основную группу и снижающая возможность распространять результаты исследования на широкий контингент больных — обобщаемость (внешнюю обоснованность, применимость) результатов исследования.

*Пример:* отбор участников только с определенной стадией заболевания.

Следует отметить, что систематические ошибки, возникающие при отборе, весьма характерны для исследований случай—контроль.

**2. Систематическая ошибка, возникающая при измерении<sup>1</sup>.** Она возникает, когда больные в сравниваемых группах

---

<sup>1</sup> Measurement bias (англ.).

обследованы в разной степени (разные методы диагностики, частота обследований) или используются нестандартизованные схемы получения данных и субъективные оценки.

*Пример:* различия в степени подробности сбора анамнеза в группе больных и группе здоровых.

Эта систематическая ошибка может быть также связана с неодинаковой степенью точности оценок различных характеристик у больных основной и контрольной групп, а также с неточностью, неполнотой и возможными искажениями данных, получаемых из медицинских документов.

*Пример:* при ретроспективном исследовании факторов риска обычно оказывается, что больные люди лучше помнят факты неблагоприятных воздействий, чем здоровые лица.

**3. Систематическая ошибка, обусловленная воздействием вмешивающихся факторов<sup>1</sup>** возникает в тех случаях, когда изучаемые факторы взаимосвязаны, причем один из них искажает эффект другого. Это может произойти из-за систематической ошибки при отборе, под действием случайности или из-за реально существующей связи факторов. Такого рода ошибки должны учитываться при анализе данных.

*Пример:* разная распространенность второго фактора риска (например, курения) в сравниваемых группах, не учтенная при проведении исследования влияния количества потребления овощей в пищу на возникновение какого-либо сердечно-сосудистого заболевания.

Если при формировании выборок возникли систематические ошибки (смещения) указанных или других типов, то формируется так называемая смещенная (нерепрезентативная) выборка — выборка, которая систематическим образом отличается от популяции, представляющей объект исследования, или от популяции, по отношению к которой должны применяться результаты исследования. Следовательно, результаты, полученные на смещенной выборке, представляют существенно меньшую ценность. Другим следствием систематических ошибок, возникающих при формировании групп, является отсутствие их сопоставимости по основным характеристикам, предположительно влияющим на результат исследования, например полу, возрасту.

<sup>1</sup> Confounding bias (англ.).

Одним из способов обеспечения сопоставимости групп является использование рандомизации при распределении участников по группам (как в одномоментном, так и в динамическом проспективном исследованиях).

Для оценки сопоставимости групп чаще всего проводится сравнение их основных исходных характеристик с помощью статистических критериев, или тестов. Отсутствие исходных различий групп позволяет предполагать, что преимущество в эффективности испытываемого нового вмешательства или диагностического метода объясняется именно им, а не какими-либо различиями (несопоставимостью) исследуемых групп.

Если же при формировании групп правильно проводилась рандомизация, то все их различия заведомо носят случайный характер, и поэтому статистический анализ может лишь подтвердить или опровергнуть эффективность рандомизации. Статистический анализ устанавливает лишь наличие или отсутствие сопоставимости сравниваемых групп. В случае несопоставимости групп, несмотря на проведенную рандомизацию, следует использовать специальные подходы к анализу данных (в этой ситуации рекомендуется обращаться к статистике).

Случайные ошибки возникают на любом этапе исследования. В отличие от систематических ошибок случайные ошибки нельзя устранить, но можно свести к минимуму. Для этого необходимы правильное планирование исследования (определение необходимых объемов выборок), а также оценка вероятности возникновения случайной ошибки с использованием статистических методов.

Однако следует помнить, что никакая статистическая обработка данных не может устранить неизвестную систематическую ошибку. Существует английская поговорка: “garbage in, garbage out”, т.е. “из мусора конфетку не сделаешь”. Существует и другая опасность: статистический анализ может придавать исследованию наукообразность, т.е. “обманчивую внешность”, вводящую в заблуждение читателя. Поэтому уважающему себя статистику лучше не браться за обработку заведомо некачественных исходных данных.

Сведение к минимуму случайных ошибок является одной из основных целей проведения статистического анализа данных, полученных в медико-биологических исследованиях.

## 1.5. Некоторые вопросы планирования клинических испытаний

В настоящее время методология проведения клинических испытаний является наиболее разработанной областью клинической эпидемиологии. Утверждены и действуют Единые стандарты представления результатов клинических испытаний [10—12], в которых регламентировано применение статистических методов.

В данном разделе мы рассмотрим следующие особенности планирования РКИ, к которым имеет отношение статистика:

- виды структуры клинических испытаний;
- способы распределения участников испытания по группам;
- способы маскирования вмешательств (ослепления)<sup>1</sup>;
- этические последствия ошибок в клинических испытаниях.

### 1.5.1. Виды структуры клинических испытаний

Выделяют следующие виды структуры клинических испытаний:

- **параллельная**<sup>2</sup> (группы контроля и активного лечения исследуются параллельно, независимо друг от друга; это наиболее распространенная структура исследования);
- **перекрестная**<sup>3</sup> (осуществляется последовательная смена методов лечения в одной группе больных);
- **парная**<sup>4</sup> (проводится анализ групп, сформированных путем подбора пар, т.е. каждому участнику основной группы соответствует участник в контрольной группе);
- **последовательная**<sup>5</sup> (исследование проводится до момента выявления различий между группами);
- **факторный протокол**<sup>6</sup> (исследование групп, в которых применяются комбинации вмешательств; например, при факторном протоколе 22 для двух видов лечения формируются

<sup>1</sup> Вместо терминов “слепой метод” или “ослепление” лучше использовать термин “маскирование вмешательства”, это позволяет избежать ассоциаций с истинной слепотой, хотя все эти термины давно используются в медицинской литературе и должны быть понятны читателям отчетов об РКИ.

<sup>2</sup> Parallel design (англ.).

<sup>3</sup> Cross-over design (англ.).

<sup>4</sup> Paired design (англ.).

<sup>5</sup> Sequential design (англ.).

<sup>6</sup> Factorial design (англ.).

четыре группы, в двух из которых применяется один из видов лечения, в третьей — ни один из них, в четвертой — оба);

- **адаптивная**<sup>1</sup> (по ходу исследования набор участников в группу, получающую худшее, по предварительным оценкам, лечение, уменьшается);
- **структура Зелена**<sup>2</sup> (участникам, распределенным в группу изучаемого лечения, предоставляется возможность отказаться от него и перейти в группу контроля).

Все перечисленные способы планирования исследования (кроме первого) достаточно сложны. Для планирования испытаний с такими типами структуры, а также для анализа получаемых при этом данных настоятельно рекомендуется консультация статистика.

### **1.5.2. Распределение участников по группам в рандомизированных клинических испытаниях**

Основными способами распределения участников по группам в ходе РКИ являются следующие:

- рандомизация;
- минимизация;
- псевдорандомизация.

**Рандомизация** — это случайное распределение участников РКИ, соответствующих критериям включения, по группам. Целью рандомизации является достижение сопоставимости групп по характеристикам, способным влиять на изучаемый результат РКИ. Таким образом сводятся к минимуму систематические ошибки в результатах РКИ, связанные с различиями групп как по известным, так и по неизвестным факторам.

Говорят, что рандомизация эффективна, если полученные группы сопоставимы.

Выделяют несколько типов рандомизации:

- простая;
- блоковая;
- стратифицированная (другое название — стратификационная);
- кластерная.

---

<sup>1</sup> Adaptive design (англ.).

<sup>2</sup> Zelen's design (англ.).

*Простая рандомизация* может проводиться несколькими способами:

1) с помощью подбрасывания монеты — это наиболее простой способ случайного распределения, если необходимо распределять участников РКИ по двум группам. В случае необходимости создания больших выборок этот способ недостаточно надежен, так как может проявиться асимметричность, т.е. деформация, монеты;

2) с применением открытой таблицы случайных чисел (из книг по статистике), если необходимо распределение на две группы и более. Однако и этот способ рандомизации недостаточно надежен;

3) с использованием компьютерной программы генератора случайных чисел (получение последовательности случайных чисел с помощью ППП **STATISTICA** описано в разделе 4.4). Этот способ является оптимальным.

Полученная последовательность случайных чисел может использоваться разными способами:

- четные числа или цифры могут соответствовать одной группе, а нечетные — другой (в случае двух групп);
- если интервал возможных случайных чисел от 0 до 99, то числа, меньшие 50, могут соответствовать одной группе, а большие или равные 50 — другой (в случае двух групп);
- в случае трех групп может быть принято такое правило: числа от 1 до 33 соответствуют первой группе, от 34 до 66 — второй группе, от 67 до 99 — третьей группе (аналогично для четырех групп и более);
- если необходимое соотношение объемов групп заранее известно, но не равно 1, то случайное распределение также легко сделать, разбив интервал значений последовательности случайных чисел на 2 (или более) интервала с необходимым соотношением их размеров.

В результате простой рандомизации группы могут оказаться значительно различающимися по числу участников, причем различие оказывается весьма существенным, если выборки невелики по объему. В связи с этим простую рандомизацию рекомендуется использовать лишь в больших РКИ.

Для поддержания равенства числа участников в группах разработан способ *рандомизации внутри блоков*. Это метод рандомизации, при котором больных, подлежащих включению в ис-



следование, условно разделяют на несколько равных групп (блоков). В пределах каждого блока методы лечения распределяются между больными с использованием рандомизации таким образом, чтобы в итоге каждым методом лечилось заранее определенное соотношение больных. Проведение такой рандомизации достаточно сложно, поэтому с этой целью мы рекомендуем обращаться к статистике.

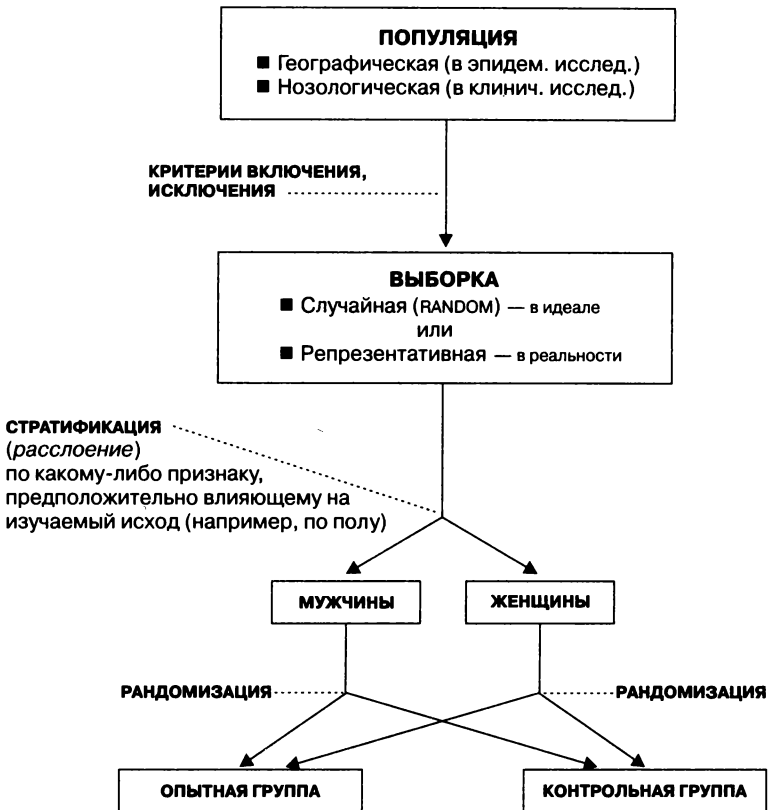
Несмотря на случайный характер распределения участников РКИ по группам, простая и даже блоковая рандомизация не гарантирует получение сопоставимых групп, особенно если размеры групп малы. Хотя обнаруживаемые при этом различия групп случайны, они (несмотря на свой случайный характер) также могут влиять на результаты исследования. Для получения сопоставимых групп целесообразнее всего использовать стратифицированную рандомизацию.

*Стратифицированная рандомизация* — рандомизация, которой предшествует стратификация. Стратификация (расслоение) — это выделение подвыборок (подгрупп) по какому-либо признаку (например, полу, возрасту), который предположительно может влиять на результаты исследования. Затем в каждой из этих подгрупп рандомизация проводится независимо.

Стратификация необходима в исследованиях с ограниченным количеством участников для достижения сопоставимости групп по основным характеристикам.

Стратифицированную рандомизацию целесообразно проводить в подгруппах, подобранных по одному или двум признакам, предположительно влияющим на исход. В случае большего числа прогностических признаков и соответственно большого количества подгрупп достижение сопоставимости становится маловероятным, особенно если общее число участников исследования невелико. Конечно, стратифицированная рандомизация возможна только по признакам, которые заранее известны и подлежат измерению. Пример схемы формирования групп в РКИ приведена на рис. 1.3.

*Кластерная рандомизация* — это вариант стратифицированной рандомизации, при котором рандомизации подвергаются не отдельные участники РКИ, а медицинские центры (поликлиники, больницы) или регионы, где она проводится (т.е. группы участников). Фактически в многоцентровых исследованиях стратифицирующим признаком является сам медицинский центр.



**Рис. 1.3.** Пример схемы формирования групп с использованием стратифицированной рандомизации.

**Минимизация** является единственным альтернативным способом формирования групп, обеспечивающим их сопоставимость по нескольким прогностическим признакам (факторам), предположительно влияющим на исход. Минимизация основана на других, нежели рандомизация, принципах:

- очередного поступающего участника (кроме первого) относят в ту или иную группу не случайным образом, а в зависимости от ранее набранных в группы участников РКИ;
- очередного участника относят в ту группу, в которой существует на данном этапе набора участников исследования

дисбаланс будет сведен к минимуму в результате этой процедуры.

Техника минимизации достаточно сложна, поэтому для ее проведения мы рекомендуем обращаться к статистику.

**Псевдорандомизация** — это неслучайное распределение по группам, которое достаточно широко распространено и часто ошибочно принимается за истинную рандомизацию.

Некоторые широко распространенные способы псевдорандомизации:

- по дате рождения участника;
- по дате вхождения в РКИ;
- по номеру истории болезни;
- поочередно.

На первый взгляд эти способы почти не отличаются от способов простой рандомизации. Однако основное отличие состоит в том, что они являются открытыми, т.е. исследователь, проводящий псевдорандомизацию, может предсказать, в которую из групп будет распределен очередной участник испытания. Вследствие этого появляется возможность влиять на отнесение участников в ту или иную группу, и выборка получается смещенной.

Если несмотря на правильно проведенную рандомизацию группы оказались несопоставимыми по некоторым признакам, т.е. неоднородными, то можно использовать специальные методы анализа данных для учета их влияния на исход (например, ковариационный анализ, множественный регрессионный анализ — см. главу 14).

### 1.5.3. Маскирование вмешательства

Маскирование вмешательства (ослепление) — это способ сведения к минимуму искажений в ходе проведения испытаний. Искажения могут возникать в связи с субъективностью (тенденциозностью) оценки эффективности лечения больным, врачом, статистиком.

Выделяют следующие виды исследований по степени маскирования:

- **простое слепое исследование;** в данном случае больной не знает, какой из методов лечения к нему применяется. Такой вид маскирования может использоваться и при испытании хирургических методов лечения;

- **двойное слепое исследование;** о методе лечения не знают ни больной, ни врач, оценивающий результаты лечения;
- **тройное слепое исследование;** о методе лечения, применяемом в той или иной группе, не знает также и статистик, анализирующий данные.

Для маскирования, например, терапевтического вмешательства необходимо использование плацебо, по внешним свойствам (виду, запаху, вкусу) неотличимого от лекарственной формы изучаемого препарата.

## Глава 2. Типы данных

Первым шагом, предваряющим статистический анализ данных, является анализ (на основе здравого смысла) типов данных<sup>1</sup>. Это необходимо делать для того, чтобы определить правомочность использования в дальнейшем того или иного способа представления данных и статистического метода.

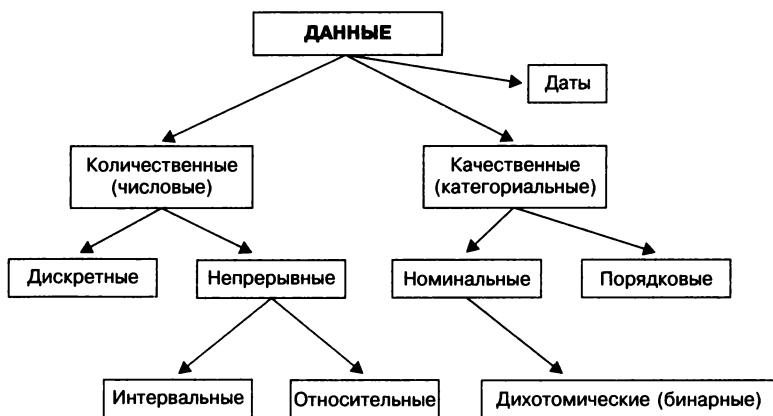


Рис. 2. Типы данных.

Тип получаемых данных (рис. 2.) необходимо также учитывать на этапе планирования исследования при определении необходимых объемов выборок (подробнее см. раздел 15.3).

<sup>1</sup> Здесь следует напомнить, в чем заключается разница между понятиями “признак” и “данные”. Признаки (их также называют переменными) — это названия какой-либо характеристики объекта исследования (например, рост). Данные — набор конкретных чисел для признака, полученный в исследовании. Один и тот же признак может быть описан разными данными и даже иногда данными разных типов.

Принято выделять в качестве основных типов данных количественные и качественные.

**Количественные данные** в свою очередь подразделяются на непрерывные и дискретные.

**Непрерывные данные** — это данные, которые получают при измерении на непрерывной шкале, т.е. теоретически они могут иметь дробную часть (но это не всегда возможно в связи с низкой точностью измерений). Примерами могут служить масса тела, рост, артериальное давление. Непрерывные данные бывают интервальными и относительными.

**Интервальные данные** — вид непрерывных данных, которые измеряются в абсолютных величинах, имеющих физический смысл.

**Относительные данные** — вид непрерывных данных, отражающих долю изменения (увеличения или уменьшения) значения признака по отношению к исходному (или какому-либо другому) значению этого признака. Являются безразмерными величинами или выражаются в процентах.

**Дискретные данные** — количественные данные, которые не могут иметь дробную часть. Пример: количество детей.

**Качественные данные** подразделяются на номинальные и порядковые.

**Номинальные данные** (иногда их также называют *номинативными*) — вид качественных данных, которые отражают условные коды неизмеряемых категорий (например, коды диагноза).

**Порядковые данные** — вид качественных данных, которые отражают условную степень выраженности какого-либо признака (например, стадии онкологического заболевания, степени сердечной недостаточности). Их основное отличие от дискретных количественных данных заключается в отсутствии пропорциональной шкалы для измерения выраженности признака.

**Бинарные (дихотомические) данные** — особо выделяемый вид качественных данных. Признак такого типа имеет лишь два возможных значения (например, пол, наличие или отсутствие какого-либо заболевания).

Как для номинальных, так и для порядковых дискретных данных не могут быть определены “расстояния” между значениями, так как для них не существуют интервальные измерительные шкалы.

Обычно в научном исследовании встречаются данные разных типов. Для определения типа данных компьютер обычно не нужен. Достаточно просто задуматься над природой цифр. Это необходимо делать для того, чтобы выбирать правильные способы описания параметров выборок и методы статистического анализа (подробнее этот вопрос будет рассмотрен в соответствующих главах).

Первый вопрос, на который должен ответить исследователь по поводу каждого из изучаемых признаков — количественный он или качественный. Обычно это бывает сделать достаточно легко. Иногда затруднение вызывает дифференцирование количественных дискретных и качественных порядковых данных. В этой ситуации принципиальное значение имеет природа шкалы измерения: если “расстояния” между значениями на шкале равны, то такие данные являются количественными дискретными (например, количество беременностей). Если же “расстояния” между значениями на шкале не равны, то такие данные являются качественными порядковыми (например, число баллов для оценки интенсивности головной боли).

Второй вопрос — как определить, является ли количественный признак дискретным или непрерывным. Здесь граница достаточно условна. На практике реализуются разные подходы. Один из них состоит в том, что при значительном (превышающем 20) числе различных значений дискретного признака его можно приблизительно считать непрерывным и использовать соответствующие способы описания распределения и методы анализа, адекватные виду его распределения на выборке (например, скорость оседания эритроцитов определяется с точностью до миллиметров, т.е. в дискретной шкале, но поскольку число возможных значений больше 20, то приблизительно можно считать данный признак количественным непрерывным).

В отношении качественного признака вопрос состоит в том, является ли этот признак порядковым или номинальным. Порядковые признаки с большим количеством различных значений (больше 20) также на практике принято считать приблизительно непрерывными (например, сумма баллов какой-либо шкалы для оценки неврологического дефицита, качества жизни и пр.). Для номинальных признаков такой подход исключен.

Особым типом данных являются *даты*. Поскольку в ряде случаев бывает необходимо произвести с ними некоторые арифме-

тические действия (например, вычисление абсолютного периода времени между двумя событиями по датам этих событий), следует обратить внимание на то, в каком формате в используемом Вами статистическом пакете необходимо представлять даты.

Иногда выделяют также некоторые особые подтипы данных, являющиеся частными случаями вышеперечисленных типов:

- ранги;
- очки;
- баллы;
- визуальные аналоговые шкалы;
- цензурированные данные.

Ранги и очки преимущественно используются для полуколичественной оценки параклинических параметров, а баллы — для суммарных оценок по клиническим шкалам. О цензурированных данных см. более подробно в разделах 3.1.6 и 15.4.



## Глава 3. Подготовка данных к статистическому анализу

Подготовка данных может осуществляться в любой программе, поддерживающей электронные таблицы (*STATISTICA*, MS Excel и др.), а также в любых реляционных базах данных, поддерживающих стандартные форматы экспорта данных.

Данные следует располагать в строках и столбцах электронной таблицы. В строках располагаются наблюдения (объекты исследования), в столбцах — переменные (признаки). Качественные данные могут быть представлены текстовыми значениями, которые автоматически кодируются числовыми значениями, однако такое представление не рекомендуется из-за возрастания вероятности ошибок.

Ввод данных в ППП *STATISTICA* может производиться в любом из его модулей. Файлы данных имеют формат \*.sta. Обычные приемы работы с электронными таблицами, принятые в среде MS Windows (в том числе в пакете MS Excel), применимы и в ППП *STATISTICA*. Однако для создания нового файла мы рекомендуем вызвать модуль “Управление данными” (“Data Management”).

### **STATISTICA:**

- “Управление данными” (“Data Management”)
- Меню “Файл” (“File”) или “Анализ” (“Analysis”)
- “Создание файла данных” (“Create new data file”)

В окне (рис. 3.1) необходимо указать число переменных<sup>1</sup> (признаков), т.е. столбцов, и наблюдений (объектов исследования), т.е. строк<sup>2</sup>. Не следует думать, что указанные количества столб-

---

<sup>1</sup> Variables (англ.).

<sup>2</sup> Cases (англ.).

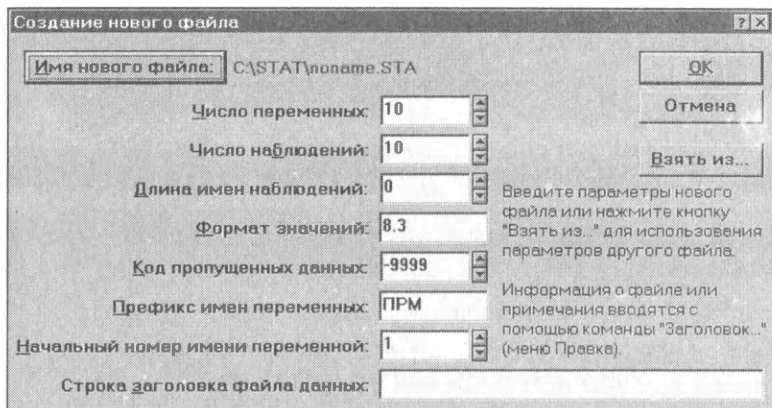


Рис. 3.1. Окно указания спецификаций нового файла.

цов и строк являются окончательными. В дальнейшем их число может быть либо уменьшено, либо увеличено.

Другой способ — обратиться к команде “Файл — Создать данные” и начать ввод данных в файл с исходным названием “new.sta”, увеличивая по мере необходимости число строк и столбцов путем использования опций “Переменные” (“Vars”) — “Добавить” (“Add”) и “Наблюдения” (“Cases”) — “Добавить” (“Add”).

Текстовые имена наблюдений (объектов исследования) — идентификаторы, например фамилии, номера историй болезни, порядковые номера объектов исследования — можно вводить как в любой из столбцов, так и в заголовки строк<sup>1</sup>.

Имя столбца можно задать, активизировав окно его спецификации одним из трех способов:

- щелкнув 2 раза на заголовке столбца (по умолчанию он имеет шаблон “VAR...”);
- вызвав опцию “Текущие спецификации” (“Current Specs”) в меню “Переменные” (“Vars”);
- щелкнув правой кнопкой и выбрав ту же опцию.

<sup>1</sup> Не забывайте о конфиденциальности медицинских сведений. Обычно для статистического анализа нет необходимости вводить в таблицу данных такие идентификаторы, как имя, адрес. Предпочтительным является кодирование участников исследования номерами, по которым исходные данные можно при необходимости найти, например, в архиве.

Имя столбца может включать до 8 букв и цифр, а также знак пробела. Рекомендуется задавать имена с использованием латинских букв, если Вы пользуетесь англоязычной версией ППП **STATISTICA**.

Подробное описание операций управления данными можно найти в справочных файлах ППП **STATISTICA** и “Кратком руководстве”, поставляемом в комплекте с программным обеспечением, а также литературе справочно-учебного характера по данному пакету, например в книгах В.П. Боровикова и соавт. [13—16].

На некоторых наиболее часто используемых процедурах модуля “Управление данными” и функциях, доступных из любого модуля пакета, мы кратко остановимся в главе 4.

### **3.1. Подготовка первичных данных**

Данные для статистического анализа рекомендуется готовить в виде таблицы, в строках которой перечисляются отдельные объекты исследования (например, больные, или участники исследования), а в столбцах — наблюдаемые признаки.

На этом этапе подготовки в одну таблицу мы рекомендуем внести все данные, которые предполагается анализировать, т.е. относящиеся ко всем объектам исследования в разных группах и подгруппах (в том числе в группах контроля), а также все исследуемые признаки для каждого объекта исследования (несмотря на обычно большое количество пропусков данных). Рекомендуется также стремиться вносить в левую часть таблицы все качественные данные (обычно это клинические характеристики больных), а в правую часть — количественные данные (обычно результаты лабораторных и инструментальных обследований). Это облегчает в процессе статистического анализа выделение списков признаков во многих процедурах.

В случае пропущенного (отсутствующего) значения необходимо оставить клетку электронной таблицы пустой.

#### **3.1.1. Оценка точности данных**

Точность количественных данных, которые вносятся в таблицу, определяется точностью измерений того метода или прибора, с которого сняты показания. Например, если активность фермента определяется с точностью до десятых долей единицы измерения, то они так и должны вноситься в таблицу. Если артери-

альное давление (АД) измеряется с точностью до 5 единиц измерения (мм рт. ст.), то также необходимо вносить данные без предварительного округления. Таким образом предотвращаются потери информации, которая может оказаться существенной для результатов статистической обработки данных.

Заметим, что результаты статистического анализа (средние, медианы, средние квадратические отклонения, границы процентильных интервалов, коэффициенты регрессионного уравнения и т.д.) рекомендуется описывать с той же точностью, что и исходные данные. Допускается лишь один дополнительный десятичный разряд. Избыточная точность не имеет смысла. О необходимой точности при расчете параметров см. раздел 7.3.1.

### 3.1.2. Предварительные расчеты

Часто исследователи проводят предварительные расчеты следующих типов:

1. Предварительная разбивка области значений количественного признака на отрезки, т.е. создание на основе непрерывных количественных признаков дискретных количественных признаков. Например, область значений АД делят на интервалы;
2. Вычисление различных расчетных индексов (коэффициентов, отношений и т.п.).

Проводить такого рода предварительные расчеты мы не рекомендуем по следующим причинам.

В первом случае происходит потеря информации. Возможности любого статистического пакета позволяют легко осуществлять любую разбивку областей значений количественных признаков на интервалы автоматически. Кроме того, в дальнейшем может потребоваться разбивка на другие интервалы, а в случае если ранее были введены значения интервалов, а не первичные данные, это сделать будет невозможно. Кроме того, при разбивке на интервалы необходимо (в соответствии с требованиями доказательной медицины) привести обоснование выбора именно таких границ интервалов. В оптимальном случае при этом руководствуются общепринятыми международными классификациями или шкалами. Если же Вы делаете это на основании каких-либо своих представлений или опыта, такую разбивку необходимо убедительно обосновать.

Второй тип предварительных расчетов также нецелесообразен по двум причинам. Во-первых, возможности статистического пакета позволяют сделать такие операции в программном режиме

(автоматически). Во-вторых, при автоматическом расчете точность его выше, а число ошибок меньше, чем при ручном расчете.

### **3.1.3. Идентификация участников (объектов) исследования**

При внесении данных необходимо обозначить, к какой из исследуемых групп относится каждый участник (объект) исследования. В этом случае каждой исследуемой группе присваивается свой код (обычно натуральное, т.е. целое положительное число) — 1, 2, 3 и т.д. Для этого в таблице создают специальный столбец. Для каждого объекта исследования (больного, или участника исследования) в этот столбец заносят соответствующий код группы.

Группирующим называется качественный признак порядкового или номинального типа, значения которого определяются (детерминируются) исследуемой группой. Значения группирующего признака — это коды групп. Иногда такой признак называют также индикатором группы.

Достаточно часто больные из исследуемых групп различаются также по другим дополнительным признакам, т.е. существуют подгруппы, которые в дальнейшем планируется сопоставлять по каким-либо еще параметрам. Подгруппа (или несколько подгрупп) аналогично должны быть закодированы натуральными числами в дополнительном столбце (или столбцах).

Таким образом, один файл данных может содержать несколько группирующих признаков. Вообще говоря, любой из качественных признаков может рассматриваться как группирующий. Любой из количественных признаков может также быть превращен в группирующий после разбивки области значений этого признака на интервалы, каждому из которых может быть приписано условное значение (код). Такая разбивка на интервалы может быть проведена в любом модуле ППП **STATISTICA** путем обращения к опции “Перекодировать” (“Recode”) в меню “Переменные” (“Variables”) (см. раздел 4.4).

### **3.1.4. Пропущенные (отсутствующие) значения данных**

Если информация у какого-либо больного по какому-либо признаку отсутствует, следует оставить соответствующую клетку

пустой. Если значение какого-либо признака равно нулю, то в клетке должна стоять цифра "0". Например, в таблице данных, представленной на рис. 3.2, диагноз закодирован цифрами 1 и 2 (столбец 3), тип течения заболевания определяется только для диагноза 1 (столбец 4) и закодирован цифрами 1, 2, 3, значения фибриногена измерены только у 1, 2, 4 и 5-го больных, причем у 5-го его значение равно нулю.

№	Фамилия больного	Группа (диагноз)	Подгруппа (тип течения)	Уровень В-фибриногена
1	Федоров	1	1	1
2	Климентьев	2		1
3	Некрасов	1	3	
4	Суворов	1	2	3
5	Кузнецов	2		0

Рис. 3.2. Пример структуры таблицы данных в пакете прикладных программ STATISTICA.

### 3.1.5. Данные динамических исследований

Если исследование является динамическим, т.е., например, каждый больной обследовался в период наблюдения дважды, трижды и т.д., то следует использовать один из двух способов размещения данных в электронной таблице.

**Первый способ.** Данные о каждом больном занимают в таблице число строк, равное числу его обследований. При этом вводится дополнительный столбец, в котором могут фиксироваться (в зависимости от необходимости) следующие данные:

- число дней от начала заболевания;
- дата;
- фаза исследования (до или после лечения).

**Второй способ.** Данные о каждом больном располагают в одной строке таблицы, при этом количество столбцов с изменяющимися в динамике данными увеличивается во столько раз, сколько раз обследован больной. При таком способе размещения данных каждый столбец должен иметь уникальное название, следовательно,

столбцы с изменяющимися данными должны иметь какие-либо индексы, например “Холест1”, “Холест2” и “Холест3”.

Таким образом, при использовании любого из перечисленных способов размещения данных вся информация, которой располагает исследователь, должна быть внесена в одну таблицу.

Выбор одного из описанных способов размещения динамических данных может определяться следующим соображением. Если количество измеряемых в динамике признаков составляет более 50% от общего числа признаков, то эффективнее использовать первый способ, если наоборот, то второй способ.

Существенную роль при выборе того или иного способа размещения данных играют те статистические методы, которыми исследователь будет в дальнейшем пользоваться. Однако для суждения об этом исследователь должен иметь определенный опыт выполнения статистического анализа данных в том или ином ППП, поэтому на этих аргументах мы здесь не останавливаемся. Отметим лишь, что при всестороннем анализе данных динамических исследований приходится пользоваться обоими перечисленными способами для применения разных методов анализа, т.е. реформатировать данные в ходе исследования.

### **3.1.6. Данные для анализа времени до наступления исхода**

В медицинских исследованиях обычно оценивается какой-либо наблюдаемый исход — возникновение осложнения, рецидива, выздоровление, смерть и др. Таким образом, понятие “анализ времени до наступления исхода” имеет более широкий смысл, чем понятие “анализ дожития” — термин, ранее более распространенный.

Следует заметить, что момент включения больного в исследование должен быть особо оговорен и обоснован, т.е. больные должны включаться в исследование строго на одном и том же этапе их заболевания. Например, в момент постановки диагноза или в момент проведения операции. В этом случае выборка называется синхронизированной когортой. Только такая когорта может исследоваться с целью анализа времени до наступления исхода.

Данные для такого рода анализа готовятся особым образом.

В двухмерной таблице данных в строки (как и в других случаях) вносятся отдельные объекты исследования (клинические случаи, больные), а в столбцы — их признаки. Однако поскольку

не все объекты исследования бывают прослежены до момента наступления исхода (поскольку больные выбывают по ходу исследования), то в отдельный столбец (индикатор цензурирования; это признак, содержащий сведения о свойстве наблюдения — цензурированное оно или нет) вносится следующая информация о каждом из объектов исследования:

- если исход известен — это так называемое завершённое наблюдение. В данном случае вносится значение “0”;
- если исход неизвестен (больной выбыл из исследования до наступления исхода или на момент окончания исследования у больного исход не наступил) — это незавершённое (цензурированное) наблюдение, вносится значение “1”.

В принципе коды могут быть и другими. Мы привели лишь обычно используемые значения.

В ещё один специальный столбец вносится время от момента включения больного в исследование до момента наступления исхода или выбытия из исследования.

Другой способ — указание в одном из столбцов даты включения, а в другом столбце — даты наступления исхода или выбытия из исследования. В этом случае период времени рассчитывается путем вычитания дат. При этом надо строго соблюдать формат представления даты, чтобы результат вычитания оказался правильным.

### 3.1.7. Проверка данных

После внесения данных в электронную таблицу следует провести проверку данных с целью выявления возможных ошибок двух типов:

- ошибки ввода (набора);
- ошибки измерений.

Наиболее просто установить ошибки набора. Для этого можно поступить следующим образом. Щелкнув на имени столбца два раза, Вы увидите окно спецификации признака, внесённого в этот столбец (рис. 3.3). Нажав на кнопку “Знач./статист.”, можно увидеть все значения этого признака.

Среди них иногда можно заметить некоторые опечатки (например, “2,,34”). Это явные ошибки набора. Их следует исправить, предварительно сверившись с первоисточником данных (например, лабораторным журналом).



Переменная 3

Имя:  Код ПД:

Формат отображения

Ширина столб.:  Дес. разр.:

Тип:	Представление:
Число	1 000.000; -1 000.000
Дата	1 000.000; -1 000.000
Время	1 000.000; (1 000.000)
Научный	1 000.000; (1 000.000)
Денежный	
Проценты	

Длинное имя (метка, связь или формула с [Функциями](#)):

Примеры: Метка: Валовой доход в 1991      Формулы: =v1 + v2;  
Связь: @Excel\c:\file.xlsr2c2.r4c4      = (v1>0)\*AGE + v3

OK  
Отмена  
<< >>  
Все парам.  
Текст. знач.  
Знач./статист.  
Графики

Рис. 3.3. Окно спецификаций переменной (признака).

Другие значения могут быть слишком малыми или слишком большими, абсурдными с точки зрения биологических законов или методических подходов (например, систолическое артериальное давление 0.11). Это возможно в двух ситуациях:

1) при ошибке набора; ее легко устранить, сверившись с первоисточником данных;

2) при ошибке измерения (артефакт). В этом случае следует обсудить ситуацию с сотрудниками, непосредственно проводившими измерения. При определенных обстоятельствах может быть принято решение об исключении этого значения из исследования. Тогда необходимо просто удалить это значение из электронной таблицы, оставив ячейку пустой.

Сложнее в тех случаях, когда в принципе какое-либо слишком малое или, наоборот, слишком большое значение возможно, но оно резко отличается от остальных значений признака в данной выборке (например, в 10 раз). В первую очередь здесь надо также на всякий случай удостовериться, что это не ошиб-

ка набора. Если же такого рода ошибки не произошло, то это, вероятнее всего, так называемое выпадающее значение (выброс). О работе с такими значениями признака см. раздел 7.3.5.

### 3.1.8. Перенос данных, подготовленных в других программах, в таблицу данных пакета STATISTICA

Возможен перенос данных из файлов всех наиболее распространенных форматов, в том числе форматов \*.dbf, \*.db, \*.xls, ASCII и ряда других. Ниже мы рассмотрим два наиболее распространенных способа внесения данных из других программ.

**Использование буфера обмена MS Windows.** Для переноса данных из открытого окна другой электронной таблицы какого-либо из приложений, работающих в среде MS Windows (например, таблицы MS Excel или MS Word), может использоваться буфер обмена Windows, т.е. использование операций “Копирование” (“Copy”) и “Вставка” (“Paste”). Такой перенос неудобен потому, что при этом не переносятся заголовки столбцов таблицы MS Excel в заголовки таблицы данных ППП STATISTICA, поэтому в дальнейшем необходимо заново задать названия столбцов в таблице данных ППП STATISTICA.

**Перенос данных из программы MS Excel.** Если данные уже подготовлены в пакете MS Excel версий 7.0, 8.0 или 2000, то для последующего переноса данных в ППП STATISTICA версии 4.\* или 5.\* потребуется предварительно сохранить их в формате Excel 3.0 или 4.0. Для переноса в STATISTICA'99 (версия 5.5) предварительного сохранения в форматах более ранних версий MS Excel не требуется.

**Замечание.** *Файлы данных, созданные или отредактированные в ППП STATISTICA версии 5.5, не могут быть потом открыты в ППП STATISTICA более ранних версий (например, 4.3). Это аналогично ситуации с невозможностью открытия в более ранних версиях MS Word документов, созданных в более поздних версиях того же текстового редактора. Однако в отличие от MS Word в ППП STATISTICA 5.5 не предусмотрена возможность сохранения файлов данных в формате более ранних версий этой программы.*

Перенос данных осуществляется в модуле “Управление данными”.

**STATISTICA:**

- Модуль “Управление данными” (“Data management”)
- Меню “Анализ” (“Analysis”)
- Процедура “Импорт данных” (“Import”)  
(или меню “Файл — Импорт”)

или

- Модуль “Управление данными” (“Data management”)
- Меню “Файл” (“File”)
- Процедура “Импорт — Быстрый”  
(“Import Data”)

После обращения к указанным командам необходимо следующее:

- 1) указать файл, подлежащий переносу (“импорту”);
- 2) подтвердить автоматически определенный формат, отредактировав опции (рис. 3.4), в том числе рекомендуется выбрать опцию “Имена переменных из 1-й строки диапазона” (“Variables names” — “from first row of selected range”);
- 3) подтвердить выбранные опции (“OK”).

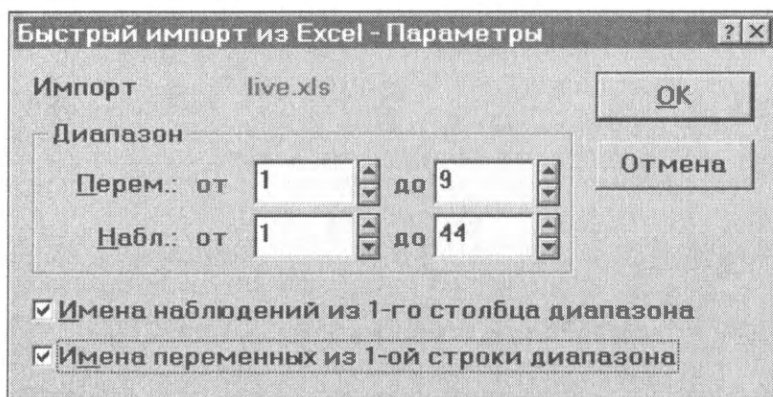


Рис. 3.4. Диалоговое окно импорта данных.

## 3.2. Подготовка вторичных данных

В ряде случаев исследователь не имеет на руках данных о каждом больном, а располагает вторичными (обобщенными)

данными, которые могут быть представлены следующим образом:

- в виде частот наблюдений (объектов исследования); такие данные подлежат дальнейшему статистическому анализу;
- в виде рассчитанных средних значений и каких-либо параметров вариабельности; такие данные не следует подвергать дальнейшему анализу, поскольку они часто уже содержат ошибки, возникшие при описании выборки (см. главу 7).

В первом случае для статистического анализа данные должны быть представлены одним из двух способов:

1. В виде таблицы; каждому столбцу таблицы соответствует какая-либо из групп (две или более), а в строках приведены частоты значений изучаемых признаков. Ими могут быть, например повозрастное распределение участников исследования, исход заболевания, наличие или отсутствие определенного симптома. Данные такого типа подлежат статистическому анализу лишь одного типа — проверке гипотезы о различии распределений изучаемого признака в изучаемых группах (см. главу 10).

2. В виде процентов (долей) и общего числа наблюдений (объектов исследования), принятых за 100%. В случае двух групп и двух значений признака данные не нужно вносить в таблицу данных. Они вводятся в диалоговое окно непосредственно при выполнении анализа (см. раздел 11.2).

## **Глава 4. Управление данными и сохранение результатов анализа в пакете прикладных программ STATISTICA**

Управление данными в ППП *STATISTICA* производится в модуле “Управление данными” (“Data Management”), а также частично во всех остальных модулях ППП. Обычные приемы работы с электронными таблицами, принятые в среде Windows (в том числе в пакете MS Excel), применимы и в ППП *STATISTICA*.

Подробное описание операций управления данными можно найти в справочных файлах ППП *STATISTICA* и “Кратком руководстве”, поставляемом в комплекте с программным обеспечением, а также литературе справочно-учебного характера по данному пакету, например в книгах В.П. Боровикова и соавт. [13—16].

Мы лишь очень кратко остановимся на некоторых наиболее часто используемых процедурах модуля “Управление данными” и функциях, доступных из любого модуля пакета.

### **4.1. Отбор строк таблицы данных по условию**

В большинстве случаев исходный файл данных содержит сведения об объектах исследования, относящихся к разным группам. При этом каждый объект исследования описан в отдельной строке таблицы данных, а номер группы заносится в отдельный столбец (такой признак называется группирующим). Для выполнения статистического анализа для какой-либо одной группы объектов или объектов, объединенных каким-либо иным общим

свойством, в ППП *STATISTICA* существует опция, доступная из всех модулей анализа. В правой части первого диалогового окна любого модуля анализа находится кнопка “У” (“Условия выбора наблюдений” (объектов исследования), “Select cases”). Нажав на нее, можно задать условие отбора объектов исследования. Например, `age < 50` или `group = 3`.

## 4.2. Формирование подгрупп (разделение файлов)

В ряде случаев стоит задача разделения одного файла данных на несколько файлов в соответствии с каким-либо условием. Операция выполняется следующим образом:

**Данные:** разделяемый файл.

**STATISTICA:**

- Модуль “Управление данными” (“Data management”)
- Меню “Анализ” (“Analysis”)
- Процедура “Подмножество” (“Create subset from data file”) (рис. 4.1)

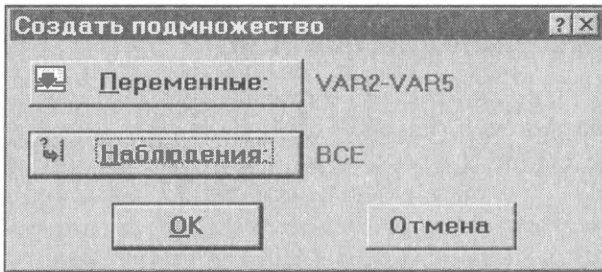


Рис. 4.1. Диалоговое окно процедуры “Подмножество” (модуль “Управление данными”).

Затем необходимо указать те переменные (признаки), которые необходимо вынести в подгруппу (мы рекомендуем указывать здесь “Все”), а также сформулировать условие, по которому будет формироваться подгруппа (совокупность объектов исследования, соответствующих заданному условию). Обычно таким

условием является выбор объектов с определенным значением какого-либо признака.

*Пример:* указание “gender=1” в опции “Наблюдения” (объекты исследования) означает, что в подгруппу будут отобраны все объекты исследования, у которых значение признака “gender” равно единице.

После выполнения этой операции в текущем окне данных окажется файл, содержащий искомое подмножество исходного файла. Его необходимо сохранить (“Файл — Сохранить как...”, “File — Save as...”) под каким-либо именем, отличающимся от имени исходного файла.

Если после этого необходимо выделить другую подгруппу из исходного файла, следует вновь открыть исходный файл данных и повторить процедуру с другим условием отбора и т.д. Таким образом будет создана серия файлов, образованных из исходного файла данных, которые будут отличаться друг от друга значением переменной (признака), по которой проходило разделение исходного файла.

*Пример:* из общего файла данных будут созданы несколько файлов с объектами исследования, соответствующими разным возрастным группам (например, 20—29 лет, 30—39 лет и т.д.).

### 4.3. Объединение файлов

Существуют два способа объединения файлов:

1. **Объединение столбцов** (переменных, признаков). Этот способ используется в тех случаях, когда одни и те же объекты исследования описаны разными признаками в разных файлах. При этом второй файл присоединяется к первому справа.

Для объединения столбцов предлагаются три режима:

- *нереляционный*; при этом файлы просто объединяются путем добавления столбцов присоединяемого файла справа. При этом происходит выравнивание сверху, т.е. к первой строке первого файла присоединяется первая строка второго файла. Если число строк в объединяемых файлах различно, то в новом файле число строк соответствует наибольшему из этих чисел;
- *реляционный*; файлы объединяются с учетом ключевого признака (идентификатора) строк таблицы данных. Такой режим должен использоваться, если одни и те же объекты ис-

следования в файлах расположены в разном порядке (последовательности);

- *реляционный иерархический*; это особый вариант реляционного объединения. Используется, если одной строке первого файла соответствует несколько строк второго файла.

2. **Объединение строк** (объектов исследования). Используется для объединения разных групп объектов исследования, описанных одинаковыми признаками. При этом второй файл присоединяется к первому снизу.

**Внимание!** Необходимо проследить, чтобы в объединяемых файлах число и порядок признаков в столбцах были одинаковыми.

**Данные:** один из объединяемых файлов.

#### STATISTICA:

- Модуль “Управление данными” (“Data management”)
- Меню “Анализ” (“Analysis”)
- Процедура “Объединить” (“Merge two data files”) (рис. 4.2)

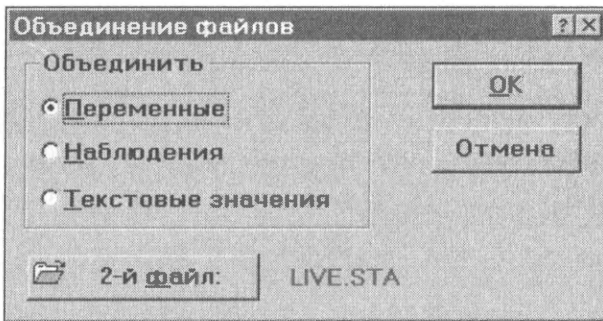


Рис. 4.2. Диалоговое окно процедуры “Объединить” (модуль “Управление данными”).

## 4.4. Операции со столбцами

Операции со столбцами доступны в любом модуле пакета путем активизации кнопки “Переменные” (“Vars”) (рис. 4.3):

- “Добавить” (“Add”) — добавление указанного числа столбцов после указанного столбца;





Рис. 4.3. Меню “Переменные” (может быть вызвано в любом из модулей пакета). Фрагмент интерфейса.

- “Переместить” (“Move”) — перенос столбца вправо или влево;
- “Копировать” (“Copy”) — копирование значений столбца и вставка их в другой столбец;
- “Удалить” (“Delete”) — удаление одного столбца и более;
- “Текущие спецификации” (“Current specs”) — диалоговое окно для указания имени переменной (признака), формата переменной (признака), комментария или формулы для расчета значений переменной (признака) и другие опции (вызывается также двойным щелчком на заголовке столбца в таблице данных);
- “Все спецификации” (“All specs”) — описание текущих спецификаций всех столбцов;
- “Текстовые значения” (“Text values”) — присвоение текстовых значений числовым признакам (кодам);

- “Значения дат” (“Date values”) — операции над датами, заданными в одном или нескольких столбцах;
- “Пересчитать” (“Recalculate”) — пересчет по заданной формуле значений текущего или другого столбца;
- “Сдвинуть значения столбца по вертикали” (“Shift (lag)”);
- “Ранжировать” (“Rank”) в порядке возрастания или убывания значений признака;
- “Перекодировать” (“Recode”) — переменной приписываются коды в соответствии с указанными условиями (например, при разбиении количественной переменной, или признака, на интервалы, т.е. превращение количественной переменной, или признака, в качественную порядковую).

С помощью редактирования в окне определения текущих спецификаций можно производить различные вычисления, в том числе:

- пересчитывать значения признака путем умножения на какие-либо коэффициенты  
*Пример синтаксиса:* =Var1×100;
- рассчитывать значения новых признаков по значениям двух признаков и более  
*Пример синтаксиса:* =(Var2–Var1)/Var1;
- получать столбец, содержащий случайные числа (функция генератора случайных чисел), распределенные по равномерному закону.

*Пример синтаксиса:* =rnd(100), в этом случае в текущем столбце будут выведены случайная последовательность чисел от 0 до 99.

**Замечание.** Подробное описание способов написания формул для различных вычислений можно найти в документации ППП STATISTICA.

## 4.5. Операции со строками

Операции со строками доступны в любом модуле пакета путем активизации кнопки “Наблюдения” (“Cases”) (рис. 4.4):

- “Добавить” (“Add”) — добавить указанное число строк после указанной строки;
- “Перенести” (“Move”) строки;
- “Копировать” (“Copy”) значения в строке;
- “Удалить” (“Delete”) — удалить указанный диапазон строк;



Рис. 4.4. Меню “Наблюдения” (может быть вызвано в любом из модулей пакета). Фрагмент интерфейса.

— “Имена” (“Names”) — указать имена строк (объектов исследования), например фамилии больных.

## 4.6. Сохранение результатов статистического анализа

Результаты всех видов статистического анализа в ППП *STATISTICA* представляются в отдельном окне (“Scrollsheet”), представляющем собой обычно таблицу. Эти результаты могут быть сохранены для последующего представления следующими способами:

1) сохранение в форматах TXT или RTF. Производится путем выделения (“Select”) нужного фрагмента таблицы и копирования (“Copy”) его в буфер обмена, а затем вставки (“Paste”) в необходимый раздел документа, например в окне “Вывод: NEW.RTF” (“new.txt”). Данное окно можно открыть, отметив опцию “Окно” (“Window”) в меню “Сервис — Принтер” (“Options — Print”). Недостатком этого способа является ограниченный объем памяти текстового окна вывода результатов в ППП *STATISTICA*;

2) копирование в буфер обмена нужного фрагмента таблицы результатов и вставка его в открытый в другом окне документ любого текстового редактора (например, MS Word или MS Excel).

При копировании фрагментов таблицы рекомендуется выделять все ячейки нужного столбца или строки. В этом случае сохраняются названия столбцов и строк таблицы результатов. Лишние результаты затем можно удалить непосредственно в текстовом редакторе. При этом предпочтительнее копировать таблицы результатов в MS Excel, чем в MS Word, так как в последнем случае содержание таблицы будет перенесено как текст вне таблицы.

При выведении результатов описательной статистики обычно бывает необходимо представление всей таблицы из окна результатов.

При выведении результатов проверки статистических гипотез обычно бывает необходимо представление только столбцов, содержащих рассчитанные значения статистического критерия и значения  $p$ .

Более подробно необходимая форма представления результатов для каждого вида анализа обсуждается в соответствующих разделах книги.

## **Глава 5. Основные принципы и методы статистического анализа**

Цель исследования всегда заключается в выявлении некоторых закономерностей на выборке и дальнейшей экстраполяции полученных результатов на всю генеральную совокупность (популяцию), из которой получена исследуемая выборка. Выявление закономерностей на выборке производится обычно путем решения следующих основных задач:

- описание группы (групп) объектов исследования;
- сравнение групп (или одной группы в разные моменты времени);
- исследование взаимосвязей признаков.

В статистике для решения этих задач существуют соответственно следующие основные подходы:

- статистическая оценка параметров распределения;
- проверка статистических гипотез;
- статистическое моделирование.

Рассмотрим основные принципы этих подходов.

### **5.1. Описательная статистика и статистическая оценка**

Статистический анализ основан на следующем допущении: то, что верно для случайной выборки, верно и для генеральной совокупности (популяции), из которой эта выборка получена. Например, если в ходе исследования выявлено, что применение тромболитиков при определенном типе ишемического инсульта улучшает прогноз для больного, то можно считать, что для всех больных с ишемическим инсультом данного типа это будет справедливо. Однако сделать истинно случайную выборку из генеральной совокупности бывает практически невозможно. Это, например, может быть связано с тем, что объекты исследования (больные) проживают на разных континентах. Поэтому обычно исследователю следует стремиться к тому, чтобы выборка была

репрезентативной по отношению к изучаемой популяции, т.е. достаточно адекватно отражающей все возможные аспекты изучаемого состояния или заболевания в популяции. Для достижения этой цели необходимо четко сформулировать и в дальнейшем строго соблюдать критерии включения и исключения, а также избегать тенденциозности в решении вопроса о включении либо о не включении того или иного больного в исследование.

Обычно популяционное значение параметра (среднее значение, медиану, долю и т.д.) узнать невозможно (исключение составляют случаи, когда исследование проводится на группе, которая включает **всех** членов популяции). Однако популяционное значение параметра можно оценить по выборке. Точность такой оценки зависит от метода измерения (ошибки измерения), объема и репрезентативности выборки (ошибка выборки) и биологической вариации.

Описание распределения признака в выборке проводится путем оценки значений ее параметров, характеризующих центральную тенденцию и рассеяние наблюдений (объектов исследования) по области значений признака. Способы такого описания, зависящие от типа признака и его распределения, описаны в главах 7 и 9.

*Распределением признака в выборке* называется совокупность частот наблюдений (объектов исследования) для каждого интервала значений признака в конкретной выборке.

*Распределением признака в генеральной совокупности (популяции)* называется совокупность частот наблюдений (объектов исследования) для каждого интервала значений признака в генеральной совокупности.

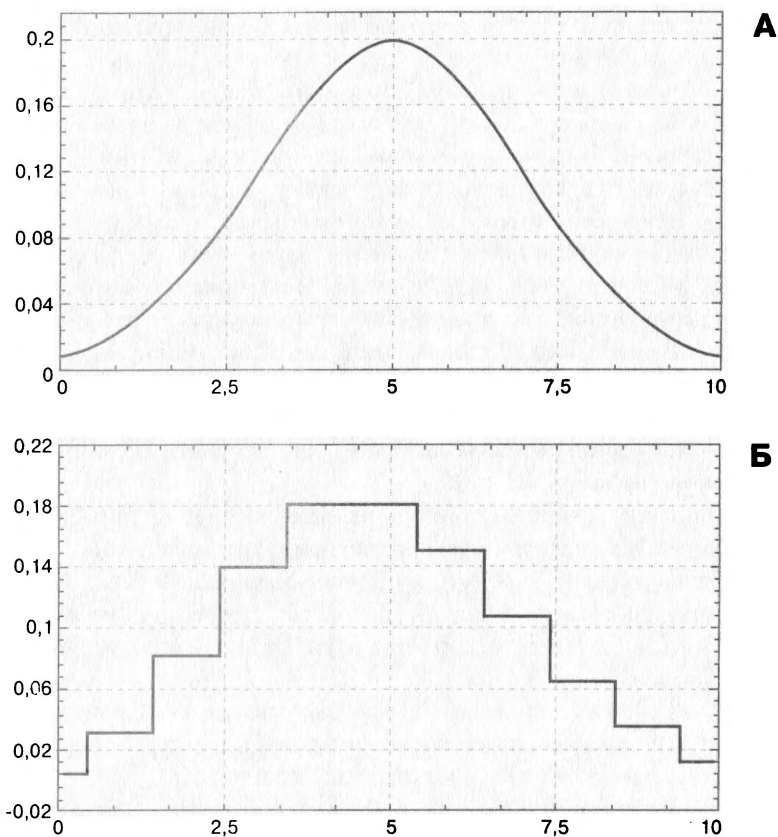
Основными типами и видами распределений признаков являются следующие (примеры на рис. 5.1).

1. Дискретные (для дискретных признаков):

- биномиальное;
- распределение Пуассона;
- распределение Бернулли.

2. Непрерывные (для непрерывных признаков):

- нормальное (гауссово, или распределение Гаусса);
- логнормальное;
- постоянное;
- экспоненциальное;
- хи-квадрат ( $\chi^2$ ).



**Рис. 5.1. Распределения признаков.**

А — непрерывное нормальное; Б — дискретное биномиальное.

Очевидно, что если из одной и той же генеральной совокупности будут независимо набраны несколько выборок, то значения их параметров (например, средние значения некоего признака) будут несколько различаться в силу случайности. При этом каждое из них будет несколько отличаться от значения этого параметра в генеральной совокупности, или популяции (так называемого популяционного значения). Исследователю же важно знать, насколько может быть велика его уверенность в том, что рассчитанное им по выборке значение (оценка) близко к популяционному.

Вариабельность значений параметра в разных выборках из одной генеральной совокупности тем ниже, чем больше объем выборок. Значение вариации оценки может быть выражено с помощью доверительного интервала.

**Доверительный интервал (ДИ)** — интервал значений признака, рассчитанный для какого-либо параметра (например, среднего значения признака) по выборке и с определенной вероятностью (например, 95%) включающий истинное значение этого параметра во всей генеральной совокупности.

ДИ всегда связан с каким-либо уровнем доверия, уверенности. Напомним, что все оценки параметров признаков генеральной совокупности, полученные на основе анализа данных выборки, не являются абсолютно истинными. Они истинны лишь с некоторой долей вероятности. Так, если мы выбираем доверительный коэффициент (ДК; степень уверенности, выраженная в процентах; вероятность того, что данный интервал содержит истинное значение параметра) равным 95%, то это означает, что в 95 выборках из 100, сделанных таким же способом из генеральной совокупности объектов исследования, оценка параметра признака будет находиться в рассчитанном нами ДИ.

Ширина ДИ зависит от объема выборки и вариабельности в выборке. В свою очередь, чем шире ДИ, тем менее точной оказывается выборочная оценка популяционного среднего, и наоборот. При увеличении числа наблюдений (объектов исследования) ДИ сужается, и точность оценки увеличивается.

Построение ДИ для различных параметров проводится по-разному, что описано в соответствующих главах книги.

В настоящее время в зарубежной научной медицинской литературе представление результатов исследования с использованием ДИ получает все большее распространение, а в ряде изданий представление ДИ для основных результатов исследований стало обязательным требованием. Это обусловлено следующим:

1. ДИ для какого-либо параметра позволяет наглядно представить спектр возможных значений этого параметра, которые могут быть рассчитаны по другим выборкам из той же генеральной совокупности.

2. ДИ, помимо использования с целью оценки параметров описания выборки, может также использоваться для сравнения выборок (способ, аналогичный проверке статистических гипотез — см. ниже).



3. При указании ДИ нет необходимости следить за видом распределения количественного признака, так как такое описание спектра популяционных значений параметра подходит как для нормально распределенных, так и распределенных по другим законам данных. В случае нормального распределения ДИ для среднего значения симметричен относительно этого среднего значения, в других случаях — несимметричен относительно него.

Указание ДИ экономит место в таблицах в случае негауссовых распределений, так как отпадает необходимость указания медианы и границ интерпроцентильного интервала с указанием процентилей.

После того как определен тип данных, необходимо выяснить, какие способы описания выборок и методы статистического анализа следует применять. Если признак является качественным, то его распределение следует описывать способами, изложенными в главе 9. Если же признак количественный, то возникает задача определения вида его распределения в зависимости от его типа — непрерывный или дискретный. На практике важно бывает знать хотя бы то, подчиняется ли распределение непрерывного признака закону нормального (гауссова) распределения или нет.

## 5.2. Проверка статистических гипотез

Наиболее частыми задачами медицинских и биологических исследований, для решения которых оказывается необходимым сформулировать статистические гипотезы, являются следующие:

- анализ соответствия распределения значений признака в изучаемой группе какому-либо определенному закону (например, анализ соответствия распределения нормальному закону);
- сравнение групп по параметрам распределений признака (например, по средним значениям, дисперсиям).

Для решения любой подобной задачи формулируются две статистические гипотезы:

- нулевая гипотеза  $H_0$  — гипотеза об отсутствии различий между группами, либо гипотеза об определенных значениях параметров, либо гипотеза о соответствии распределения нормальному закону;
- альтернативная гипотеза  $H_1$  — гипотеза о существовании различий между группами либо гипотеза об отличающихся от

заданных значениях параметров, либо гипотеза о несоответствии распределения нормальному закону.

Обычно нулевая гипотеза формулируется таким образом, чтобы она была противоположна той исследовательской (медицинской, биологической) гипотезе, которая послужила поводом для проведения исследования. Для проверки нулевой гипотезы применяют статистические методы (тесты, критерии).

В результате проверки статистических гипотез возникают следующие ситуации (рис. 5.2):

- $H_0$  неверна и отклонена согласно статистическому критерию — истинноположительный результат;
- $H_0$  верна, но ошибочно отклонена согласно статистическому критерию — ложноположительный результат (ошибка первого рода, или  $\alpha$ -ошибка);
- $H_0$  неверна, но ошибочно не отклонена согласно статистическому критерию — ложноотрицательный результат (ошибка второго рода, или  $\beta$ -ошибка);
- $H_0$  верна и не отклонена согласно статистическому критерию — истинноотрицательный результат.

		В генеральной совокупности (популяции)	
		$H_0$ неверна	$H_0$ верна
В статистическом тесте	$H_0$ отклонена	Истинно-положительный (ИП) результат	Ложно-положительный (ЛП) результат (ошибка первого рода, или $\alpha$ -ошибка)
	$H_0$ не отклонена	Ложно-отрицательный (ЛО) результат (ошибка второго рода, или $\beta$ -ошибка)	Истинно-отрицательный (ИО) результат

Рис. 5.2. Возможные решения при различных соотношениях результатов статистического теста и истинной ситуации в генеральной совокупности (популяции).

Конечно, желательно минимизировать обе ошибки одновременно, но при заданном объеме выборки это невозможно,

поскольку при уменьшении величины одной из них неизбежно будет возрастать величина другой.

Ошибка первого рода иначе называется уровнем статистической значимости (обозначаемого как  $p_0$ ). Уровень значимости — это максимально приемлемая для исследователя вероятность ошибочно отклонить нулевую гипотезу, когда на самом деле она верна, т.е. допускаемая исследователем величина ошибки первого рода. Величина уровня значимости устанавливается исследователем произвольно, однако обычно принимается равным 0,05 либо 0,01, либо 0,001.

В ходе применения статистического метода вычисляется значение тестовой статистики (например, при применении критерия Стьюдента — значение  $t$ ), а также соответствующее ему и числу степеней свободы значение  $p$ .

Значение  $p$  — это рассчитанная в ходе статистического теста вероятность ошибочного отклонения нулевой гипотезы об отсутствии различий. Другое определение: значение  $p$  — это вероятность получить данные анализируемых выборок в случае справедливости нулевой гипотезы (в частности, в случае отсутствия различий групп). Третье определение: значение  $p$  — это вероятность справедливости нулевой гипотезы.

Метод проверки статистических гипотез заключается в сравнении полученного значения  $p$  с принятым уровнем значимости:

— если рассчитанное в статистическом тесте значение  $p$  оказывается больше принятого (на усмотрение исследователя) уровня значимости (обычно 0,05), то нулевую гипотезу  $H_0$  не отклоняют. Различия групп называются статистически незначимыми. Подчеркнем особо, что в этом случае нельзя утверждать, что нулевая гипотеза верна;

— если же значение  $p$  оказывается меньше уровня значимости, то нулевую гипотезу  $H_0$  отклоняют, при этом следует принять альтернативную гипотезу  $H_1$ . В данном случае различия групп называют статистически значимыми (при  $p < 0,05$ ) или статистически высокозначимыми (при  $p < 0,01$ ).

**Замечание 1.** К сожалению, до настоящего времени очень часто вместо термина “статистически значимый” в отечественных публикациях ошибочно используется термин “достоверный”, имеющий в статистике другой смысл.

**Замечание 2.** Часто статистически незначимые результаты несправедливо рассматриваются исследователями как не-

*удача работы, что подчеркивается такими распространенными выражениями как “отрицательный результат”, “не удалось достичь статистической значимости”. Такие исследования часто не публикуются, что приводит к возникновению систематической ошибки, обусловленной преимущественным опубликованием положительных результатов исследования, когда опубликованными оказываются лишь те исследования, в которых результаты оказались статистически значимыми, т.е. лишь часть от всех выполненных исследований по данной проблеме. Вместе с тем статистически незначимые результаты являются не менее важными в контексте общенаучного процесса.*

Как уже упоминалось выше, обычно за величину уровня статистической значимости принимаются 0,05, 0,01 или 0,001 (т.е. допускается  $\alpha$ -ошибка в 5%, 1% или 0,1%). Однако в связи с тем, что выбор критического уровня статистической значимости — произвольный акт исследователя, в настоящее время в научных публикациях (статьях, отчетах, диссертациях) рекомендуется указывать точное значение  $p$ , а не описывать результаты как статистически значимые (например, при  $p < 0,05$ ) или незначимые (при  $p > 0,05$ ), т.е. указывать лишь интервал, в котором находится вычисленное значение  $p$ . Очевидно, например, что результаты  $p = 0,051$  и  $p = 0,049$  следует интерпретировать практически одинаково. Указание точного значения  $p$  позволяет читателю самостоятельно интерпретировать статистическую значимость результата. Значение  $p$  принято указывать в тексте статей с точностью до трех десятичных знаков, и только в случае, если  $p$  меньше 0,001, то в формате “ $p < 0,001$ ”, т.е. в формате указания лишь интервала значений.

Величина  $\beta$ -ошибки зависит от величины ожидаемого эффекта и объема выборки. Часто величина  $\beta$ -ошибки задается значением 0,2 (20%). С учетом допустимого уровня этого параметра существует возможность рассчитать объем выборки, необходимой для выявления эффекта определенной величины (подробно см. раздел 15.3).

Чем меньше величина  $p$ , тем менее вероятна справедливость нулевой гипотезы, однако величина  $p$  никак не отражает величину различий между группами. Именно поэтому получил широкое распространение другой подход к сравнению групп — подход, основанный на построении и сравнении ДИ для оценки различий между группами или изменений в одной группе во времени.

По сути два упомянутых подхода сравнения групп — проверка статистических гипотез и сравнение ДИ — основаны на одних и тех же статистических моделях и предположениях. Однако способ сравнения групп с помощью ДИ зачастую упрощает исследователям анализ и облегчает интерпретацию результатов. Это обусловлено следующими обстоятельствами.

— Поскольку ДИ имеет такие же единицы измерения, что и изучаемый признак (в отличие от безразмерной величины  $p$ , вычисляемой при проверке статистических гипотез), интерпретация сопоставления выборок с использованием ДИ легче, чем при статистической проверке гипотез.

— Если значение  $p$  обычно интерпретируется либо как статистически значимый (позитивный), либо как статистически незначимый (негативный) результат, то ДИ, представляющий интервал значений, на котором истинное изменение должно произойти, позволяет читателю интерпретировать значение изменений на любом конце ДИ. Например, если один конец ДИ содержит клинически важные значения, а другой — нет, результаты могут быть оценены как неоднозначные (ни позитивные, ни негативные).

По этим причинам ДИ предпочтительнее, чем  $p$ , хотя обычно при представлении результатов исследования желательно приводить и ту, и другую величину.

**Замечание.** *Поскольку построение ДИ возможно не всегда, проверка статистических гипотез с помощью статистических критериев остается основным подходом при сравнении групп.*

Если признак определен как количественный, то задача исследователя заключается в том, чтобы установить, является ли его распределение нормальным (гауссовым). По данным В.П. Леонова, П.В. Ижевского (1995), лишь около 20% распределений количественных признаков, встречающихся в медико-биологических исследованиях, являются приближенно нормальными. Статистическая гипотеза о соответствии распределения нормальному может быть проверена специальными статистическими методами (см. раздел 7.1). Это необходимо для того, чтобы выбрать наиболее адекватные и мощные методы статистического анализа данных

Проверка нормальности распределения признака важна еще и по другой причине. Известно, что нормальное распределение возникает в том случае, когда вариабельность значений этого признака обусловлена влиянием множества причин, каждая из кото-

рых вносит минимальный вклад. Для конкретного признака это означает его взаимосвязь со многими подсистемами организма, а не с одной-двумя из них.

### 5.3. Статистическое моделирование

Статистическое моделирование — это построение математических моделей взаимодействия между собой двух признаков и более. Цель статистического моделирования — прогнозирование значения одного признака по значениям других признаков. Статистическая модель не позволяет выявлять биологические закономерности, но может лишь имитировать “поведение” одного признака при известном “поведении” других признаков.

Методы статистического моделирования описаны в главах 13—14.

При использовании методов статистического моделирования необходимо учитывать следующие два основных положения:

- 1) каждый метод статистического моделирования имеет определенные предположения в отношении распределений признаков. Если эти предположения не обоснованы (не выполняются), то модель необоснованна и ее не следует использовать;
- 2) статистические модели необходимо проверять на работоспособность (валидизировать; см. раздел 13.6).

### 5.4. Статистическая и клиническая значимость полученных результатов

Термин “статистически значимый” часто ошибочно смешивается с термином “клинически значимый”. Однако статистически значимый результат следует интерпретировать как клинически значимый лишь в том случае, если он потенциально может изменить клиническую практику, т.е. тактику диагностики, лечения. Часто статистически значимые, но клинически незначимые результаты получают на больших выборках. Так, например, в ходе большого клинического испытания может быть получен статистически значимый результат, заключающийся в снижении систолического АД на 3 мм рт. ст. в группе лечения новым препаратом. Однако этот результат не является клинически значимым хотя бы потому, что находится за пределами разрешающей способности рутинного метода измерения АД.

Возможна и обратная ситуация, когда статистически незначимый результат может оказаться клинически значимым. Такие результаты часто получают на небольших выборках. Например, какой-либо новый хирургический прием позволяет снизить летальность на 30%, хотя результат и не является статистически значимым. Однако новый факт наверняка заставит исследователей продолжить работу в этом направлении.

Следует также помнить о том, что интерпретация результата как клинически значимого или незначимого зависит от содержательного вопроса исследования. Так, если какой-либо препарат позволяет снизить на 50% частоту эпилептических приступов у некоторых больных, а на других никак не действует, то он вряд ли может рассматриваться как препарат выбора по отношению к другому лекарственному средству, применение которого позволяет снизить частоту приступов на 20%, но зато у всех больных. Указание значения  $p$  в такой ситуации может оказаться совершенно недостаточным для оценки клинической значимости результатов исследования. Именно в этих случаях полезно приводить ДИ для основных результатов.

## 5.5. Первичный и вторичный анализ данных

Различают два типа статистического анализа данных — первичный (запланированный) и вторичный (незапланированный).

*Первичный анализ данных* — это изучение биологических и медицинских закономерностей, существование которых предполагается исследователем и которые являются собственно предметом исследования. Перечень задач исследования должен включать описание признаков, изучение которых учтено при планировании исследования. Первичный анализ данных служит обычно для проверки заранее (априори) сформулированных научных гипотез исследователя.

Однако в связи с тем что получение данных в области клинической медицины сопряжено со значительными организационными, временными, финансовыми затратами, обычно у исследователя возникает стремление сделать анализ всех данных настолько полно, чтобы можно было выявить и заранее неизвестные (и не предполагавшиеся) закономерности. Такой анализ данных

называется “просеивание данных<sup>1</sup>” и обычно квалифицируется как исследовательский, поисковый, разведочный, *вторичный*. Результаты такого анализа следует интерпретировать более осторожно. Осторожность должна соблюдаться в связи тем что результаты вторичного анализа в большинстве случаев не свободны от систематических ошибок в связи с несопоставимостью подгрупп и другими факторами. Несопоставимость групп является естественным следствием того, что априори неизвестные закономерности не могли быть учтены при планировании исследования. Результаты вторичного анализа данных не могут служить в качестве доказательств той или иной научной гипотезы, но обычно воспринимаются как основание для выдвижения научных гипотез.

## 5.6. Классификация статистических методов

Методы статистического анализа данных принято классифицировать по нескольким принципам.

### 1. По количеству анализируемых признаков:

- одномерные (анализ каждого признака в отдельности);
- двумерные (одновременный анализ двух признаков, например анализ связей изучаемых признаков — ассоциации или корреляции);
- многофакторные (анализ трех признаков и более одновременно, например многофакторный дисперсионный анализ, многофакторный регрессионный анализ, дискриминантный анализ).

### 2. По статистическим принципам, лежащим в основе методов:

**методов:**

- параметрические. Применяются главным образом для анализа нормально распределенных количественных признаков;
- непараметрические. Применяются в остальных случаях:
  - а) для анализа количественных признаков независимо от вида их распределения;
  - б) для анализа качественных признаков.

В целом непараметрические методы (в случае использования их на малых выборках) являются менее мощными по сравнению с параметрическими, т.е. иногда не позволяют выявить

<sup>1</sup> Data dredging (англ.).



статистические закономерности, которые могут быть выявлены с помощью параметрических методов. В то же время непараметрические методы более надежны в случаях, когда есть сомнения в том, что анализируемый признак имеет нормальное распределение. Для нормально распределенных признаков параметрические и непараметрические методы дают близкие результаты.

### **3. По возможности учета имеющихся априори предположений:**

- односторонние тесты. Это тесты, учитывающие исходное (априорное) предположение о том, что в одной из групп распределение признака смещено в определенную сторону (в сторону увеличения либо уменьшения) по отношению к другой. Однако для того чтобы воспользоваться таким тестом, необходимо обосновать свое предположение;
- двусторонние тесты. Эти тесты используются в отсутствие исходного (априорного) предположения о том, что в одной из групп распределение признака смещено в определенную сторону (в сторону уменьшения или увеличения) по отношению к другой. Вычисляемое при этом значение  $p$  обычно примерно в 2 раза больше, чем для одностороннего теста. Двусторонние тесты рекомендуется использовать как можно шире.

### **4. По зависимости или независимости сопоставляемых выборок:**

- тесты для независимых выборок. Используются в случае, если при формировании выборок объекты исследования набирались в группы независимо друг от друга (например, были рандомизированы в две группы или лечились в разных больницах);
- тесты для зависимых выборок. Используются в двух случаях:
  - а) когда в динамике анализируются данные, полученные в одной группе больных (например, до и после лечения);
  - б) когда анализируются выборки, сформированные путем подбора пар по полу, возрасту, стадии заболевания и т.д.

Классификация наиболее важных статистических методов, которая может быть использована при выборе теста для решения конкретной задачи, представлена на рис. 5.3.

Заметим, что для двухмерного и одномерного анализа параметрические методы могут применяться лишь при условии, что все анализируемые одновременно признаки являются нормально распределенными. В противном случае должны применяться не-

Задача	Методы	
	параметрические (для количественных нормально распределенных признаков)	непараметрические (для количественных признаков независимо от вида распределения, а также для качественных — порядковых или номинальных — признаков)
Выполнение описательной статистики	Вычисление средних значений, средних квадратических отклонений и т.д.	Вычисление медиан и интерквартильных интервалов, пропорций
Сравнение двух независимых групп по одному признаку	<i>t</i> -критерий Стьюдента для независимых выборок	Критерии Манна—Уитни, Колмогорова—Смирнова, Вальда—Вольфовица, $\chi^2$ , точный критерий Фишера
Сравнение двух зависимых групп по одному признаку	<i>t</i> -критерий Стьюдента для зависимых выборок	Критерий Вилкоксона, критерий знаков, критерий МакНемара
Сравнение трех независимых групп и более по одному признаку	ANOVA	ANOVA по Краскелу—Уоллису, медианный критерий, критерий $\chi^2$
Сравнение трех зависимых групп и более по одному признаку	Критерий Кокрана	ANOVA по Фридмену, критерий Кокрана
Анализ взаимосвязи двух признаков	Корреляционный анализ по Пирсону	Критерий $\chi^2$ , корреляционный анализ по Спирмену, Кендаллу, гамма и др.
Одновременный анализ трех признаков и более	Регрессионный анализ Дискриминантный анализ Факторный анализ Кластерный анализ	Логистический регрессионный анализ, логлинейный анализ, анализ древовидных диаграмм, анализ конъюнкций и др.

Рис. 5.3. Рекомендуемые к использованию статистические критерии (методы) в зависимости от задачи исследования и типа данных.

параметрические методы. Например, если исследователя интересует корреляция двух признаков, один из которых является нормально распределенным, а второй — нет, то следует использовать непараметрические методы анализа корреляции (Спирмена, Кендалла, гамма).

Наше дальнейшее изложение построено в соответствии с логикой, которой рекомендуется придерживаться при анализе данных оригинальных исследований — от более простых методов к более сложным.

Обычно перед исследователем встает ряд вопросов, для ответа на которые необходимо воспользоваться различными методами статистического анализа данных. Заметим однако, что в одном исследовании для решения его задач необходимо использовать как можно меньшее число различных методов. Чрезмерное увлечение разнообразными способами анализа данных не оправдано. Анализ данных не может быть самоцелью, он служит лишь для выдвижения или проверки научных гипотез. В конечном счете объем и глубина анализа данных во многом определяются знаниями и навыками того, кто непосредственно выполняет этот этап исследования.

---

## Глава 6. Методы статистического анализа, содержащиеся в пакете прикладных программ *STATISTICA*

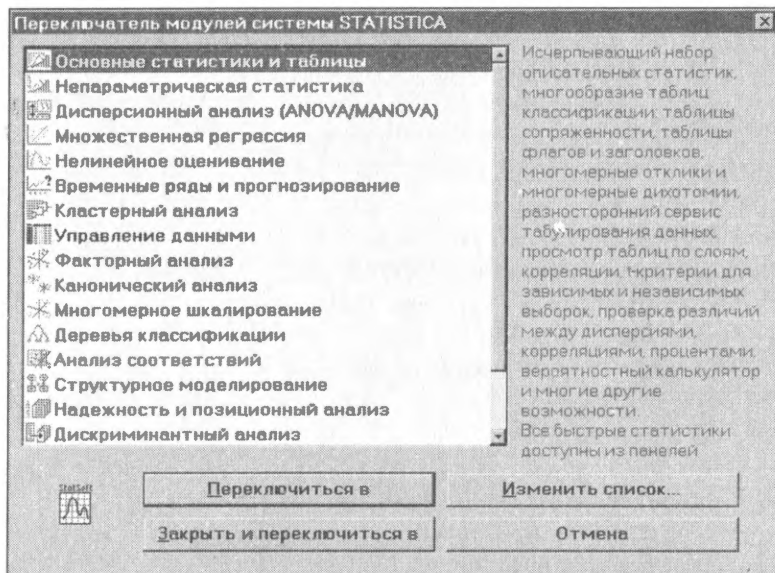
ППП *STATISTICA* состоит из 24 модулей:

- Основные статистики и таблицы ✓
- Непараметрические статистики ✓
- Дисперсионный анализ (ANOVA/MANOVA) ✓
- Множественный регрессионный анализ ✓
- Нелинейное оценивание ✓
- Временные ряды и прогнозирование
- Кластерный анализ ✓
- Управление данными ✓
- Факторный анализ ✓
- Канонический анализ
- Многомерное шкалирование
- Деревья классификации
- Анализ соответствий
- Структурное моделирование
- Надежность и позиционный анализ
- Дискриминантный анализ ✓
- Логлинейный анализ
- Анализ выживаемости (анализ времени до наступления исхода) ✓
- Карты контроля качества
- Анализ процессов
- Компоненты дисперсионного анализа
- Планирование эксперимента
- Командный язык
- Сервер файлов системы *STATISTICA*

Каждый из модулей представляет собой набор статистических методов для решения той или иной исследовательской задачи. В данном издании мы останавливаемся лишь на тех из них,

которые наиболее часто бывают необходимы исследователям в области медицины и биологии. Эти методы отмечены в списке модулей галочками.

На рис. 6.1 показано окно переключения модулей, которое загружается при запуске ППП (путем запуска файла `sta_win.exe`).



**Рис. 6.1.** Окно переключения модулей (может быть вызвано в любом из модулей пакета).

Вначале следует выбрать необходимый модуль. В дальнейшем существует возможность в рамках одного сеанса работы вызвать любые другие модули, каждый из которых открывается в отдельном окне. Переключение между модулями в любом из уже открытых модулей производится следующим образом.

### STATISTICA:

- Меню “Анализ” (“Analysis”)
- Строка меню “Другие виды анализа” (“Other Statistics”)

В любом из диалоговых окон статистических процедур существует возможность указания одного или нескольких анализируемых признаков (кнопка “Переменные”, “Variables”).

---

## Глава 7. Описание количественных признаков

Для описания количественных признаков используют так называемые описательные статистики, т.е. параметры распределения. Если исследование проведено с использованием малой выборки, то в таком случае необходимо приводить в тексте статьи или отчета таблицу данных. При этом зачастую отсутствует необходимость в описательной статистике.

Для того чтобы выбрать способ описания количественного признака, следует сначала установить, соответствует ли вид распределения значений изучаемого признака закону нормального распределения.

### 7.1. Анализ соответствия вида распределения признака закону нормального распределения

Первым этапом анализа количественных данных является анализ вида их распределения. Реально ни в одной выборке не может быть строго нормального распределения признака. Однако необходимо установить, отобрана ли эта выборка из генеральной совокупности, в которой изучаемый признак имеет нормальное распределение. Для простоты мы далее будем считать, что вид распределения признака в генеральной совокупности установлен по выборке.

Существуют несколько способов решения этой задачи. Первые два помогают визуально оценить близость распределения данного признака к нормальному:

**Первый способ.** Построить гистограмму распределения признака. Для этого в таблице данных достаточно щелкнуть правой кнопкой мыши на шапку соответствующего столбца, выбрать опцию “Быстрые стат. графики” (“Quick stats graphs”), затем последовательно опции “Гистограмма” (“Histogram of ...”) и

“Нормальная подгонка” (“Normal fit”). После этого можно визуально оценить, насколько диаграмма близка к колоколу нормального распределения, прорисованному красной линией.

**Второй способ.** Можно построить график функции распределения в специальных координатах. Из этого же меню “Быстрые стат. графики” (“Quick stats graphs”) выбираются опция “Вероятностный график” (“Probability plot of ...”), а затем опция “Нормальный” (“Normal probability”). После этого можно также визуально оценить, насколько близко к прямой лежат точки, отражающие отдельные объекты исследования.

Однако оба эти способа лишь качественные. Одним из количественных способов является оценка симметричности распределения признаков, имеющих только положительные значения, т.е. **третий способ.** Если  $s <^M / 2$  (среднее квадратическое отклонение меньше половины среднего арифметического), то распределение можно считать симметричным (симметричность — одна из основных характеристик нормального распределения).

Однако третий способ не позволяет оценить эксцесс — степень крутизны графика, или его плосковершинность. Поэтому для получения надежной оценки соответствия изучаемого распределения признака закону нормального распределения следует пользоваться **четвертым способом** — проверкой статистических гипотез о виде распределения, т.е. о том, извлечена ли выборка из генеральной совокупности, в которой изучаемый признак имеет нормальное (гауссово) распределение. Ведь от вида распределения зависят как выбор способа описания центральной тенденции и рассеяния значений признака, так и выбор методов дальнейшего анализа данных. По результатам анализа можно принять одну из гипотез:

- нулевую гипотезу (о том, что распределение исследуемого признака в генеральной совокупности соответствует закону нормального распределения);
- альтернативную гипотезу (о том, что распределение исследуемого признака в генеральной совокупности не соответствует закону нормального распределения).

Для принятия решения о виде распределения можно применять следующие критерии:

- Колмогорова—Смирнова. Применяется в тех случаях, когда среднее значение и среднее квадратическое отклонение признака известны априори, а не вычисляются по выборке;

- Лиллиефорса. Применяется в тех случаях, когда среднее значение и среднее квадратическое отклонение признака априори неизвестны и вычисляются по выборке;
- Шапиро—Уилка. Применяется также при исходно неизвестных среднем значении и среднем квадратическом отклонении признака. Данный критерий предпочтителен, так как является наиболее мощным и универсальным и при этом наиболее “строгим” из перечисленных.

### STATISTICA:

→ Модуль “Основные статистики и таблицы”

    (“Basic Statistics/Tables”) (рис. 7.1)

    → Подмодуль “Описательные статистики”

        (“Descriptive Statistics”) (рис. 7.2)

        → Процедуры:

            — “Критерии нормальности Колмогорова—Смирнова и Лиллиефорса”

            — “Критерий Шапиро—Уилка W”

Отметив галочками необходимые критерии, следует активизировать процедуры нажатием кнопки “Таблицы частот” (“Frequency tables”). Результаты анализа распределения каждого из анализируемых признаков (значения  $p$ ) будут выдаваться в отдельных окнах (рис. 7.3).

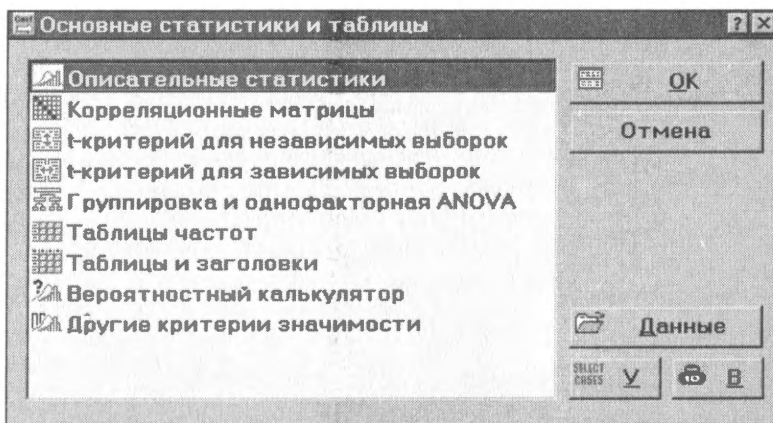


Рис. 7.1. Первое диалоговое окно модуля “Основные статистики и таблицы”.



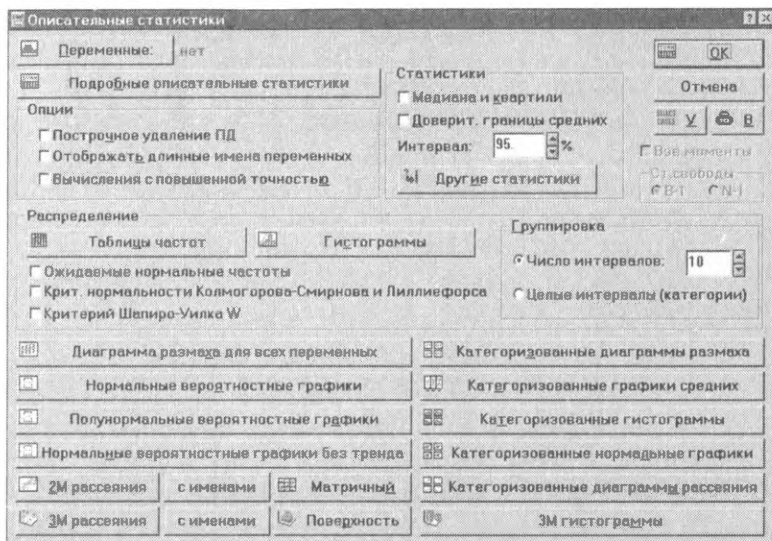


Рис. 7.2. Диалоговое окно подмодуля “Описательные статистики” (в модуле “Основные статистики и таблицы”).

Группа	Частота	Кумул. частота	Процент допуст.	Кумул. % допуст.	% всех наблщ.	Кумул. % от всех
-20.00 < x <= 0.0000	5	5	17.24138	17.2414	11.62791	11.6279
0.0000 < x <= 20.000	22	27	75.86207	93.1034	51.16279	62.7907
20.000 < x <= 40.000	0	27	0.00000	93.1034	0.00000	62.7907
40.000 < x <= 60.000	1	28	3.44828	96.5517	2.32558	65.1163
60.000 < x <= 80.000	0	28	0.00000	96.5517	0.00000	65.1163
80.000 < x <= 100.00	1	29	3.44828	100.0000	2.32558	67.4419
Пропущ.	14	43	48.27586		32.55814	100.0000

Рис. 7.3. Окно результатов критериев нормальности подмодуля “Описательные статистики”.

Процедуры для анализа вида распределения признака доступны также в другом подмодуле.

**STATISTICA:**

- Модуль “Основные статистики и таблицы” (“Basic Statistics/Tables”) (см. рис. 7.1)
- Подмодуль “Таблицы частот” (“Frequency tables”) (рис. 7.4)

→ Процедуры

- “Критерий К.—С., ср./ст. откл. известны”
- “Крит. Лиллиефорса, ср./ст. откл. неизвестны”
- “Критерий Шапиро—Уилка W”

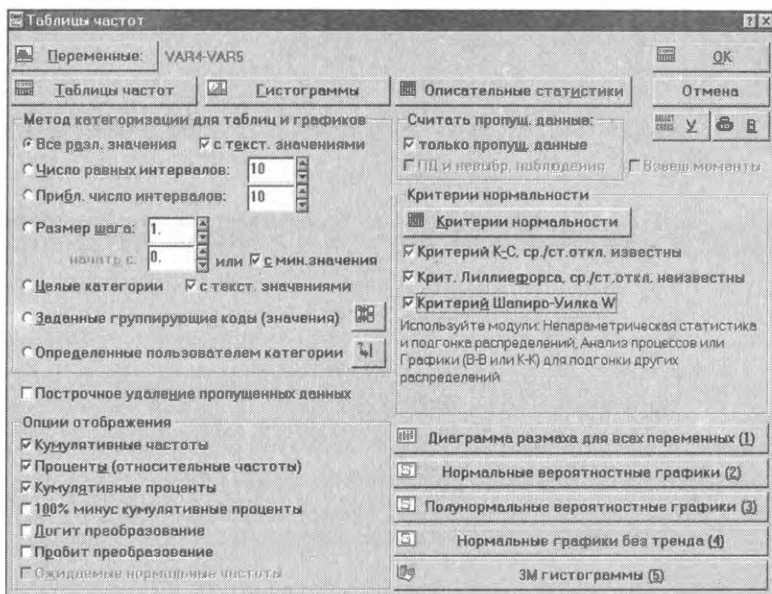


Рис. 7.4. Диалоговое окно подмодуля “Таблицы частот” (в модуле “Основные статистики и таблицы”).

Отметив галочками необходимые критерии, следует активировать процедуры нажатием кнопки “Критерии нормальности” (“Tests for normality”).

В этом случае результаты анализа (рассчитанные значения  $p$ ) могут быть выданы сразу для всех анализируемых признаков в одном окне (рис. 7.5), что более удобно.

**Интерпретация результатов.** Напомним основные свойства нормального распределения признака:

- нормальное распределение признака симметрично относительно своего среднего значения;

Перемен.	N	Вес.	p
VAR2	26	.880514	.005903
VAR3	26	.728023	.000013
VAR4	30	.394670	.000000
VAR5	29	.342907	.000000
VAR6	29	.218990	.000000
VAR7	29	.353381	.000000
VAR8	29	.337202	.000000

Рис. 7.5. Окно результатов применения критериев нормальности подмодуля “Таблицы частот”.

- среднее значение, мода и медиана распределения признака совпадают;
  - приблизительно 68% значений признака (объектов исследования) находятся в интервале  $M \pm s$ ,  
95% — в интервале  $M \pm 2s$ ,  
99% — в интервале  $M \pm 3s$ ,
- где  $M$  — среднее значение,  $s$  — среднее квадратическое отклонение (СКО).

Критерии нормальности проверяют следующую нулевую гипотезу: распределение изучаемого признака не отличается от нормального распределения. Альтернативная гипотеза — распределение изучаемого признака отличается от нормального.

Если полученное значение  $p$  для данного статистического критерия (теста) больше критического (принятого исследователем) уровня статистической значимости (например, 0,05), то нулевую гипотезу не отклоняют, т.е. распределение исследуемого признака приближенно можно считать нормальным. Из этого следуют два вывода:

- 1) для описания распределения этого признака можно использовать такие параметры, как среднее значение и СКО;
- 2) при сравнении групп по этому признаку мы вправе пользоваться параметрическими методами. В этом случае также можно использовать и непараметрические методы, однако в связи с

тем что они обладают несколько меньшей статистической мощностью (чувствительностью), т.е. реже, чем параметрические методы, обнаруживают существующие различия выборок, предпочтительнее применять параметрические методы.

Если при использовании критериев нормальности  $p < 0,05$  (или другой принятой критической величины), то следует отклонить нулевую гипотезу и принять альтернативную, т.е. распределение признака считать отличающимся от нормального. В этом случае сравнение групп по этому признаку должно проводиться с использованием исключительно непараметрических методов.

**Замечание 1.** Иногда критерии Лиллиефорса и Шапиро—Уилка дают разные результаты. В этом случае мы рекомендуем опираться на результат теста Шапиро—Уилка как на более надежный.

**Замечание 2.** Если гипотеза о нормальном распределении отклонена, но известно, что выборка содержит аномальные, или выпадающие, значения признака (“выбросы”), то можно предполагать, что после их удаления повторное применение теста на нормальность даст другой результат. Подробнее о работе с выпадающими значениями признака см. раздел 7.3.5.

**Внимание!** Если необходимо представить описание центральных тенденций и дисперсий признака не только в целом по всей группе, но и в подгруппах, а также в дальнейшем проводить сопоставление подгрупп, то необходимо анализировать вид распределения признака в каждой из этих подгрупп (предварительно указав условие отбора строк либо выделив их в отдельные файлы из исходного файла). Описание процедуры разбиения файла на несколько файлов (подгрупп) см. в разделе 4.2.

## 7.2. Описание количественных данных в зависимости от вида их распределения

При представлении данных в научной статье или отчете обычно следует обобщить наблюдения (объекты исследования) выборки, т.е. привести описательную статистику (параметры распределения) для количественных признаков. Это необходимо для того, чтобы читатель мог составить представление о том, каковы ос-

новые характеристики выборки — центральная тенденция, рассеяние по каждому из изучаемых признаков. Без этого читатель не сможет оценить, в какой степени близка изучаемая выборка к его контингенту больных (т.е. насколько обобщаемы результаты исследования). При описании малых выборок (в которых число наблюдений, или объектов исследования, не больше 20) рекомендуется приводить таблицу с исходными данными, так как для таких выборок методы статистики, в том числе описательной, не всегда адекватны.

При описании больших выборок требуется привести таблицу, в которой описаны центральные тенденции и рассеяние значений количественных признаков в данной выборке, или соответствующий рисунок.

**Меры центральной тенденции (меры локализации, меры положения)** показывают наиболее типичное значение для данной выборки. К мерам центральной тенденции относятся:

- среднее значение ( $M$ ) — среднее арифметическое;
- медиана ( $Me$ ) — значение, справа и слева от которого на оси значений признака располагаются равные количества значений признака данной выборки;
- мода ( $Mo$ ) — наиболее часто встречающееся значение признака в выборке;
- среднее геометрическое значение — антилогарифм среднего арифметического для логарифмированных данных.

**Меры рассеяния** показывают разброс значений признака в выборке. К мерам рассеяния относятся:

- размах — разность максимального и минимального значений признака;
- интерпроцентильный размах (интервал) — значения каких-либо процентилей распределения, например 10-го и 90-го;
- интерквартильный размах (интервал) — значения 25-го и 75-го процентилей. Такой интервал независимо от вида распределения включает 50% значений признака в выборке;
- среднее квадратическое (стандартное) отклонение (СКО,  $s$ ,  $SD$ ). Показывает разброс данных по интервалу значений признака относительно среднего значения.

Центральные тенденции и рассеяния количественных признаков, имеющих приблизительно нормальное распределение (и только таких признаков!), следует описывать средним значением ( $M$ ) и средним квадратическим отклонением ( $s$ ) в формате  $M$

(s). Широко распространенное представление в виде  $M \pm s$  в настоящее время использовать не рекомендуется.

Центральные тенденции и дисперсии количественных признаков, не имеющих приблизительно нормального распределения (а это подавляющее большинство — около 80%! — распределений медико-биологических признаков), следует описывать медианой и интерквартильным размахом (25-й и 75-й процентиля) или другим интерпроцентильным размахом (например, 80% интерпроцентильный размах — между 10-й и 90-й процентилями).

*Пример:* “Медиана 240 ммоль/л (интерквартильный размах от 200 до 290 ммоль/л)” или “уровень холестерина (медиана и интерквартильный размах) составил 240 ммоль/л (от 200 до 290 ммоль/л)” или “медиана 240 ммоль/л (25-й процентиль = 200 ммоль/л, 75-й процентиль = 290 ммоль/л)”.

## STATISTICA:

- Модуль “Основные статистики и таблицы”  
 (“Basic statistics / Tables”) (см. рис. 7.1)
- Подмодуль “Описательные статистики”  
 (“Descriptive statistics”) (см. рис. 7.2)
- Кнопка “Другие статистики”  
 (“More statistics”)
- Диалоговое окно для выбора  
 необходимых статистик (рис. 7.6)

Рекомендуется выполнить расчет следующих параметров, отметив (галочками) соответствующие опции:

- “Число наблюдений N” (“Valid N”)
- “Среднее” (“Mean”)
- “Медиана” (“Median”)
- “Стандартное отклонение” (“Standard Deviation”)
- “Стандартная ошибка среднего” (“Standard error of mean”)
- “95% доверительные границы для среднего” (“95% confidence limits of mean”)
- “Минимум и максимум” (“Minimum and maximum”)
- “Нижние и верхние квартили” (“Lower and upper quartiles”)
- “Размах” (“Range”)

Для описания выборочного нормального распределения количественных признаков необходимо указывать как минимум следующие параметры:

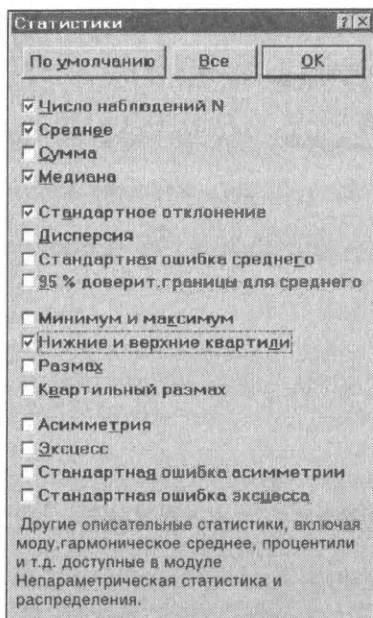


Рис. 7.6. Диалоговое окно “Статистики” (подмодуль “Описательные статистики” в модуле “Основные статистики и таблицы”).

- число наблюдений (объектов исследования);
- среднее значение признака;
- среднее квадратическое отклонение.

Для описания выборочного распределения количественных признаков, которое отличается от нормального, рекомендуется указывать:

- число объектов исследования;
- медиану;
- верхний и нижний квартили.

Выявление **максимального и минимального значений признака** необходимо для контроля за соответствием данных допустимым значениям признака. Эти значения позволяют обнаружить:

- 1) ошибки, возникшие при вводе данных (естественно, ошибки необходимо исправить, а затем выполнить расчеты заново);
- 2) выпадающие значения. Их обработка описана в разделе 7.3.5.

**Размах** представляет собой разность максимального и минимального значений признака. Этот параметр бывает необходим в том случае, если нужно принять решение о возможности рассматривать дискретный признак (число возможных значений которого достаточно велико) приближенно непрерывным. Так, если размах превышает 20, то допустимо приближенно считать такой признак непрерывным. Например, число эритроцитов является по физическому смыслу дискретным признаком, так как при измерениях их числа дробные значения не фиксируются и даже не фиксируются значения с точностью менее миллиона. Однако поскольку число возможных значений достаточно велико, то такой признак вполне обоснованно можно анализировать как непрерывный.

**Замечание.** Напомним, что параметры (статистики) вычисляются для всей выборки, которая в данный момент представлена в файле, если не заданы условия отбора строк (см. раздел 4.1).

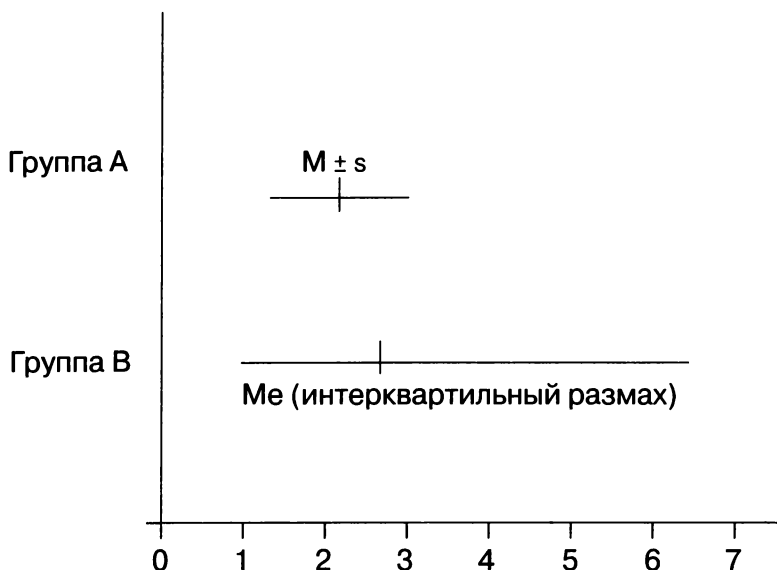
Результаты, полученные с помощью методов описательной статистики, удобно представлять в виде рисунков — диаграмм размахов и диапазонов. Они удобны для представления как нормальных, так и других распределений признаков. Приведем примеры подобного представления.

На рис. 7.7 показана диаграмма размахов в виде линейного графика. Нормальные распределения переменных (признаков), таких как в группе А, удобно представлять в виде отрезка  $[M - s; M + s]$ . Распределения как нормальные, так и отличающиеся от нормальных (группа В), можно представлять в виде интерквартильного интервала между 25-м и 75-м процентиллями, включающим 50% значений признака в выборке.

На рис. 7.8 показан другой способ представления распределений количественных данных — диаграмма диапазонов (так называемый коробочный график<sup>1</sup>) — для нормально (группа А) и ненормально (группа В) распределенных переменных (признаков). Такой график наглядно демонстрирует сразу несколько параметров распределения: центральные тенденции (среднее значение, медиана) и характеристики рассеяния объектов исследования (минимальное и максимальное значение признака, 25-й и 75-й процентиль, 10-й и 90-й процентиль).

<sup>1</sup> Box-and-whisker (англ.).





**Рис. 7.7.** Представление распределений количественных данных с использованием диаграмм размахов (линейный график) при нормальном распределении (группа А) и распределении, отличном от нормального (группа В).

Параметрами, представляющими наибольший интерес и описание которых должно сопровождаться указанием ДИ, обычно являются:

- среднее значение признака (в случае нормального распределения признака);
- медиана (в случае распределения признака, отличного от нормального; см. раздел 8.1.2);
- относительная частота (пропорция; см. раздел 9.3).

Чтобы вычислить границы ДИ для среднего значения признака в случае нормального его распределения, используют значение **стандартной ошибки среднего** значения ( $m$ )<sup>1</sup>. В отличие от СКО стандартная ошибка среднего не является характеристикой, описывающей рассеяние наблюдений (объектов исследования)

<sup>1</sup> Standard error of mean (SEM; англ.).

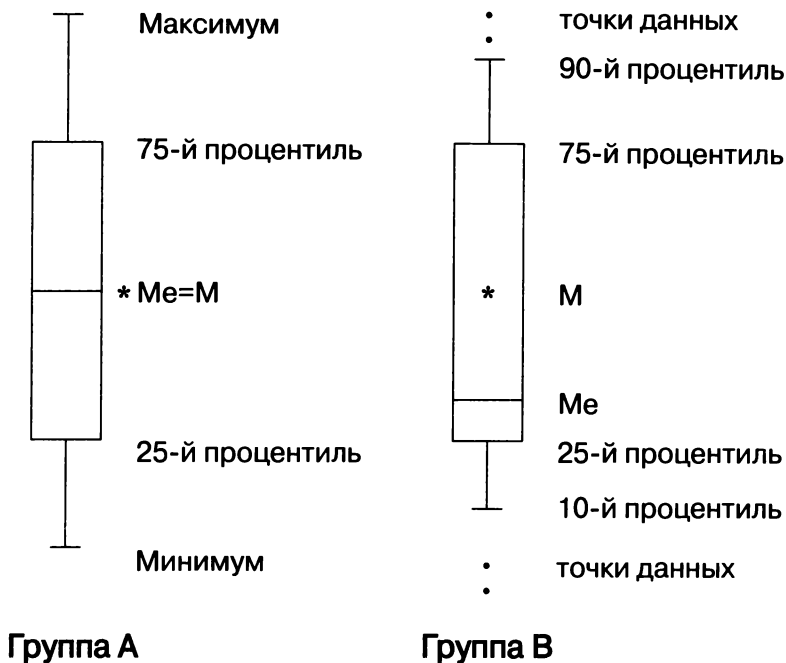


Рис. 7.8. Представление распределений количественных данных с использованием диаграмм диапазонов (“коробочный” график) при нормальном распределении (группа А) и распределении, отличном от нормального (группа В).

выборки по области значений, т.е. описательной статистикой, а представляет собой меру точности оценки среднего значения в генеральной совокупности на основании данных исследуемой выборки.

Интерпретация ДИ для среднего значения основывается на следующих предположениях:

- выборка является случайной или хотя бы репрезентативной;
- в генеральной совокупности анализируемый признак имеет приближенно нормальное распределение;
- объекты исследования независимы, т.е. отобраны из генеральной совокупности по одному, независимо от ранее отобранных объектов исследования.

ДИ рекомендуется представлять в виде указания нижней и верхней его границ. В общем виде границы ДИ для среднего вычисляют по следующим формулам:

$M - t \times m$  — нижняя граница;  $M + t \times m$  — верхняя граница, где  $t$  — значение критерия Стьюдента, соответствующее выбранному уровню статистической значимости и числу степеней свободы (число степеней свободы равно  $n-1$ , где  $n$  — число объектов исследования в выборке). Это значение может быть найдено с помощью программы “Probability calculator” (“Вероятностный калькулятор”<sup>1</sup>).

### STATISTICA:

→ Модуль “Основные статистики и таблицы”  
 (“Basic statistics/Tables”)

→ Подмодуль “Вероятностный калькулятор”  
 (“Probability calculator”)

В диалоговом окне (рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “ $t$  Стьюдента” (“Student  $t$  test”)
- выбрать опции “Обратная ф.р.”<sup>2</sup>, “Двусторонняя”<sup>3</sup> (“two-sided”) и “1— ф.р.”
- задать число степеней свободы “Ст. св.” (“ $df$ ”<sup>4</sup>)
- задать значение уровня значимости  $p$  (например, 0,05 для вычисления границ 95% ДИ)
- нажать кнопку “Вычислить” (“Compute”)

Искомое значение  $t$  появится в окне “ $t$ ”.

В частности, ДИ для среднего значения признака в выборке (в случае нормального распределения) из 100 наблюдений (объектов исследования) может быть вычислен следующим образом:

- 99% ДИ — интервал между  $(M-2,63m)$  и  $(M+2,63m)$ ;
- 95% ДИ — интервал между  $(M-1,98m)$  и  $(M+1,98m)$ ;
- 90% ДИ — интервал между  $(M-1,66m)$  и  $(M+1,66m)$ .

Формат представления ДИ в виде  $(M \pm 1,98m)$  допускается использовать только в таблицах (с целью экономии места) и только в том случае, если ДИ симметричен. Однако такой формат пред-

<sup>1</sup> Калькулятор для расчета вероятностей (Примеч. ред.).

<sup>2</sup> Функция распределения (Примеч. ред.).

<sup>3</sup> Имеется в виду двусторонняя функция распределения (Примеч. ред.).

<sup>4</sup> Degrees of freedom (англ.).

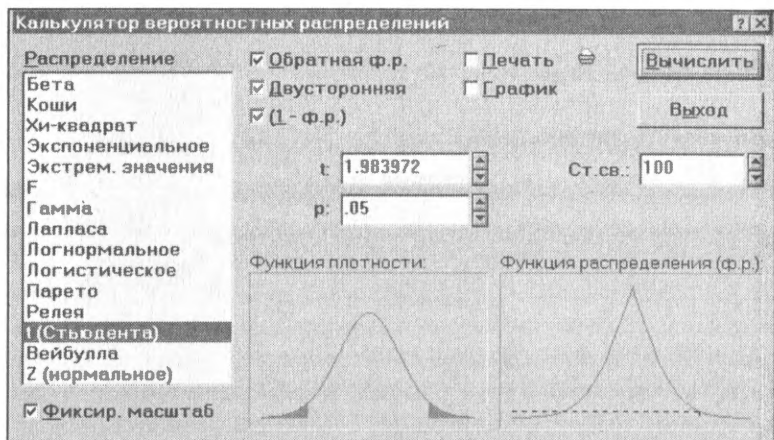


Рис. 7.9. Диалоговое окно подмодуля “Вероятностный калькулятор” для вычисления значения  $t$ -критерия по числу степеней свободы и уровню значимости.

ставления ДИ вообще говоря не рекомендуется использовать, так как в этом формате традиционно принято представлять описательные статистики  $M \pm s$ .

**Внимание!** Редакциями зарубежных научных изданий в настоящее время признается некорректным описание рассеяния признака в выборке с использованием стандартной ошибки среднего значения “ $m$ ”. Часто именно стандартную ошибку среднего “ $m$ ” ошибочно приводят в роли описательной статистики (обычно в формате “ $M \pm m$ ”), пытаясь продемонстрировать тем самым малую вариабельность своих данных, так как всегда (по определению)  $m < s$ . Другим частым поводом для использования “ $m$ ” вместо “ $s$ ” является также то, что исследователи сталкиваются с ситуацией, когда  $s$  превышает среднее значение  $M$  (так происходит в случае негауссова распределения), и соответственно запись “ $M \pm s$ ” в большинстве случаев (когда признак принимает только положительные значения) оказывается бессмысленной. Не зная, каким образом описывать распределения, отличные от нормального, авторы приводят запись “ $M \pm m$ ”, так как всегда  $m < s$ , и это позволяет избежать заведомо некорректной ситуации.

## **7.3. Некоторые частные аспекты представления количественных данных**

### **7.3.1. Точность представления описательных статистик количественных данных**

Обычно результаты вычислений описательных статистик содержат большее число значащих цифр, чем это было в исходных данных. В этом случае числовые данные должны округляться на основании правила арифметического округления<sup>1</sup>. Принято приводить оценки параметров (например, среднего значения и СКО) с той же точностью, с которой были представлены исходные данные. Например, если артериальное давление измерялось с точностью до разряда единиц, то следует приводить параметры, например, не в виде  $145,36 \pm 27,427$  мм рт. ст., а в виде  $145 \pm 27$ .

### **7.3.2. Данные связанных групп**

Данные связанных групп — это данные, полученные от одного участника исследования (до и после лечения, справа и слева и т.п.) или от разных участников, но подобранных в пары по определенным характеристикам. Такие данные могут быть представлены на графиках (например, зависимость значений изучаемого признака от времени) или в таблицах. Зачастую средние значения изучаемого признака в группах таких парных наблюдений (объектов исследования) не различаются. Поэтому в случае исследования динамических изменений количественных данных часто возникает необходимость оценить их по отношению к исходному уровню. При этом интервальные количественные данные преобразуются в относительные количественные данные. Для такого преобразования традиционно рекомендуется пользоваться следующей формулой:

---

<sup>1</sup> **Правило арифметического округления чисел:** чтобы округлить число до  $n$  значащих цифр, необходимо отбросить все цифры, стоящие после  $n$ -го разряда. При этом, если за последней сохраняемой цифрой следует цифра 0, 1, 2, 3 или 4, при округлении сохраняются все цифры до последней сохраняемой включительно (*округление с недостатком*). Если за последней сохраняемой цифрой следуют 9, 8, 7, 6 или 5, то к последней сохраняемой цифре прибавляется единица (*округление с избытком*).

$$[(A_1 - A_0) / A_0] \times 100\%,$$

где  $A_0$  — исходное значение признака,  $A_1$  — последующее его значение.

Если результат такого преобразования является отрицательным числом, то знак минус опускают и говорят об уменьшении значений параметра. Если результат является положительным числом, то говорят об увеличении значения параметра. В таблице можно привести значения со знаками минус или плюс.

О сравнении связанных групп по количественным признакам см. разделы 8.3 и 8.5.

### 7.3.3. Преобразование количественных данных

В ряде случаев возникает необходимость в преобразовании полученных данных, например:

- если изучаемый признак распределен не в соответствии с законом нормального распределения, а для использования определенного статистического метода необходимо именно нормальное распределение признака. В этом случае целью преобразования является приведение распределения к нормальному;
- если необходимо обеспечить независимость дисперсии от среднего значения, поскольку для использования ряда методов требуется обеспечить такую независимость;
- если необходимо оптимизировать множественную регрессионную модель.

В этих случаях могут быть применены следующие методы преобразования данных:

- логарифмирование;
- извлечение квадратного корня;
- обратное преобразование и т.д.

В результате некоторых преобразований распределение полученных данных может превратиться из исходно ненормального в нормальное, что дает основание для использования параметрических статистических методов.

Однако следует помнить, что при преобразовании данных преобразовываются и единицы их измерения, теряя при этом свой физический смысл. Кроме того, результаты анализа преобразованных данных бывает трудно интерпретировать.

### 7.3.4. Описание данных, полученных в малых выборках

Малыми условно принято называть такие выборки, в которых число наблюдений (объектов исследования) меньше 20. Эта граница условна. Ограничение основано на том, что при использовании методов описательной статистики для описания небольшого числа наблюдений (объектов исследования) можно получить искажение реальности, т.е. ввести в заблуждение исследователей, авторов, читателей. Например, имеющихся данных может быть недостаточно для того, чтобы проанализировать вид распределения изучаемого признака. В то же время необходимо заметить, что среднее значение и СКО формально можно рассчитать и из двух наблюдений (объектов исследования). При описании малых выборок принято в текстах статей и диссертаций представлять данные в исходном виде, а методы описательной статистики использовать лишь как вспомогательный способ описания выборок.

### 7.3.5. Анализ выпадающих данных

Выпадающие значения (“выбросы”) — крайние значения признаков, не характерные для данной выборки, слишком большие или слишком малые значения. Они могут исказить распределение исходных данных и оказывать непропорционально большой эффект на результаты всех типов статистического анализа.

Причины возникновения выпадающих значений могут быть следующими:

- ошибка при получении данных (артефакт);
- ошибка при подготовке данных (опечатка);
- аномальное значение признака.

Первые две ситуации могут быть обнаружены на этапе получения описательных статистик (например, при вычислении минимальных и максимальных значений признака).

В последнем же случае от исследователя требуется особое внимание.

Бытует мнение (например, [17, с. 42]), что на основании “правила трех сигм”<sup>1</sup> можно исключать из анализа наблюдение

---

<sup>1</sup> **Правило трех сигм:** при нормальном распределении признака 99,7% наблюдений (объектов исследования) располагаются внутри интервала с границами  $M \pm 3s$  ( $s$  — это СКО, оцененное по выборке). Для описания признаков, относящихся к генеральной совокупности, в математической статистике используются буквы греческого алфавита, а для описания признаков, относящихся к выборке, — латинские буквы.

(объект исследования), если значение признака не укладывается в интервал  $M \pm 3s$  (причем  $M$  и  $s$  рассчитываются без учета резко отклоняющегося значения признака), и в дальнейшем анализировать данные без него.

Однако в настоящее время предпочтительным считается другой подход. Тактика работы с выпадающими значениями следующая. Данные необходимо анализировать два раза — вместе с этим выпадающим значением, а затем без него. После этого необходимо сравнить результаты, и если они устойчивы к выпадающему значению (т.е. результаты различаются незначительно), то взять первый результат. Если же они различаются, необходимо привести и прокомментировать и тот, и другой.

В любом случае факт исключения выпадающих данных должен упоминаться при описании и обсуждении полученных результатов, а также должно быть приведено обоснование такого исключения.



## Глава 8. Сравнение групп по количественному признаку

К количественным признакам относятся непрерывные и дискретные (см. главу 2). Дискретные признаки занимают условно промежуточное положение между непрерывными и качественными порядковыми. Методы анализа непрерывных признаков могут применяться и для анализа дискретных признаков при условии, что число возможных значений дискретного признака достаточно велико.

В настоящее время в зарубежных медицинских исследованиях используются два подхода к сравнению двух групп по количественному признаку:

- 1) с использованием доверительных интервалов (ДИ);
- 2) путем проверки статистических гипотез.

Эти два подхода различаются по вопросам, на которые с их помощью могут быть получены ответы. Сравнение с помощью ДИ отвечает на вопрос “Насколько велики различия генеральных совокупностей (популяций)?”, а сравнение с помощью статистической проверки гипотез — на вопрос “В какой степени можно быть уверенным, что различия между генеральными совокупностями (популяциями) действительно существуют?” Перечисленные подходы основаны на одних и тех же допущениях и статистических принципах, поэтому они лишь дополняют друг друга. Вообще говоря, в настоящее время при описании результатов исследования рекомендуется представлять результаты применения обоих подходов.

Выбор подходящего метода сравнения определяется несколькими факторами:

- число сопоставляемых групп;
- зависимостью или независимостью выборок (групп);
- видом распределения признаков.

Как уже упоминалось выше, группы являются независимыми (несвязанными), если набор объектов исследования (участни-

ков) в каждую из групп осуществлялся независимо от того, какие объекты исследования (участники) включены в другую группу. Так, в частности, происходит при рандомизации, когда распределение участников исследования по группам происходит случайным образом.

Группы являются зависимыми (связанными), например, в следующих случаях:

- в исследовании случай—контроль, когда набор в группы осуществляется с использованием подбора пар;
- в динамических исследованиях, когда изучаются одни и те же объекты в разные моменты времени.

От вида распределения и типа исследуемого признака зависит выбор между двумя классами статистических методов — параметрическими и непараметрическими методами.

Параметрические методы применимы к количественным признакам, имеющим нормальное распределение.

Непараметрические методы применяются к количественным признакам независимо от вида их распределения (в том числе и к нормально распределенным признакам).

В данной главе рассматривается решение следующих задач:

- сравнение одной группы с популяцией (см. раздел 8.1):
  - в случае нормально распределенного признака;
  - в случае любого распределения признака;
- сравнение двух независимых (несвязанных) групп (см. раздел 8.2):
  - ДИ для разности средних;
  - параметрический метод;
  - непараметрические методы;
- сравнение двух зависимых (связанных) групп (см. раздел 8.3):
  - ДИ для средней разности;
  - параметрический метод;
  - непараметрические методы;
- сравнение трех независимых (несвязанных) групп и более (см. раздел 8.4):
  - параметрический дисперсионный анализ;
  - непараметрические методы;
- сравнение трех зависимых (связанных) групп и более (см. раздел 8.5):
  - непараметрический дисперсионный анализ.

## 8.1. Сравнение одной группы с популяцией

**Задача.** Известно среднее значение изучаемого признака в популяции (популяционное среднее значение), необходимо сопоставить с ним среднее значение изучаемой группы.

**Пример:** обследованы дети в возрасте 10 лет; средний рост (по данным популяционных исследований) для детей такого возраста составляет 140 см. Необходимо выяснить, не отстают ли в росте дети исследуемой группы.

**Решение** зависит от того, каково распределение изучаемого количественного признака:

- нормальное (см. раздел 8.1.1);
- отличное от нормального или неизвестное (см. раздел 8.1.2).

### 8.1.1. Случай нормально распределенного признака

**Способ I — оценка с помощью 95% доверительного интервала для среднего значения.** Если изучаемый признак распределен в изучаемой группе в соответствии с законом нормального распределения, то границы 95% ДИ для среднего определяются как  $[M - t \times m; M + t \times m]$ , где  $M$  — среднее значение,  $m$  — стандартная ошибка среднего,  $t$  — значение  $t$ -критерия для соответствующего числа степеней свободы и выбранного уровня статистической значимости. Число степеней свободы  $df = n - 1$ , где  $n$  — число объектов исследования в группе.

Значение  $t$ -критерия для конкретного случая можно узнать, воспользовавшись опцией “Вероятностный калькулятор”.

#### STATISTICA:

- Модуль “Основные статистики и таблицы”  
 (“Basic statistics/Tables”)
- Подмодуль “Вероятностный калькулятор”  
 (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “ $t$  Стьюдента” (“Student  $t$ -test”)
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1 — ф.р.”

- задать число степеней свободы “Ст. св.” (“ $df$ ”)
- задать значение  $p$  (например, 0,05 для вычисления границ 95% ДИ)
- нажать кнопку “Вычислить” (“Compute”)

Искомое значение  $t$  появится в окне “ $t$ ”.

**Интерпретация результатов.** Если рассчитанный ДИ не включает популяционное среднее, то с определенной долей уверенности (соответствующей доверительному коэффициенту, например 95%) можно считать, что выборка статистически значимо отличается от генеральной совокупности (популяции).

**Представление результатов.** Привести следующую информацию:

- популяционное значение;
- среднее значение, рассчитанное по выборке;
- границы 95% ДИ.

**Способ II — проверка гипотез.** Можно применить  $t$ -критерий Стьюдента, если изучаемый признак имеет нормальное распределение в выборке. При этом проверяется нулевая гипотеза об отсутствии различий средних. Альтернативная гипотеза — существование различий средних.

## STATISTICA:

- Модуль “Основные статистики и таблицы” (“Basic statistics/Tables”)
- Подмодуль “Другие критерии значимости” (“Other significance tests”) в версиях ППП STATISTICA 5.\* (“Probability calculator” в версиях 4.\*.)
  - Процедура “Различие между двумя средними (нормальное распределение)” (“Difference between two means (normal distribution)”) (рис. 8.1)

Включить режим “Среднее выборки 1 по сравнению со средним популяции 2” (“Single mean 1 vs. population mean 2”).

Затем нужно внести значения среднего, СКО и число наблюдений (объектов исследования) в соответствующие окошки, а после этого нажать кнопку “Вычислить” (“Compute”). Будет рассчитано значение  $p$ .

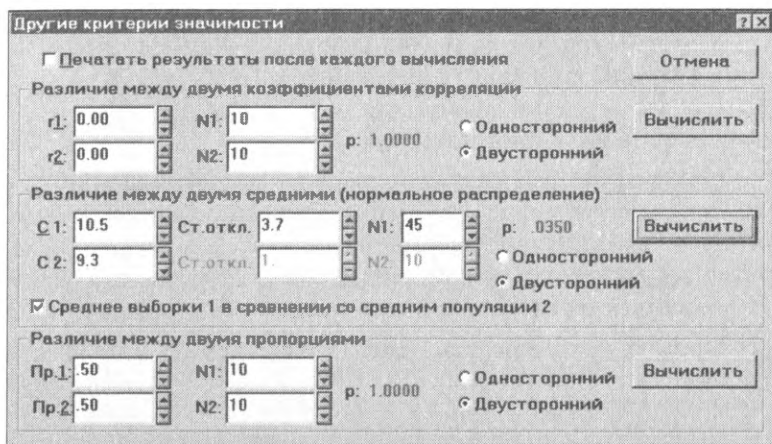


Рис. 8.1. Диалоговое окно подмодуля “Другие критерии значимости” (“Other significance tests”).

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если вычисленное значение  $p > 0,05$ , то нулевая гипотеза об отсутствии различий средних не отклоняется.
- Если  $p < 0,05$ , то следует отклонить нулевую гипотезу и принять альтернативную гипотезу о том, что различия выборочного среднего и популяционного среднего значения существуют.

**Представление результатов.** Привести следующую информацию:

- среднее значение и СКО;
- название критерия;
- точное значение  $p$ .

### 8.1.2. Случай любого распределения признака

К сожалению, в ППП *STATISTICA* пока не реализованы способы для решения такой задачи, и мы приводим описание  $\chi^2$  возможных подходов в общем виде.

**Способ I — вычисление 95% доверительного интервала для медианы.** Если признак имеет распределение, отличающееся от нормального, или просто неизвестно, каково его распреде-

ление, то можно воспользоваться способом вычисления 95% ДИ для медианы. Для этого нужно выполнить следующее:

- 1) по таблице (см. Приложение 3) с учетом объема выборки и необходимой доверительной вероятности найти ранги<sup>1</sup>, значения которых соответствуют границам ДИ;
- 2) ранжировать значения изучаемого признака в порядке возрастания [это легко сделать с использованием опции “Ранжировать” (“Rank”) в меню “Переменные” (“Variables”)];
- 3) установить, которые из значений изучаемого признака соответствуют табличным рангам. Эти значения признака и являются границами ДИ.

**Представление результатов.** Привести следующую информацию:

- популяционное среднее или медиану;
- вычисленные медиану и границы 95% ДИ.

**Способ II — проверка статистической гипотезы об отсутствии различий.** Для этого необходимо использование критериев, алгоритмы которых, к сожалению, не содержатся в ППП *STATISTICA*, а формулы для расчетов довольно сложны, поэтому мы рекомендуем пользоваться I способом.

## 8.2. Сравнение двух независимых (несвязанных) групп

Статистический анализ не может ответить на вопрос “Случайны ли различия между группами?” Он может ответить на близкий вопрос: “Какова вероятность ошибки, если считать, что группы различаются, хотя на самом деле различие между ними отсутствует?” При сравнении групп с помощью проверки статистических гипотез вычисляется значение  $p$  — вероятность того, что новые случайные выборки из тех же популяций будут различаться в той же или большей степени, если нулевая гипотеза (об отсутствии различий между популяциями) верна.

Существуют три способа сравнения двух независимых групп по количественным признакам:

---

<sup>1</sup> Ранг — порядковый номер объекта исследования в общем списке объектов исследования, упорядоченном (ранжированном) по возрастанию значений изучаемого признака.

- вычисление ДИ для разности средних (см. раздел 8.2.1);
- параметрический метод проверки статистических гипотез —  $t$ -критерий Стьюдента для независимых выборок (см. раздел 8.2.2);
- непараметрические методы проверки статистических гипотез — критерии Манна—Уитни, Колмогорова—Смирнова, Вальда—Вольфовица (см. раздел 8.2.3).

### 8.2.1. Доверительный интервал для разности средних

**Задача.** Сравнить две независимые (несвязанные) группы по количественному признаку.

*Пример:* сравнить систолическое АД в группах, получавших разное гипотензивное лечение.

**Решение.** Сравнение групп с использованием ДИ получает в последние годы все большее распространение, так как оно обладает рядом преимуществ по сравнению с методом проверки статистических гипотез:

- 1) проще, так как не требуется использование статистических критериев;
- 2) показывает величину различий между группами;
- 3) легче интерпретировать результаты.

Ограничениями на применение этого метода является нарушение хотя бы одного из следующих условий:

- выборки случайны или хотя бы репрезентативны;
- выборки имеют равные СКО;
- объекты исследования в каждой из выборок независимы;
- признак непрерывен;
- признак имеет нормальное распределение в выборках (если число объектов исследования меньше 100).

К сожалению, программа для таких вычислений в законченном виде в ППП **STATISTICA** отсутствует, поэтому мы приводим формулы для расчета. В формулах могут быть использованы параметры выборочных распределений, которые можно получить с использованием модуля “Описательная статистика” (“Descriptive statistics”).

Сравнение выборок с использованием ДИ осуществляют следующим образом:

**1. Рассчитывают СКО для разности средних ( $m_\Delta$ ):**

- Если число объектов исследования в сравниваемых группах равно, то используют следующую формулу:

$$m_\Delta = \sqrt{m_a^2 + m_b^2} .$$

- Если число объектов исследования в группах не равно, то необходимо применять более сложные вычисления последовательно по двум формулам:

$$s_{ab} = \sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{n_a + n_b - 2}} ,$$

где  $s_a$  — СКО в группе А,  $s_b$  — СКО в группе В,  $s_{ab}$  — объединенное СКО.

$$m_\Delta = s_{ab} \sqrt{\frac{1}{n_a} + \frac{1}{n_b}} .$$

**2. Вычисляются границы ДИ для разности между средними значениями признака  $\Delta = M_a - M_b$  в двух группах:**

$$(\Delta - t \times m_\Delta),$$

где  $t$  — значение  $t$ -критерия для числа степеней свободы ( $n_a + n_b - 2$ ) и выбранного уровня статистической значимости. Его значения для конкретного случая можно узнать, воспользовавшись опцией “Вероятностный калькулятор” (“Probability calculator”).

**STATISTICA:**

- Модуль “Основные статистики и таблицы” (“Basic statistics/Tables”)
  - Подмодуль “Вероятностный калькулятор” (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “ $t$  Стьюдента” (“Student  $t$ -test”);
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1 - ф.р.”;



- задать число степеней свободы “Ст. св.” (“ $df$ ”);
- задать значение  $p$  (например, 0,05 для вычисления границ 95% ДИ);
- нажать кнопку “Вычислить” (“Compute”).

В окне “ $t$ ” появится искомое значение.

*Пример:* при наличии в двух выборках 31 объекта исследования (т.е. при числе степеней свободы 29) ДК 95% (т.е.  $p=0,05$ ) соответствует значение  $t=2,045$ , ДК 99% — значение  $t=2,756$ , ДК 90% —  $t=1,699$ .

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если ДИ для разности средних включает ноль, то следует считать, что различия между группами по анализируемому признаку отсутствуют.
- Если 95% ДИ не включает ноль, то следует считать, что различие между группами существует при уровне статистической значимости 0,05.
- Если ДИ включает как клинически значимые, так и клинически незначимые значения, то результаты недостаточно точны для того, чтобы сделать определенный вывод.

**Рекомендуемое представление результатов.** Привести следующую информацию:

- результаты вычисления параметров распределения, или описательные статистики (число наблюдений, или объектов исследования, среднее, СКО) для каждой группы;
- разность средних (точечная оценка);
- границы 95% ДИ для разности средних (интервальная оценка).

## 8.2.2. Параметрический метод ( $t$ -критерий Стьюдента для независимых групп)

**Задача.** Проверить, различаются ли средние значения изучаемого признака в исследуемых группах.

*Пример:* сравнить по возрасту две группы участников исследования.

**Решение.** Наиболее популярный метод решения такой задачи — использование  $t$ -критерия Стьюдента. Данный метод заключается в проверке нулевой гипотезы о том, что средние значения признака в сравниваемых группах не различаются. Если нулевая гипотеза по результатам теста отклоняется, то следует

принять альтернативную гипотезу о том, что средние значения в группах различны.

Однако применение этого метода в его классическом варианте правомерно лишь при одновременном соблюдении следующих двух условий:

- 1) значения признаков в каждой из сравниваемых групп должны иметь нормальное распределение (проверка распределения признака на соответствие закону нормального распределения описана в разделе 7.1);
- 2) дисперсии распределений признаков в двух сравниваемых группах равны.

В связи с вышеизложенным вычислению  $t$ -критерия должна предшествовать проверка гипотезы о том, что распределения признака в каждой из сравниваемых групп являются нормальными: модуль “Основные статистики и таблицы” (“Basic statistics/Tables”), подмодуль “Таблицы частот” (“Frequency tables”), раздел “Проверка на нормальность” (“Tests for normality”).

Выполнение второго условия (равенства дисперсий распределений признаков) может быть проверено с помощью критерия Левена в данном подмодуле либо с помощью  $F$ -критерия в процедуре дисперсионного анализа ANOVA, описанной в разделе 8.4.1.1.

Вместе с тем данный тест считается достаточно устойчивым к незначительным нарушениям своих допущений.

**Важное замечание.** Опыт показывает, что в медицинских исследованиях лишь около 20% количественных признаков удовлетворяют указанным условиям (по нашему опыту и по данным В.П. Леонова, П.В. Ижевского [18]). Таким образом, в большинстве случаев использование  $t$ -критерия неправомерно, так как приводит исследователя к неправильным результатам, что влечет за собой ложные выводы, чреватые разработкой в лучшем случае неэффективных, а в худшем — опасных методов диагностики и лечения. В статье [19] на примере исследования в области кардиологии показано, как драматически влияют нарушения требований правильного применения  $t$ -критерия на результаты исследования.

Если хотя бы одно из вышеуказанных условий не выполняется, то для проверки гипотез об отсутствии различий характеристик выборок необходимо пользоваться непараметрическими методами.

В ППП *STATISTICA* реализованы два варианта критерия Стьюдента для сравнения независимых выборок:

- 1 — для анализа исходных данных в таблице данных;
- 2 — для анализа с использованием ранее рассчитанных значений средних, СКО и числа объектов исследования при условии, что оба допущения *t*-критерия выполняются.

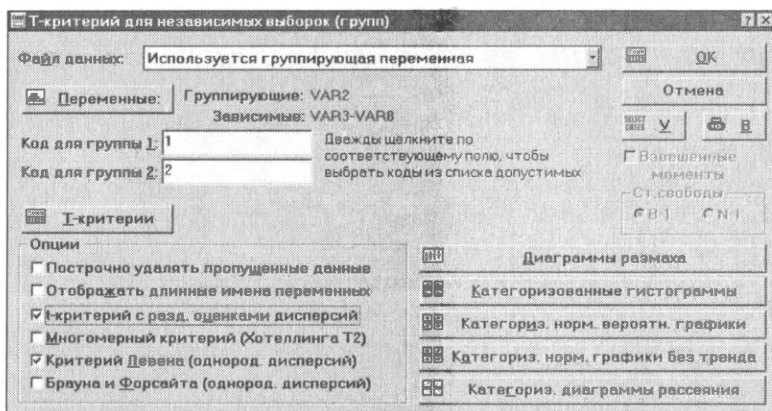
### Вариант 1

**STATISTICA:**

→ Модуль “Основные статистики и таблицы”

(“Basic statistics/Tables”)

→ Процедура “*t*-критерий для независимых выборок” (“*t*-test for independent samples”) (рис. 8.2)



**Рис. 8.2.** Диалоговое окно процедуры “*t*-критерий для независимых выборок” (в модуле “Основные статистики и таблицы”).

В диалоговом окне, активизировав кнопку “Переменные” (“Variables”), необходимо указать в левой части окна группирующий признак (столбец, содержащий коды групп), а в правой части окна — столбцы, содержащие анализируемые признаки.

Для проверки выполнения второго условия применимости критерия Стьюдента (условия равенства дисперсий распределений признаков) можно воспользоваться процедурой расчета критерия Левена, для чего необходимо активизировать опцию “Критерий Левена (однород. дисперсий)” (“Levene’s test”) в диалоговом окне (см. рис. 8.2).

Помимо классического критерия Стьюдента, ППП **STATISTICA** содержит и его модификацию, которая позволяет учесть возможное невыполнение второго условия — расчет  $t$ -критерия с учетом неравных дисперсий (“ $t$ -Критерий с разд. оценками дисперсий” (“ $t$ -Test with separate variance estimates”).

Рекомендуется сразу отметить опции “Критерий Левена” и “ $t$ -Критерий с отдельными оценками дисперсий” и произвести расчеты. Результаты выводятся в отдельное окно (рис. 8.3).

Перемен.	Среднее		t-знач.	Ст.св.	p	t отдел. оц.дисп.		p двухст.
	G 1:1	G 2:2				Ст.св.	Ст.св.	
DEC_P	18.27273	13.45745	2.504255	56	.015210	2.573943	15.53780	.020736
DEC_TA	17.59091	13.47872	2.255763	56	.027944	2.165944	14.40014	.047553

Рис. 8.3. Окно результатов сравнения двух выборок по  $t$ -критерию.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- При  $p < 0,05$  для критерия Левена следует сделать вывод о различии дисперсий распределений признаков в сравниваемых группах. В этом случае следует принимать во внимание только значения  $p$  для  $t$ -критерия с отдельными оценками дисперсий.
- При  $p > 0,05$  для критерия Левена следует сделать вывод о равенстве дисперсий в группах. В этом случае можно принимать во внимание любой из результатов — классического  $t$ -критерия или  $t$ -критерия для выборок с разными дисперсиями.
- Оценивается подсчитанное значение  $p$  для  $t$ -критерия (классического или модифицированного). Если  $p > 0,05$ , то нулевая гипотеза об отсутствии различий средних значений не отклоняется. Если  $p < 0,05$ , то следует нулевую гипотезу отклонить и принять альтернативную гипотезу о различии средних значений в сравниваемых группах.

**Вариант 2.** Если среднее и СКО были рассчитаны в ходе предыдущего анализа, а также известно, что выполняются условия применимости  $t$ -критерия, можно использовать процедуру в другом модуле ППП.

**Решение.** Проверка нулевой статистической гипотезы об отсутствии различия средних значений в группах. Альтернативная гипотеза — различие средних существует.

## STATISTICA:

- Модуль “Основные статистики и таблицы” (“Basic statistics/Tables”)
- Подмодуль “Другие критерии значимости” (или “Probability calculator” в версии STATISTICA 4.\*)
- Процедура “Различия между двумя средними (нормальное распределение)” (“Difference between two means [normal distribution]”)

Подставив соответствующие параметры для каждой из двух групп в диалоговое окно и выполнив расчет, мы получим значение  $p$ .

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза об отсутствии различий средних не отклоняется.
- Если  $p < 0,05$  (такие результаты выделяются красным цветом шрифта в таблице результатов), то следует отклонить нулевую гипотезу и принять альтернативную гипотезу о существовании различий средних значений с уровнем статистической значимости  $p$ .

**Замечание 1.** *t-Критерий Стьюдента — частный случай однофакторного дисперсионного анализа (ANOVA) для ситуации сравнения двух групп. При использовании однофакторного ANOVA и t-критерия для анализа одних и тех же данных получаются одинаковые результаты.*

**Замечание 2.** *Помните о проблеме множественных сравнений (см. раздел 15.1)! Если сопоставлений много, то рекомендуется установить более жесткий уровень статистической значимости — 0,01 или менее.*

**Представление результатов.** Привести следующую информацию:

- число объектов исследования в каждой из групп;
- средние и СКО изучаемого признака для каждой из групп;
- результаты применения критериев для оценки нормальности распределения и равенства дисперсий в случае, если используется классический критерий Стьюдента;

- результаты применения критерия для оценки нормальности распределения и указание на использование модифицированного критерия Стьюдента для групп с различными дисперсиями;
- точное значение  $p$ .

### 8.2.3. Непараметрические методы (критерии Манна—Уитни, Вальда—Вольфовица, Колмогорова—Смирнова)

**Задача.** Сопоставить две группы по одному или нескольким количественным признакам, имеющим хотя бы в одной из групп распределение, отличное от нормального или если вид распределений не анализировался.

*Пример:* сравнить две группы больных по реологическим свойствам крови в случае, если анализ видов распределений не проводился или проводился и показал, что распределения признаков отличаются от нормальных.

**Решение.** Для сопоставления двух групп по количественным признакам, которые не являются нормально распределенными или вид распределений которых не анализировался, существует несколько способов:

- с помощью  $U$ -критерия Манна—Уитни<sup>1</sup>. Необходимо указать группирующий признак, а также один или несколько признаков, по которым проводится сопоставление групп. Условия применимости метода: анализируемые признаки должны быть количественными или порядковыми. Проверяется гипотеза о равенстве средних рангов (в отличие от  $t$ -критерия, где проверяется гипотеза о равенстве средних значений);
- с помощью критерия серий Вальда—Вольфовица<sup>2</sup>. Этот критерий применяется для проверки гипотезы о том, что исследуемые группы получены из одной и той же генеральной совокупности. Условия применимости метода те же: анализируемые признаки должны быть количественными или порядковыми;
- с помощью двухвыборочного критерия Колмогорова—Смирнова<sup>3</sup>. Критерий применяется для проверки гипотезы о том,

<sup>1</sup> Mann—Whitney  $U$ -test (англ.).

<sup>2</sup> Wald—Wolfowitz runs test (англ.).

<sup>3</sup> Kolmogorov—Smirnov two-sample test (англ.).

что исследуемые группы получены из одной и той же генеральной совокупности. Условия применимости метода — те же.

С помощью каждого из этих методов проверяется нулевая статистическая гипотеза об отсутствии различий групп. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий групп.

### STATISTICA:

→ Модуль “Непараметрические статистики”  
 (“Nonparametrics/Distrib.”)

→ Процедуры:

- “Критерий серий Вальда—Вольфовица”  
 (“Wald—Wolfowitz runs test”);
- “U-критерий Манна—Уитни”  
 (“Mann—Whitney U-test”) (рис. 8.4);
- “Двухвыборочный критерий  
 Колмогорова—Смирнова”  
 (“Kolmogorov—Smirnov two-sample test”)

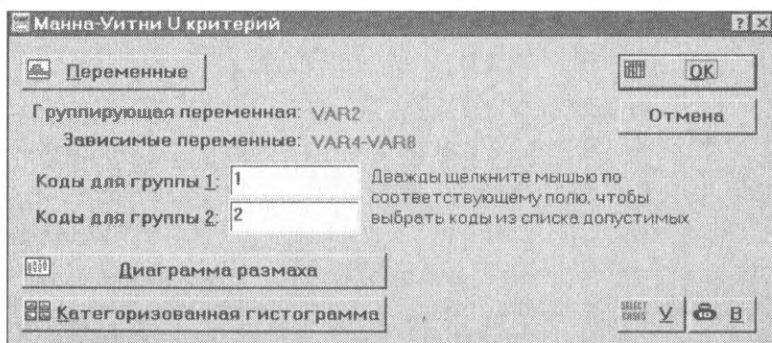


Рис. 8.4. Диалоговое окно процедуры “Манна—Уитни U-критерий” (в модуле “Непараметрические статистики”).

В ходе анализа будут рассчитаны различные параметры и выведены в окно результатов (рис. 8.5). В этом окне следует обратить внимание на столбец “ $p$ -уров.”, в который выводятся рассчитанные значения  $p$ .

При применении критериев Вальда—Вольфовица и Колмогорова—Смирнова результаты выглядят аналогично.

Перем.	Сум. ранг Группа 1	Сум. ранг Группа 2	U	Z	p-уров.	Z скорр.	p-уров.
INBUR 1	512.5000	813.5000	187.5000	-2.59081	.009579	-2.60768	.009120
ATROP 1	562.5000	763.5000	237.5000	-1.64869	.099220	-1.67803	.093351
BLESK 1	555.0000	720.0000	255.0000	-1.10685	.268368	-1.12484	.260664
MOBIL 1	143.5000	321.5000	88.5000	-.50593	.612907	-.50650	.612512

Рис. 8.5. Окно результатов процедуры “U-критерий Манна—Уитни” (в модуле “Непараметрические статистики”).

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза об отсутствии различий групп не отклоняется.
- Если  $p < 0,05$ , то следует отклонить нулевую гипотезу и принять альтернативную гипотезу о существовании различий групп с уровнем статистической значимости  $p$ .

**Замечание.** Помните о проблеме множественных сравнений (см. раздел 15.1)! Если сопоставлений много, то рекомендуется установить более жесткий уровень статистической значимости — 0,01 или менее.

**Представление результатов.** Привести следующую информацию:

- медианы и ДИ для каждой из групп;
- название статистического критерия;
- точное значение  $p$ .

### 8.3. Сравнение двух зависимых (связанных) групп

Ниже описаны три подхода к сопоставлению двух зависимых (связанных) групп:

- с использованием ДИ для средней разности значений признака (для нормальных распределений признаков);
- параметрический метод —  $t$ -критерий для зависимых выборок (в случае нормальных распределений признаков);
- непараметрические методы — вычисление критерия Вилкоксона и критерия знаков (для выборок с любыми распределениями признаков).



В процедурах анализа зависимых групп отбрасываются неполные наблюдения (т.е. объекты, или участники, исследования с пропущенными значениями, например выбывшие участники исследования, для которых отсутствует информация о конечных значениях признаков), следовательно, количества объектов исследования в сопоставляемых столбцах оказываются равными.

### **8.3.1. Доверительный интервал для средней разности**

Задача определения ДИ для средней разности возникает в случаях, когда значение исследуемого признака у участника исследования из первой группы может быть ближе к значению признака у определенного участника исследования из второй группы, чем к значению случайно выбранного участника исследования из второй группы. Такие ситуации возникают в следующих случаях:

- при обследовании участников исследования до и после изучаемого воздействия (вмешательства);
- при изучении групп, сформированных из подобранных пар участников исследования;
- при обследовании близнецов или детей с родителями;
- в лабораторных исследованиях с параллельным контролем.

При анализе двух зависимых выборок могут быть вычислены разности значений признака для каждой пары наблюдений (например, разность значений признака до и после исследования), а затем для группы в целом вычислена средняя разность.

#### **Условия применимости метода:**

- пары участников исследования должны быть случайно выбраны из генеральной совокупности или по крайней мере репрезентативны;
- разности значений изучаемого признака для пары участников исследования в генеральной совокупности должны быть распределены нормально. Чем больше выборка, тем менее существенно это ограничение.

К сожалению, программа для вычисления ДИ для средней разности значений признака в законченном виде в ППП **STATISTICA** отсутствует, поэтому мы приводим формулы для расчета. В формулах могут использоваться параметры выборочных распределений, значения которых можно оценить с использованием модуля “Описательная статистика”.

Итак, чтобы вычислить ДИ для средней разности значений парных признаков, можно действовать следующим образом:

1. Вычислить разности значений  $\Delta$  для каждой пары значений признака, причем уменьшаемое всегда должно соответствовать одной группе, а вычитаемое — другой. В динамических исследованиях принято всегда вычитать значение “до” из значения “после”. В таких случаях положительное значение разности соответствует увеличению значений признака во времени, а отрицательное — уменьшению. Вычисление разностей удобно сделать в ППП **STATISTICA**, создав новый столбец и задав его значения формулой:

«=A2-A1», где A1 — название столбца, содержащего начальные значения признака, A2 — название столбца, содержащего конечное его значение.

2. Вычислить среднюю разность  $\Delta$ , а также ее стандартную ошибку  $m_{\Delta}$ . Это удобно сделать с использованием модуля “Описательная статистика”.

3. Вычислить границы ДИ для средней разности значений признака в двух группах по формулам:

$$(\Delta - t \times m_{\Delta}); (\Delta + t \times m_{\Delta}),$$

где  $t$  — значение  $t$ -критерия для числа степеней свободы  $(n-1)$  и выбранного уровня статистической значимости. Его значение для конкретного случая можно узнать, воспользовавшись опцией “Вероятностный калькулятор”.

### **STATISTICA:**

→ Модуль “Основные статистики и таблицы”  
 (“Basic statistics/Tables”)

→ Подмодуль “Вероятностный калькулятор”  
 (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “ $t$  Стьюдента” (“Student  $t$  test”);
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1- ф.р.”;
- задать число степеней свободы “Ст. св.” (“ $df$ ”);
- задать значение  $p$  (например, 0,05 для вычисления границ 95% ДИ);
- нажать кнопку “Вычислить” (“Compute”).

Искомое значение  $t$  появится в окне “ $t$ ”.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если ДИ для средней разности включает ноль, то следует считать, что различий между группами нет.
- Если 95% ДИ не включает ноль, то следует считать, что различия групп существуют с уровнем статистической значимости 0,05.
- Если ДИ включает как клинически значимые, так и клинически незначимые значения, то результаты недостаточно точны для того, чтобы сделать определенный вывод.

**Рекомендуемое представление результатов.** Привести следующую информацию:

- число участников исследования в каждой из групп;
- результаты, полученные с помощью методов описательной статистики (число объектов исследования, среднее, СКО) для каждой группы;
- среднюю разность (точечная оценка);
- границы 95% ДИ для средней разности (интервальная оценка).

### 8.3.2. Параметрический метод (*t*-критерий Стьюдента для зависимых групп)

**Задача.** Проверить, различаются ли средние значения количественного признака до и после лечения, если известно, что в обоих случаях распределение признака является нормальным.

*Пример:* сопоставить вязкость крови до и после лечения.

**Решение.** Использование *t*-критерия для связанных выборок (зависимых групп) возможно при условии выполнения его допущения — распределения признаков в каждом из сопоставляемых столбцов должны быть нормальными, а дисперсии равными. Поэтому применению *t*-критерия должна предшествовать проверка гипотезы о том, что распределения признака в каждом из сравниваемых столбцов являются нормальными: модуль “Основные статистики и таблицы” (“Basic statistics/Tables”), подмодуль “Описательная статистика” (“Descriptive statistics”), опция “Распределение” (“Distributon”) или подмодуль “Таблицы частот” (“Frequency tables”), опция “Проверка на нормальность” (“Tests of normality”).

С помощью *t*-критерия осуществляют проверку нулевой статистической гипотезы  $H_0$  об отсутствии различий групп. Если

нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий групп.

**Замечание.** В большинстве случаев использование *t*-критерия неправомерно, так как условия применимости метода не выполняются.

### STATISTICA:

→ Модуль “Основные статистики и таблицы”  
 (“Basic Statistics/Tables”)

→ Процедура “*t*-критерий для зависимых выборок” (“*t*-test for dependent samples”)  
 (рис. 8.6)

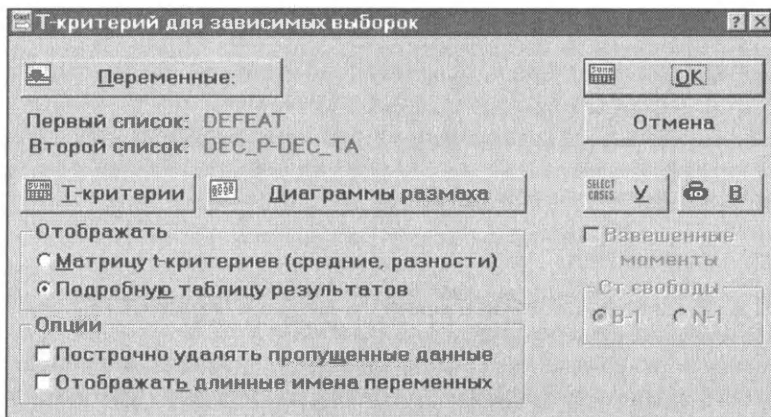


Рис. 8.6. Диалоговое окно процедуры “*t*-критерий для зависимых выборок” (в модуле “Основные статистики и таблицы”).

В окне результатов (рис. 8.7) следует обратить внимание на значения  $p$  в столбце “ $p$ ” ( $p$  — вероятность возникновения ошибки при отклонении нулевой гипотезы, т.е. вероятность ошибочно выявить различия между группами, хотя их на самом деле нет).

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза о том, что средние значения изучаемого признака в двух выборках не различаются, не отклоняется.

Перем.	Среднее G 1:1	Среднее G 2:2	t-знач.	Ст. св.	p	t отдел. оц. дисп.	Ст. св.	p
DEC P	18.27273	13.45745	2.504255	56	.015210	2.573943	15.53780	.020736
DEC TA	17.59091	13.47872	2.256763	56	.027944	2.165944	14.40014	.047553

Рис. 8.7. Окно результатов сравнения двух выборок по  $t$ -критерию для зависимых выборок.

- Если  $p < 0,05$ , то следует отклонить нулевую гипотезу и принять альтернативную гипотезу о существовании различий средних в двух выборках с уровнем статистической значимости  $p$ .

**Представление результатов.** Привести следующую информацию:

- число объектов исследования для каждой из выборок;
- утверждение о выполнении условий применимости метода (желательно аргументированное);
- средние значения изучаемого признака и СКО для каждой из групп;
- название статистического критерия;
- точное значение  $p$ .

### 8.3.3. Непараметрические методы (критерий знаков, критерий Вилкоксона)

**Задача.** Сравнить две связанные (зависимые) выборки по какому-либо признаку (или нескольким признакам) в любом из следующих случаев:

- анализируемый признак является количественным, но вид его распределений в выборках не анализировался;
- анализируемый признак является количественным, но хотя бы в одной из выборок его распределение не является нормальным;
- анализируемый признак является качественным порядковым.

**Пример:** сравнить показатель агрегации тромбоцитов у больных до и после лечения антиагрегантами в случае, если анализ видов распределений не проводился или проведенный анализ показал, что распределения признаков отличаются от нормальных.

**Решение.** Для сопоставления двух выборок по количественным признакам (которые не являются нормально распределенными в

исследуемых выборках, или вид распределений которых не анализировался) или порядковым признакам в ППП **STATISTICA** алгоритмизированы два метода:

- критерий знаков<sup>1</sup>. Условие применимости метода: анализируемые признаки должны быть количественными;
- критерий Вилкоксона для парных сравнений<sup>2</sup>. Этот критерий мощнее, чем критерий знаков.

Для применения любого из указанных методов сопоставляемые признаки (или один признак, наблюдаемый в разные моменты времени в одной группе объектов исследования) должны располагаться в таблице данных в разных столбцах (подробнее см. раздел 3.1.5). В диалоговом окне, активизировав кнопку “Переменные” (“Variables”), необходимо указать, которые из столбцов необходимо сопоставлять (например, столбец, в котором приведены значения признака до лечения, и столбец, в котором приведены значения признака после лечения).

При применении указанных критериев проверяется нулевая статистическая гипотеза  $H_0$  об отсутствии различий групп. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий групп.

### **STATISTICA:**

→ Модуль “Непараметрические статистики”  
 (“Nonparametrics/Distrib.”)

→ Процедуры:

- “Критерий знаков” (“Sign test”)
- “Критерий Вилкоксона для парных сравнений” (“Wilcoxon matched pairs test”) (рис. 8.8)

В окне результатов любой из этих процедур будет выведен столбец “ $p$ -уров.”, содержащий рассчитанные значения  $p$  при проверке статистической гипотезы.

**Интерпретация результатов.** Полученные результаты следует интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза об отсутствии различий групп по изучаемому признаку не отклоняется.

<sup>1</sup> Sign test (англ.).

<sup>2</sup> Wilcoxon matched pairs test (англ.).

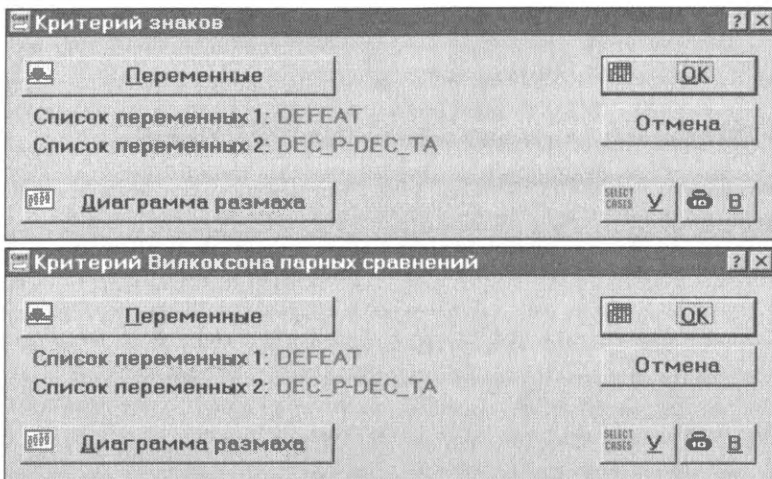


Рис. 8.8. Диалоговые окна процедур “Критерий Вилкоксона” и “Критерий знаков” (в модуле “Непараметрическая статистика”).

- Если  $p < 0,05$ , то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о существовании различий групп по изучаемому признаку.

**Представление результатов.** Привести следующую информацию:

- число объектов исследования для каждой из групп;
- медианы и границы интерквартильного отрезка для каждой из групп;
- название статистического теста;
- точное значение  $p$ .

## 8.4. Сравнение трех независимых (несвязанных) групп и более

**Задача.** Установить, различаются ли три группы или более по какому-либо одному количественному признаку.

**Пример:** определить, зависит ли активность фермента аспарагинаминотрансферазы от стадии заболевания.

**Решение.** Казалось бы, что в случае, если имеется три группы или более, нужно просто попарно сравнить их друг с другом. Однако такие сопоставления чреваты возникновением так называемой проблемы множественных сравнений (об этом подробно

см. в разделе 15.1). Поэтому для ситуации сравнения нескольких групп следует использовать другой метод — однофакторный дисперсионный анализ, состоящий из нескольких этапов.

Приведем классификацию методов дисперсионного анализа.

### 1. По количеству анализируемых признаков:

- однофакторный (анализ различий групп по одному признаку)<sup>1</sup>;
- многофакторный (анализ различий групп одновременно по двум признакам и более)<sup>2</sup>.

### 2. По принципам анализа:

- параметрический (для анализа нормально распределенных признаков в группах);
- непараметрический (для анализа количественного признака независимо от вида его распределения в группах).

### 3. По анализируемым данным:

- данные, полученные в несвязанных (независимых) выборках (в частности, данные однократных наблюдений);
- данные, полученные в связанных (зависимых) выборках (в частности, данные повторных наблюдений).

В данном разделе описаны методы однофакторного анализа вариаций для независимых (несвязанных) выборок:

- параметрический (см. раздел 8.4.1);
- непараметрический (метод Краскел—Уоллиса; см. раздел 8.4.2).

В разделе 8.5 описаны методы однофакторного анализа вариаций для связанных (зависимых) выборок.

## 8.4.1. Параметрический дисперсионный анализ

**Задача.** Сравнить три (или более) группы по количественному нормально распределенному признаку.

*Пример:* сравнить три группы больных с разными заболеваниями по содержанию в сыворотке крови липоперекисей.

**Решение.** Если необходимо сравнить средние значения количественного, нормально распределенного признака в группах, число которых больше двух, необходимо пользоваться параметрическим однофакторным анализом вариаций (ANOVA).

<sup>1</sup> Analysis of Variance (ANOVA; англ.).

<sup>2</sup> MANOVA (англ.).



В процедуре параметрического анализа вариаций общая вариация данных рассматривается как сумма двух видов вариаций — вариации между средним каждой группы и общим средним значением всей выборки (межгрупповая вариация), а также вариации между каждым объектом исследования группы и средним значением соответствующей группы (внутригрупповая вариация). Если межгрупповая вариация оказывается статистически значимо больше внутригрупповой вариации, то можно полагать, что различия между средними значениями групп существуют. В данном методе проверяется нулевая гипотеза о том, что группирующий фактор (группирующий признак) не влияет на значения исследуемого признака. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий групп.

**Замечание.** Если анализируются две группы, то ANOVA сводится к вычислению критерия Стьюдента.

#### **Условия применимости метода:**

- анализируемый признак является количественным;
- анализируемый признак нормально распределен в каждой из групп;
- дисперсии анализируемого признака в группах равны;
- группы определяются (детерминируются) качественным признаком (группирующий признак является качественным).

Таким образом, проведению параметрического дисперсионного анализа должна предшествовать проверка гипотез о виде распределений признака в группах (см. раздел 7.1). Проверка гипотез о равенстве дисперсий входит в собственно процедуру дисперсионного анализа в качестве его I этапа (см. раздел 8.4.1.1).

В случае, если указанные условия выполняются, применение параметрического дисперсионного анализа корректно. Обычно он проводится в три этапа:

1. Проверка гипотез о равенстве дисперсий (см. раздел 8.4.1.1).
2. Собственно анализ вариаций (см. раздел 8.4.1.2).
3. Апостериорное сравнение групп с помощью специальных процедур, отличных от  $t$ -критерия Стьюдента (см. раздел 8.4.1.3).

#### **8.4.1.1. Проверка гипотез о равенстве дисперсий**

Первым этапом параметрического дисперсионного анализа является проверка одного из условий его применимости, т.е. проверка равенства дисперсий в группах.

**Решение.** При использовании любого теста сравнения групп происходит проверка нулевой статистической гипотезы  $H_0$  об отсутствии различий дисперсий в группах. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различия дисперсий в группах.

### STATISTICA:

→ Модуль “Основные статистики и таблицы”  
 (“Basic statistics/Tables”)

→ Процедура “Группировка и однофакторный ANOVA” (“Breakdown and one-way ANOVA”)

или

→ Модуль “Дисперсионный анализ (ANOVA/MANOVA)”

В стартовом диалоговом окне необходимо активизировать кнопку “Переменные” (“Variables”) и указать в открывшемся окне группирующий признак (в левой части окна в случае использования модуля “Основные статистики и таблицы”) и один или несколько анализируемых независимых признаков (в правой части окна). Далее следует начать анализ, нажав кнопку “ОК”, при этом будет выведено следующее диалоговое окно (рис. 8.9).

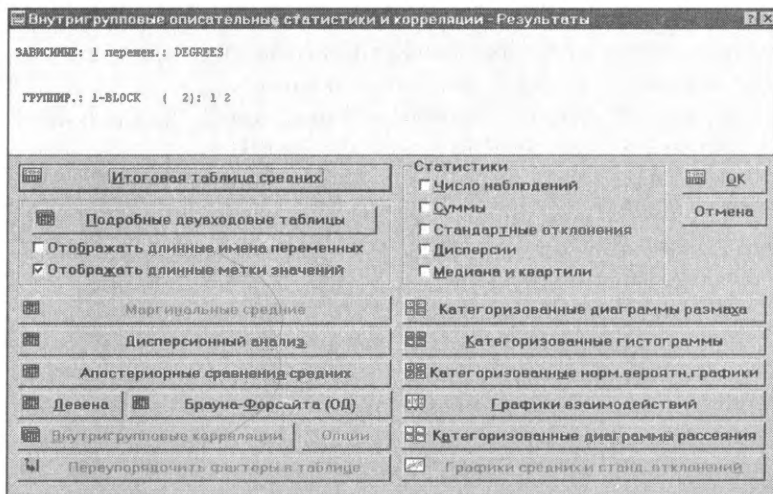


Рис. 8.9. Диалоговое окно процедуры “Группировка и однофакторный ANOVA” (в модуле “Основные статистики и таблицы”).

Далее следует вызвать опцию “Левена” (“Levene test [homogeneity of var.]”).

**Интерпретация результатов.** Тест Левена проверяет гипотезу о равенстве дисперсий исследуемого количественного признака в группах, определяемых значениями фактора (группирующего качественного признака).

- Если результат теста Левена свидетельствует об отсутствии различия дисперсий (при  $p > 0,05$ ), то применение параметрического дисперсионного анализа обосновано.
- Если результат свидетельствует, что различие дисперсий имеется ( $p < 0,05$ ), то применять параметрический дисперсионный анализ не следует. В этом случае можно воспользоваться процедурой непараметрического дисперсионного анализа (см. раздел 8.4.2).

**Представление результатов.** Привести следующую информацию:

- число объектов исследования в каждой из групп;
- средние значения и СКО изучаемого признака в каждой из групп;
- название статистического критерия;
- точное значение  $p$ .

#### 8.4.1.2. Собственно дисперсионный анализ

Для проведения собственно параметрического дисперсионного анализа существуют две возможности:

- 1) обратиться к опции “Дисперсионный анализ” (“Analysis of variance”) в диалоговом окне “Внутригрупповые описательные статистики и корреляции — Результаты” (“Descriptive statistics and correlations by groups”) (рис. 8.10, 8.11);
- 2) обратиться к модулю “Дисперсионный анализ” (“ANOVA/MANOVA”) (рис. 8.12).

При использовании данного статистического критерия проверяется нулевая статистическая гипотеза об отсутствии различий групп. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий групп.

В первом случае результаты будут представлены следующим образом (рис. 8.13). В данной таблице результатов следует обратить внимание на столбец “ $p$  двухст.”.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

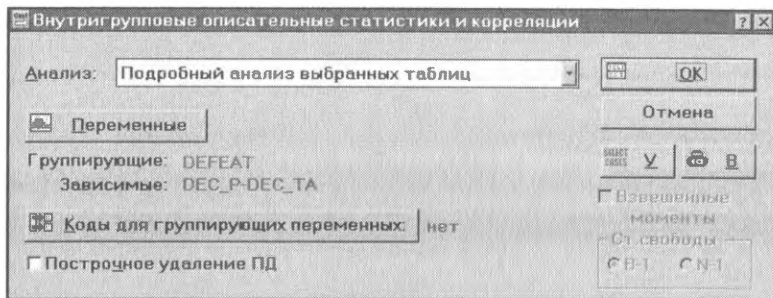


Рис. 8.10. Первое диалоговое окно дисперсионного анализа (модуль “Основные статистики и таблицы”, процедура “Группировка и однофакторный ANOVA”).

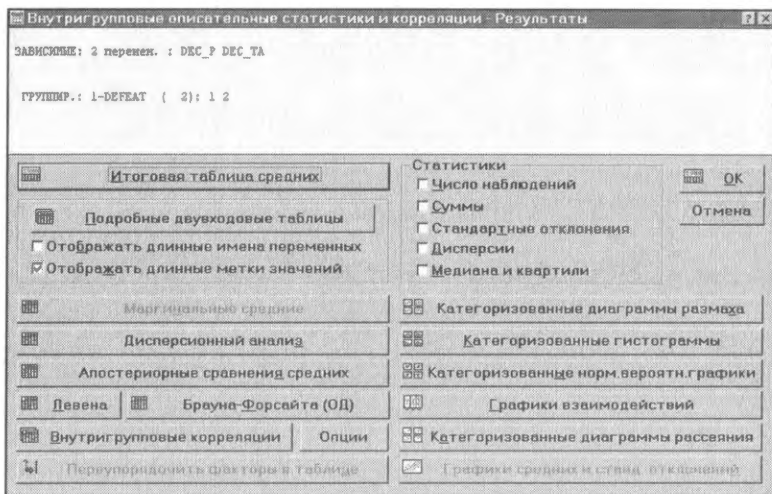


Рис. 8.11. Второе диалоговое окно дисперсионного анализа (модуль “Основные статистики и таблицы”, процедура “Группировка и однофакторный ANOVA”).

- Если  $p > 0,05$ , то нулевая гипотеза об отсутствии различий средних значений в группах не отклоняется, т.е. группы считаются не различающимися. В этом случае парный анализ групп нецелесообразен.
- Если  $p < 0,05$ , то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о существовании различий сред-

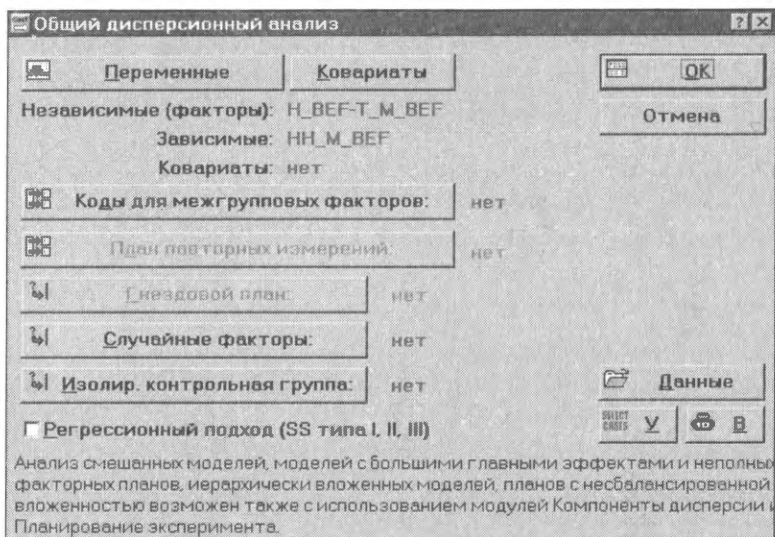


Рис. 8.12. Диалоговое окно дисперсионного анализа (модуль “Дисперсионный анализ (ANOVA/MANOVA)”).

Переменк.	Среднее G 1:1	Среднее G 2:2	t-знач.	Ст. св.	p	t отдел. оц. дисп.	Ст. св.	р двухст.
DEC P	18.27273	13.45745	2.504255	56	.015210	2.573943	13.53780	.020736
DEC TA	17.59091	13.47872	2.256763	56	.027944	2.165944	14.40014	.047553

Рис. 8.13. Окно результатов дисперсионного анализа.

них значений в группах. В этом случае следует перейти к апостериорному сравнению групп.

**Представление результатов.** Привести следующую информацию:

- число объектов исследования в каждой из групп;
- точное значение  $p$ ;
- ДИ для средних каждой группы (желательно). Способ их вычисления дан ниже;
- таблицу результатов анализа вариаций (желательно).

**Доверительные интервалы для средних в параметрическом дисперсионном анализе.** К сожалению, процедура для таких расчетов в ППП STATISTICA отсутствует, поэтому мы приводим ее математическое описание.

1. Сначала рассчитывают стандартную ошибку  $m_a$  для среднего значения в группе А:

$$m_a = s_{res} \sqrt{n_a} ,$$

где  $n_a$  — число объектов исследования в группе А,  $s_{res}$  — остаточное СКО (квадрат его значения — средний квадрат внутригрупповой вариации — выводится в таблице результатов (см. рис. 8.13) процедуры “Дисперсионный анализ” (“Analysis of variance”) в столбце “Ср. квадр. ошибки” (“MS Error”).

2. Затем вычисляют границы ДИ:

$$(M_a - t \times m_a); (M_a + t \times m_a),$$

где  $t$  — значение  $t$ -критерия для числа степеней свободы ( $n-1$ ) и выбранного уровня статистической значимости. Его значение для конкретного случая можно узнать, воспользовавшись опцией “Вероятностный калькулятор”.

### STATISTICA:

→ Модуль “Основные статистики и таблицы”  
 (“Basic statistics/Tables”)

→ Подмодуль “Вероятностный калькулятор”  
 (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “ $t$  Стьюдента” (“Student  $t$  test”);
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1- ф.р.”;
- задать число степеней свободы “Ст. св.” (“ $df$ ”);
- задать значение  $p$  (например, 0,05 для вычисления границ 95% ДИ);
- нажать кнопку “Вычислить” (“Compute”).

Искомое значение  $t$  появится в окне “ $t$ ”.

3. Аналогичные вычисления проводятся для средних значений других групп.

#### 8.4.1.3. Апостериорные сравнения групп

Если при анализе вариаций получены статистически значимые результаты о различии групп, то у исследователя возникает естественное стремление выяснить, которые же из нескольких

групп попарно отличаются друг от друга. Применение в этой ситуации обычных методов сравнения двух групп (например, критерия Стьюдента) не может быть рекомендовано, так как многократные сравнения нескольких групп попарно с помощью этого теста порождают так называемую проблему множественных сравнений (подробнее см. раздел 15.1). Для решения задачи парных сравнений нескольких групп разработаны специальные методы — “Апостериорные сравнения средних” (“Post-hoc comparisons of means”). При использовании любого из них проверяется нулевая статистическая гипотеза об отсутствии различий групп. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий групп.

В ППП **STATISTICA** представлены несколько таких процедур. В промежуточном диалоговом окне (вызываемом нажатием кнопки “Переменные”) необходимо указать анализируемые признаки, а затем выбрать одну или несколько из процедур апостериорного сравнения средних (рис. 8.14).

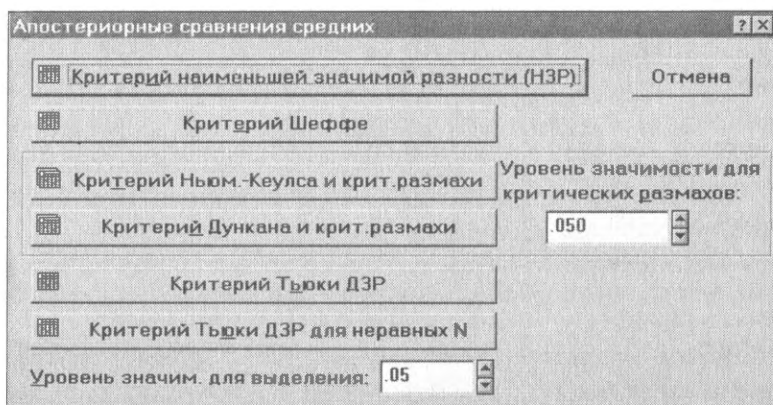


Рис. 8.14. Диалоговое окно процедур сравнения (подмодуль “Группировка и однофакторная ANOVA” модуля “Основные статистики и таблицы”).

#### STATISTICA:

- Модуль “Основные статистики и таблицы”  
 (“Basic statistics/Tables”)
- Подмодуль “Группировка и однофакторный ANOVA”  
 (“Breakdown and one-way ANOVA”)

→ Процедура “Апостериорные сравнения средних” (“Post-hoc comparisons of means”)

Воспользовавшись каким-либо из имеющихся методов, можно выявить пары групп, средние значения которых различаются статистически значимо. Результаты (значения  $p$ , полученные при сопоставлении групп попарно) представляются в таблице результатов, симметричной относительно главной диагонали (главная диагональ соединяет левый верхний и правый нижний углы таблицы).

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевую гипотезу  $H_0$  об отсутствии различий групп не отклоняют.
- Если  $p < 0,05$  для какой-либо из пар групп, то следует отклонить нулевую гипотезу и принять альтернативную гипотезу о существовании различий этих групп.
- Возможен вариант, когда, несмотря на полученное на этапе собственно дисперсионного анализа статистически значимое различие групп, ни в одной из пар групп при парном сравнении не будет получено статистически значимых различий. Это связано со значительной консервативностью методов множественного сравнения, т.е. их малой статистической мощностью (чувствительностью) в выявлении незначительных различий при небольших размерах групп. В этом случае рекомендуются следующие варианты дальнейших действий:
- переход к парному сравнению групп с помощью методов сопоставления двух выборок (например, критерию Стьюдента) с ограничением числа сравниваемых групп и применением поправки Бонферрони (т.е. с делением установленного уровня значимости  $p$  (например, 0,05) на количество парных сравнений (например, на 6) и установкой полученного частного (в нашем примере  $0,05/6=0,008$ ) в качестве нового уровня статистической значимости; подробнее см. раздел 15.1);
- объединение некоторых групп и повторение анализа.

**Представление результатов.** Привести следующую информацию:

- ДИ (желательно) для разности средних значений пар групп, которые статистически значимо различаются. Способ вычисления таких ДИ описан ниже, так как этот алгоритм в ППП *STATISTICA* отсутствует;



- название статистического критерия апостериорного сравнения групп;
- точное значение  $p$  для каждой пары сравниваемых групп.

**Вычисление доверительного интервала для разности средних любых двух групп в параметрическом дисперсионном анализе.**

Сначала рассчитывают стандартную ошибку для разности средних  $m_{\Delta}$ :

$$m_{\Delta} = s_{res} \sqrt{\frac{1}{n_a} + \frac{1}{n_b}},$$

где  $n_a$  — число наблюдений (объектов исследования) в группе А,  $n_b$  — число наблюдений (объектов исследования) в группе В,  $s_{res}$  — остаточное СКО (квадрат его значения — средний квадрат внутригрупповой вариации, который выводится в таблице результатов (см. рис. 8.13) процедуры “Анализ вариаций” (“Analysis of variance”) в столбце “Ср. квадр. ошибки” (“MS Error”).

2. Затем вычисляют границы ДИ для разности между средними значениями  $\Delta = (M_a - M_b)$  в двух группах:

$$(\Delta - t \times m_{\Delta}); (\Delta + t \times m_{\Delta}),$$

где  $t$  — значение  $t$ -критерия для числа степеней свободы ( $n_a + n_b - 2$ ) и выбранного уровня статистической значимости, обычно принимается равным 1,96 для 95% ДИ.

Значение  $t$  для конкретного случая можно узнать, воспользовавшись опцией “Вероятностный калькулятор”:

### STATISTICA:

- Модуль “Основные статистики и таблицы” (“Basic statistics/Tables”)
  - Подмодуль “Вероятностный калькулятор” (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “ $t$  Стьюдента” (“Student  $t$  test”);
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1- ф.р.”;
- задать число степеней свободы “Ст. св.” (“df”);

- задать значение  $p$  (например, 0,05 для вычисления границ 95% ДИ);
  - нажать кнопку “Вычислить” (“Compute”).
- Искомое значение  $t$  появится в окне “t”.

### 8.4.2. Непараметрические методы сравнения независимых групп (метод Краскела—Уоллиса, медианный тест)

**Задача.** Сопоставить три группы и более по одному количественному или порядковому признаку независимо от вида его распределения в группах.

*Пример:* сравнить три группы больных с разным патогенезом острого инсульта по значению неврологического дефицита, измеренного по Скандинавской шкале.

**Решение.** В ППП *STATISTICA* реализованы два метода решения такой задачи:

- 1) ранговый анализ вариаций по Краскелу—Уоллису<sup>1</sup> — обобщение метода Манна—Уитни для сравнения трех групп и более;
- 2) медианный тест<sup>2</sup> — грубый вариант метода Краскела—Уоллиса. Этот тест наиболее эффективен в случаях, если шкала измерений признака имеет искусственные границы, т.е. большое число объектов исследования приходится на крайние значения шкалы.

При использовании каждого из методов проверяется нулевая статистическая гипотеза  $H_0$  об отсутствии различий групп. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий групп.

**Условия применимости метода.** Анализируемый признак должен быть количественным или порядковым. Ограничений на вид распределения нет, поэтому перечисленные методы служат альтернативой параметрическому дисперсионному анализу в следующих случаях:

- если распределение признака не является нормальным;
- если вид распределения неизвестен (не исследовался).

<sup>1</sup> Kruskal—Wallis ANOVA (англ.).

<sup>2</sup> Median test (англ.).

## STATISTICA:

- Модуль “Непараметрические статистики” (“Nonparametrics/Distrib.”)
- Процедура “ANOVA Краскела—Уоллиса и медианный тест” (“Kruskal—Wallis ANOVA, median test”) (рис. 8.15).

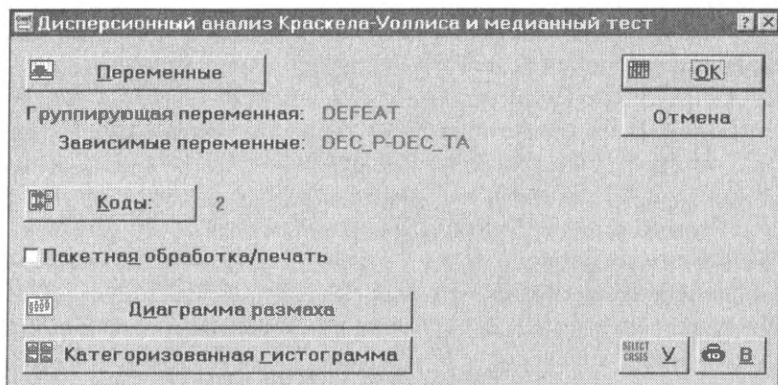


Рис. 8.15. Диалоговое окно процедуры “ANOVA Краскела—Уоллиса и медианный тест” (модуль “Непараметрические статистики”).

Сначала в диалоговом окне процедуры необходимо активизировать кнопку “Переменные” (“Variables”) и указать группирующий признак (в левой части открывшегося окна) и анализируемые признаки (в правой части этого же окна). Затем следует собственно запустить анализ. В верхних частях окон результатов (рис. 8.16) будут выведены значения  $p$  для критерия Краскела—Уоллиса и медианного теста соответственно.

**Интерпретация результатов.** В ANOVA Краскела—Уоллиса проверяется гипотеза о том, получены ли изучаемые группы из одной генеральной совокупности либо из разных генеральных совокупностей с равными медианами.

- Если  $p > 0,05$ , то нулевую гипотезу об отсутствии различий групп, т.е. о том, что они получены из одной генеральной совокупности или из генеральных совокупностей с равными медианами, не отклоняют.

Ранговый дисперсионный анализ Краскела-Уоллиса...			
НЕПАРАМ. СТАТИСТ.	Независимая переменная: DEFEAT		
Завис.:	критерий К-У: $H(1, N 58) = 10.61$		
DEC_P	Код	Число набл.	Сумм. ранг.
Группа1	1	11	489.500
Группа2	2	47	1222.500

Медианный тест, общ медиана= 17.00000 (gr23-min.sta)			
Далее...	Независимая переменная: DEFEAT		
	хи-квадрат = 10.08431, $сс = 1$ , $p = .0015$		
Зависимые:	Группа1	Группа2	Общий
DEC_P			
<= Медиана: набл.	2.00000	33.00000	35.00000
ожидаемые	6.63793	26.36207	
набл.-ожд.	-4.63793	4.63793	
> Медиана: набл.	9.00000	14.00000	23.00000
ожидаемые	4.36207	18.63793	
набл.-ожд.	4.63793	-4.63793	
Сумма: набл.	11.00000	47.00000	58.00000

Рис. 8.16. Окна результатов ANOVA Краскела—Уоллиса и медианного теста (выводятся на дисплей одновременно).

- Если  $p < 0,05$ , то следует отклонить нулевую гипотезу и принять альтернативную гипотезу о различии групп, т.е. делается вывод о том, что группы получены из разных генеральных совокупностей или генеральных совокупностей с разными медианами. В этом случае рекомендуется провести парное сравнение групп с использованием непараметрического теста Манна—Уитни (см. раздел 8.2.3), применяя поправку Бонферрони при оценке значения  $p$  (см. раздел 15.1).
- Медианный тест проверяет ту же гипотезу путем анализа генерируемых таблиц сопряженности с использованием критерия  $\chi^2$ . Интерпретация в отношении вычисляемого  $p$  аналогична ANOVA.

**Представление результатов.** Привести следующую информацию:

- число объектов исследования в каждой из групп;
- медианы и квартили для каждой группы;
- название статистического критерия;
- точное значение  $p$ .

## 8.5. Сравнение трех связанных (зависимых) групп и более (непараметрический метод Фридмена)

**Задача.** Сопоставить значения признака на разных этапах динамического исследования.

*Пример:* сравнить значения вязкости крови на трех этапах лечения препаратом, улучшающим реологические свойства крови.

**Решение.** Необходимо сопоставить значения одного и того же признака в разные моменты времени, т.е. сопоставить не группы участников исследования, а одних и тех же участников исследования в разные моменты времени (повторные наблюдения, представленные в разных столбцах таблицы данных). Для этого в ППП **STATISTICA** предлагаются следующие процедуры непараметрического дисперсионного анализа:

- ранговый дисперсионный анализ по Фридмену<sup>1</sup>;
- вычисление коэффициента конкордации Кендалла<sup>2</sup>. Этот коэффициент часто используется также для решения о том, насколько согласованы оценки экспертов.

При использовании данного статистического критерия проверяется нулевая статистическая гипотеза об отсутствии различий групп. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий групп.

**Условия применимости метода.** Анализируемые признаки должны быть количественными (с любым видом распределения).

### **STATISTICA:**

→ Модуль “Непараметрические статистики”  
 (“Nonparametrics/Distrib.”)

→ Процедура “Фридмана ANOVA и конкордация Кендалла” (“Friedman ANOVA & Kendall’s concordance”)

В открывающемся диалоговом окне необходимо указать сопоставляемые признаки (столбцы таблицы данных), а затем выполнить анализ. Будет выведено окно результатов (рис. 8.17).

<sup>1</sup> Friedman ANOVA (англ.).

<sup>2</sup> Kendall’s concordance (англ.).

Фридмана ANOVA и конкордация Кендалла (gr23-min.sta)				
ANOVA хи-квад. (N = 58, cc = 1) = 1.230769 p < .26727				
Коэфф. конкордации = .02122 Сред. ранг r = .00405				
Перем.	Средний ранг	Сумма рангов	Среднее	Ст. откл
DEC_P	1.568966	91.00000	14.37069	6.000410
DEC_TA	1.431034	83.00000	14.25862	5.632176

Рис. 8.17. Окно результатов процедуры “Фридмана ANOVA и конкордация Кендалла”.

**Интерпретация результатов.** В ходе анализа проверяется гипотеза о том, что указанные признаки получены из одной и той же генеральной совокупности или из разных генеральных совокупностей с равными медианами.

- Если  $p > 0,05$ , то нулевая гипотеза о том, что признаки получены из одной генеральной совокупности или генеральных совокупностей с равными медианами, не отклоняется.
- Если  $p < 0,05$ , то следует отклонить нулевую гипотезу и принять альтернативную гипотезу о том, что признаки получены из разных генеральных совокупностей или генеральных совокупностей с разными медианами. В этом случае рекомендуется провести парное сравнение групп с использованием непараметрического теста Вилкоксона (см. раздел 8.3.3), применяя поправку Бонферрони при оценке вычисленного  $p$ -значения (см. раздел 15.1).

Коэффициент конкордации Кендалла отражает степень связи признаков. В данной процедуре вычисляется его значение.

**Представление результатов.** Привести следующую информацию:

- число объектов исследования для каждой из групп;
- медианы и квартили для каждой группы;
- название теста;
- точное значение  $p$ .

## Глава 9. Описание качественных признаков

Как уже упоминалось ранее, качественные признаки (переменные, данные) могут быть номинальными (иногда используется термин “номинативные”), в том числе бинарными, и порядковыми.

**Номинальными** называются признаки, значения которых представляют собой условные коды неизмеряемых категорий (например, коды пола или коды диагноза). Значения таких признаков не могут быть упорядочены по какому-либо семантическому принципу.

**Порядковыми (ранговыми)** называются признаки, значения которых отражают степень выраженности какой-либо характеристики объекта исследования (например, стадии заболевания, степени выраженности симптома). При этом в отличие от количественных признаков “расстояния” между значениями порядковых признаков не могут быть оценены с помощью какой-либо известной шкалы, однако все же значения порядкового признака могут быть упорядочены (ранжированы). Порядковые признаки являются качественными (иногда их называют “полуколичественными”) оценками какой-либо характеристики.

Все объекты исследования выборки могут быть разделены на подгруппы в соответствии со значениями любого из имеющихся качественных признаков, например, по полу. Число объектов исследования с определенным значением качественного признака называется **абсолютной частотой**. Кроме абсолютных частот широко используются и относительные частоты. **Относительная частота** — это отношение числа объектов с каким-либо значением признака к объему числа объектов.

Анализ качественных данных начинается с простейшего подсчета абсолютных и относительных частот для каждого значения (градации) такого признака. Если объекты исследования в группе распределяются в соответствии со значениями двух качественных признаков, то строят прямоугольные таблицы сопряженности, или таблицы кросстабуляции (рис. 9.1).

Конституция	Мужчины	Женщины
Астеники	5	8
Нормостеники	25	30
Гиперстеники	12	11

**Рис. 9.1.** Пример таблицы сопряженности признаков пола и конституции.

В ячейках таблицы указаны количества объектов исследования, соответствующих различным сочетаниям значений признаков (абсолютные частоты).

В общем случае при наличии нескольких качественных признаков для анализа взаимосвязи между отдельными парами качественных признаков строятся таблицы сопряженности для всех возможных пар таких признаков.

Если число качественных признаков, по значениям которых распределяются объекты исследования выборки, равно трем и больше, то возможно построение таблиц сопряженности с тремя осями в многомерном пространстве признаков и более. Для анализа таких таблиц используют достаточно сложные методы статистического анализа, такие как логлинейный анализ, логистическая регрессия, анализ деревьев классификаций и т.д. Эти методы в настоящем издании не описываются, так как представляют собой достаточно сложные статистические задачи. Следует помнить, однако, что при увеличении числа признаков, по которым табулируются объекты исследования, абсолютные частоты в ячейках таблицы резко уменьшаются, а, следовательно, становится меньше шансов получить надежные статистические выводы при анализе таких таблиц.

Для порядковых признаков имеет смысл вычислять параметры распределения — медиану, моду, квартили.

Для номинальных признаков целесообразнее представление в виде относительных частот объектов исследования.

Для признаков бинарного типа (т.е. имеющих только два значения — “да” и “нет”, часто такие признаки называют еще дихотомическими) — представление в виде относительной частоты (доли, пропорции) и ее доверительного интервала.

Дальнейший анализ качественных признаков состоит обычно в сопоставлении соответствующих частот в исследуемых группах



или сравнении параметров распределения. Это может быть проверка гипотез о равенстве медиан в сравниваемых группах или сравнение долей (пропорций) в этих группах. Методы сравнения групп по качественным признакам описаны в главах 10 и 11.

**Замечание.** Частой ошибкой исследователей является анализ порядковых данных (имеющих небольшое число допустимых значений) как количественных. Действительно, никому не может прийти в голову вычислить средний пол, который является качественным номинальным признаком, в изучаемой выборке, но очень часто вычисляют, например, средний балл, если какой-либо качественный порядковый признак измеряется в баллах по какой-либо шкале (например, 5-балльной), хотя это не меньшая бессмыслица. Ведь для качественных данных не определено "расстояние" между значениями порядкового признака, поэтому интерпретировать промежуточные (дробные) значения невозможно (см. также главу 2). Наиболее распространенной ошибкой в этом случае является использование *t*-критерия Стьюдента для сравнения средних в группах. Многочисленные примеры таких ошибок читатели могут найти в Интернете по адресу <http://www.doktor.ru/doctor/biometr/kk/index.htm>

Следует также отметить, что в современной статистике разработаны, а в ряде статистических пакетов и реализованы специальные методы оцифровки качественных признаков [20].

Тем не менее, как уже упоминалось в главе 2, при достаточно большом числе возможных значений качественного порядкового признака (обычно приближающемся к 20) на практике для анализа таких признаков могут применяться те же непараметрические методы, что и для количественных признаков.

*Пример:* Скандинавская шкала оценки неврологического дефицита при инсульте имеет минимально возможное значение 0, максимальное значение 60. В подобных случаях принято применять те же методы анализа, что и для количественных признаков (предварительно анализируя, как обычно, вид распределения).

Существуют следующие основные способы описания качественных данных:

- вычисление параметров распределения качественных данных
- моды (для номинальных данных), медианы, моды и квартилей (для порядковых данных);

- вычисление абсолютных и относительных частот (пропорций, долей, процентов) и доверительных интервалов для них.

## 9.1. Вычисление параметров распределения качественных признаков

Для описания распределения качественных номинальных признаков можно использовать единственный параметр. Этим параметром (параметром положения) является **мода ( $M_o$ )** — наиболее часто встречающееся на данной выборке, типичное для нее значение номинального признака.

Моду также можно вычислять и для качественных порядковых признаков. Вычисление других параметров распределения возможно только для порядковых признаков (но не номинальных!):

- **медиана ( $M_e$ )** — центральное значение признака в выборке, слева и справа от которого расположены равные количества объектов исследования;
- **процентили распределения**, в том числе **нижний квартиль ( $LQ$ , или  $Q_1$ )** и **верхний квартиль ( $UQ$ , или  $Q_3$ )**<sup>1</sup> — для характеристики рассеяния в выборке.

Вычисление этих параметров можно осуществить с использованием следующей процедуры пакета.

### STATISTICA:

→ Модуль “Непараметрические статистики”  
 (“Nonparametrics/Distrib.”)

→ Процедура “Обычные описательные статистики (медиана, мода...)” “Ordinal descriptive statistics (median, mode...)”

Медиану и квартили также можно рассчитать, используя другой модуль пакета.

### STATISTICA:

→ Модуль “Основные статистики и таблицы”  
 (“Basic statistics/Tables”) (см. рис. 7.1)

<sup>1</sup> Напомним, что нижний квартиль отсекает 25% объектов с наименьшими значениями признака, а верхний квартиль — 25% объектов с наибольшими значениями признака.

- Подмодуль “Описательные статистики” (“Descriptive statistics”) (см. рис. 7.2)
- Кнопка “Другие статистики” (“More statistics”)
- Окно выбора необходимых статистик (см. рис. 7.6)

**Рекомендуемое представление результатов.** Привести следующую информацию:

- для номинального признака — значение моды;
- для порядкового признака — значение медианы и квартилей в формате  $Me$  ( $LQ$ ;  $UQ$ ) или значение моды и квартилей в формате  $Mo$  ( $LQ$ ;  $UQ$ ).

## 9.2. Вычисление абсолютных и относительных частот (долей, процентов, вероятностей, шансов)

Часто задачей исследования является подсчет числа объектов исследования (абсолютной частоты) или их доли (относительной частоты) в исследуемой выборке, соответствующей значению какого-либо признака. Иными словами, необходимо ответить на вопрос: сколько больных (абсолютное число, абсолютная частота) или какая доля, процент от общего числа больных (относительная частота) характеризовались тем или иным значением признака.

**Абсолютная частота признака** — это количество объектов исследования в группе, имеющих данную характеристику (значение признака).

**Относительная частота признака (распространенность)** — это доля объектов исследования в группе, имеющих данную характеристику (значение признака) по отношению к общему числу наблюдений (объектов исследования) в группе. Выражается обычно в десятичных долях единицы (например, 0,3) или процентах (например, 30%).

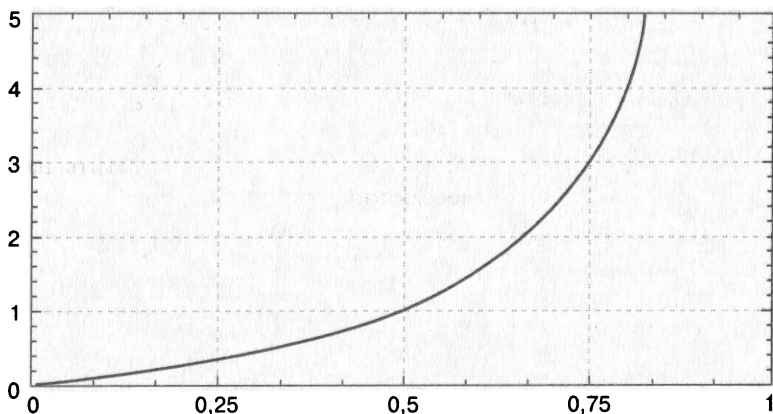
В современной литературе также широко используются понятия “пропорция”, “доля”, “вероятность” и “шанс”. Последний термин применяется преимущественно по отношению к бинарным данным.

**Пропорция**<sup>1</sup>, **доля**<sup>2</sup>, **вероятность**<sup>3</sup> — это относительная частота события, выраженная в десятичных долях единицы или в процентах. Изменяется в интервале от 0 до 1, или от 0 до 100%.

**Шанс**<sup>4</sup> — отношение вероятности того, что событие произойдет, к вероятности того, что событие не произойдет. Другое определение: шанс — это отношение частоты возникновения события к частоте его отсутствия. Значение шанса изменяется в интервале от нуля до бесконечности.

*Пример:* если вероятность возникновения осложнений после операции равна  $V$  (например, 0,3), то вероятность того, что оно не произойдет, равна  $1-V$  (в нашем примере 0,7). Шанс события равен отношению этих вероятностей (в нашем примере  $3:7=0,43$ )<sup>5</sup>.

Функциональную взаимосвязь шанса и вероятности для бинарного признака (события) можно изобразить на графике следующим образом (рис. 9.2).



**Рис. 9.2.** Функциональная взаимосвязь вероятности события (ось абсцисс) и его шанса (ось ординат).

<sup>1</sup> Proportion (англ.).

<sup>2</sup> Rate (англ.).

<sup>3</sup> Probability (англ.).

<sup>4</sup> Odds (англ.).

<sup>5</sup> В общем виде вероятность ( $V$ ) и шанс ( $Ш$ ) связаны следующими формулами:  
 $Ш = V / (1 - V)$ ;  $V = Ш / (1 + Ш)$ .

Относительные частоты удобно представлять в виде процентов. Это делается с целью дальнейшего сопоставления разных по объему выборок по каким-либо характеристикам, частота которых в выборке представлена в виде процентов (описание методов сравнения относительных частот в разных группах приведено в главах 10 и 11).

Вычисление абсолютных частот и процентов для одного качественного признака удобно производить путем построения таблиц частот и процентов в модуле “Основные статистики и таблицы”.

### STATISTICA:

→ Модуль “Основные статистики и таблицы”  
 (“Basic Statistics /Tables”)

→ Процедура “Таблицы частот”  
 (“Frequency tables”) (рис. 9.3)

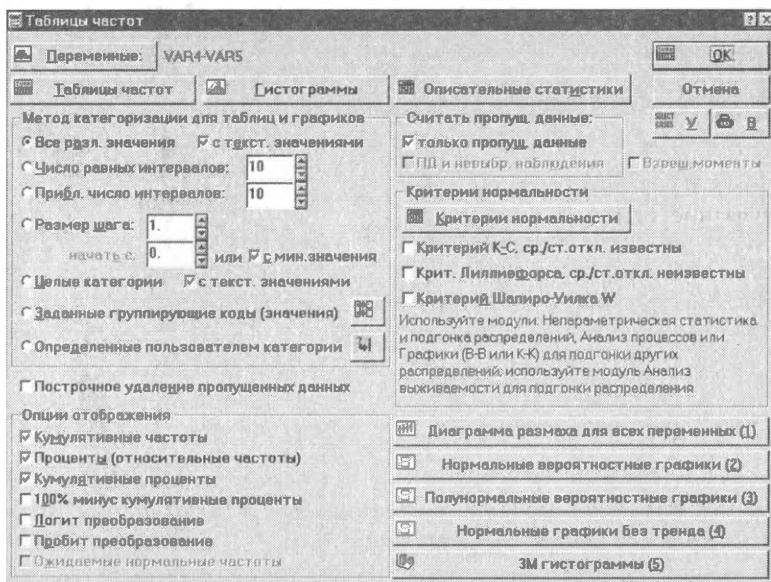


Рис. 9.3. Стартовое диалоговое окно “Таблицы частот” (в модуле “Основные статистики и таблицы”).

В диалоговом окне сначала надо активизировать кнопку “Переменные” (“Variables”) и указать анализируемые признаки.

Далее следует запустить программу расчетов, после чего в отдельном окне появится результат (рис. 9.4) — таблица частот, в которой отражены как абсолютные значения частот (количество объектов исследования с тем или иным значением качественного признака в анализируемой выборке), так и проценты объектов исследования.

Далее...	Частота	Кумул. частота	Процент	Кумул. Процент
0.00000	1	1	1.25000	1.2500
1.00000	6	7	7.50000	8.7500
2.00000	7	14	8.75000	17.5000
3.00000	3	17	3.75000	21.2500
Пропущ.	63	80	78.75000	100.0000

Рис. 9.4. Окно результатов построения таблицы частот.

**Рекомендуемое представление результатов.** Проценты необходимо представлять в тексте статьи или таблице, одновременно указывая абсолютное значение той величины, которая принята за 100%, например 25% из 120 больных. Другой способ — указание одновременно и процентов, и абсолютных значений, например “25% (30/120)” или “30 из 120 больных (25%)”.

В случае, если проводят последовательный пересчет процентов, т.е. вычисляют процент от процента (процент от числа объектов исследования в ранее описанной процентами подгруппе), необходимо понятно описать эту процедуру и представить количества объектов исследования, принимаемые последовательно за 100%.

Необходимая точность приводимых значений процентов зависит от объема выборки:

- Так называемые малые выборки (менее 20 объектов исследования) вообще не принято описывать процентами (так как значение процента оказывается в таких случаях значительно больше абсолютного числа объектов исследования). В этих случаях указываются абсолютные значения частот для значений того или иного признака.
- Если объем выборки составляет от 20 до 100 объектов исследования, то проценты представляют в виде целых чисел.

— Если объем выборки больше 100 объектов исследования, то процент указывается не более чем с одним разрядом десятичной дроби.

**Замечание.** Модуль “Таблицы частот” (“Frequency Tables”) можно использовать только для построения одномерных таблиц частот, т.е. для построения таблиц частот отдельно для каждого признака.

### 9.3. Описание относительной частоты бинарного признака с использованием доверительного интервала

В современной научной литературе относительные частоты бинарных признаков — “событий” (т.е. признаков, имеющих только два возможных значения — “да” и “нет”) — принято приводить вместе с указанием их ДИ.

Напомним, что ДИ — это интервал, в котором с некоторой вероятностью (например, 95% в случае 95% ДИ) находится истинное популяционное значение. Границы ДИ вычисляют на основании анализа данных выборки.

**Задача:** оценить популяционное значение относительной частоты какого-либо состояния по данным выборки.

*Пример:* благоприятный результат стереотаксической операции наблюдается у 67% больных идиопатическим паркинсонизмом в изучаемой выборке. Возникает вопрос: какой процент благоприятных результатов будет наблюдаться при применении данной операции у пациентов данной категории в другой аналогичной клинике, другим врачом подобной квалификации и т.д. (т.е. в общей популяции *таких* больных)? Понятно, что в общей популяции будет наблюдаться определенный разброс в результатах этого метода лечения. Вычислить границы этого разброса на основании собственных данных — и означает определить границы ДИ.

**Решение:** ДИ определяется, как всегда в статистике, с определенной долей уверенности. Обычно используют коэффициент доверительной вероятности 95%, но можно использовать 90%, 99% и др. При этом с соответствующей долей уверенности можно считать, что истинное популяционное значение находится в границах такого ДИ.

**Замечание.** Вычисление границ ДИ для бинарного признака — достаточно сложная задача. Существует несколько способов ее решения (см., например, [21, 22]). К сожалению, в ППП STATISTICA не содержится программы для вычисления таких ДИ.

Поэтому в Приложении 4 мы приводим таблицу с точно вычисленными границами ДИ, а также математическое описание одного из простых способов вычисления границ ДИ для бинарного признака.

**Условия применимости метода.** Использование значений  $t$ -критерия в описанной ниже формуле возможно в случае одновременного выполнения следующих условий:

- $nP \geq 5$ ,
- $n(1-P) \geq 5$ ,
- $0,3 \leq P \leq 0,7$ ,

где  $P$  — относительная частота события, выраженная десятичной дробью,  $n$  — общее число объектов исследования в выборке. Перечисленные ограничения применимости метода связаны с необходимостью обоснования использования нормальной аппроксимации биномиального распределения [21, 23].

Если указанные условия не выполняются, то следует пользоваться более сложными формулами, которые мы в данном издании не приводим, либо таблицей, приведенной в Приложении 4.

При вычислении соответственно верхней и нижней границ ДИ для относительной частоты бинарного признака (доли) можно использовать следующую приближенную формулу:

$$P \pm t \times \left( \sqrt{\frac{P(1-P)}{n} + \frac{1}{2n}} \right),$$

где  $P$  — относительная частота события, выраженная десятичной дробью,  $n$  — общее число объектов исследования в выборке,  $t$  — значение  $t$ -критерия, соответствующее объему исследуемой выборки и ДК (95, 90 или 99%), обычно принимается приближенно

равным 1,96 для 95% ДИ,  $\frac{1}{2n}$  — поправка на непрерывность, компенсирующая ошибку, возникающую при аппроксимации биномиального распределения нормальным.



Точное значение  $t$  для конкретного случая можно узнать, воспользовавшись опцией “Вероятностного калькулятора”:

### STATISTICA:

- Модуль “Основные статистики и таблицы” (“Basic statistics/Tables”)
- Подмодуль “Вероятностный калькулятор” (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “ $t$  Стьюдента” (“Student  $t$  test”);
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1–ф.р.”;
- задать число степеней свободы “Ст. св.” (“ $df$ ”),  $df=n-1$ ;
- задать значение  $p$  (например, 0,05 для вычисления границ 95% ДИ);
- нажать кнопку “Вычислить” (“compute”).

Искомое значение  $t$  появится в окне “ $t$ ”.

**Интерпретация результата.** С определенной долей уверенности (соответствующей ДК) можно утверждать, что истинное популяционное значение лежит в границах вычисленного ДИ.

**Замечание.** Особыми являются ситуации, когда  $P=0,00$  или  $P=1,00$ . При  $P=0,00$  вычисляется только верхняя граница ДИ,  $P=1,00$  — только нижняя граница ДИ. В этих случаях ширина ДИ будет больше (на 2,5% для 95% ДИ, на 0,5% для 99% ДИ, на 5% для 90% ДИ), так как неопределенности в одном из направлений не существует. В этих ситуациях не следует пользоваться приведенной формулой, а рекомендуется воспользоваться таблицей, приведенной в Приложении 4. Она содержит более точно рассчитанные 95% ДИ.

---

## Глава 10. Сравнение групп по качественному признаку

Выбор метода сравнения групп по качественному признаку зависит от типа этого признака, количества и связанности сопоставляемых групп. Ниже перечислены сочетания этих факторов и ссылки на соответствующие методы и разделы настоящего издания:

- **признак бинарный** (имеет только два возможных значения) — методы для анализа таких признаков описаны в главе 11.
- **признак номинальный** (имеет три неупорядоченных значения и более) — тест  $\chi^2$  (применяется для всех перечисленных ниже случаев):
  - сравнение наблюдаемых и ожидаемых частот (анализ одной группы; см. раздел 10.1);
  - сравнение двух групп и более (см. раздел 10.2);
- **признак порядковый** (имеет три упорядоченных значения и более);
  - сравнение наблюдаемых и ожидаемых частот (анализ одной группы; см. раздел 10.1);
  - сравнение двух несвязанных групп;
    - тесты Манна—Уитни, Вальда—Вольфовица, Колмогорова—Смирнова (см. раздел 8.2.3);
    - критерий  $\chi^2$  (см. раздел 10.2);
  - сравнение двух связанных групп;
    - критерий Вилкоксона, критерий знаков (см. раздел 8.3.3);
    - критерий  $\chi^2$  (см. раздел 10.2);
  - сравнение трех несвязанных групп и более;
    - анализ вариаций (ANOVA) по Краскелу—Уоллису (см. раздел 8.4.2);
    - критерий  $\chi^2$  (см. раздел 10.2)
  - сравнение трех связанных групп и более;
    - анализ вариаций (ANOVA) по Фридмену (см. раздел 8.5);
    - критерий  $\chi^2$  (см. раздел 10.2).

Как можно заметить, для порядковых признаков во многих случаях пригодны те же методы, что и для номинальных признаков. Однако помимо них для анализа порядковых данных можно также применять и непараметрические методы анализа, которые используются и при анализе количественных данных. В связи с этим мы привели выше отсылки к описанию соответствующих непараметрических методов анализа в других главах и разделах настоящего издания.

Таким образом, в данной главе мы рассмотрим наиболее универсальный метод сопоставления групп по качественным признакам. В ППП *STATISTICA* реализованы два способа применения критерия  $\chi^2$ :

- сравнение наблюдаемых и ожидаемых частот (анализ одной группы);
- сравнение двух групп.

Рассмотрим их последовательно.

## 10.1. Сравнение наблюдаемых и ожидаемых частот (анализ одной группы)

**Задача:** сравнить распределение частот значений номинального или порядкового признака с ожидаемым (на основании априорных представлений) распределений частот.

*Пример:* известно, что в популяции распределение генотипов I, II и III следующее: генотип I — 10%, генотип II — 60%, генотип III — 30%. В исследуемой выборке (300 человек, проживающих на изолированной территории) наблюдается следующее распределение генотипов — у 54 человек (18%), у 171 человека (57%) и у 75 человек (25%) соответственно. Требуется исследовать, отличается ли выборочное распределение генотипов от популяционного, или наблюдаемые колебания случайны.

**Решение:** применение критерия  $\chi^2$  для сравнения наблюдаемой и ожидаемой частот. Для реализации этого способа сначала следует сформировать следующую таблицу (рис. 10.1). Ожидаемые абсолютные частоты являются произведениями ожидаемых относительных частот на объем выборки.

В нашем примере ожидаемые частоты вычисляются следующим образом:

Генотип	Наблюдаемые абсолютные частоты	Ожидаемые абсолютные частоты
I	54	30
II	171	180
III	75	90

Рис. 10.1. Таблица наблюдаемых и ожидаемых частот бинарного признака (пример).

- для генотипа I:  $0,10 \times 300 = 30$ ;
- для генотипа II:  $0,60 \times 300 = 180$ ;
- для генотипа III:  $0,30 \times 300 = 90$ .

Данные таблицы наблюдаемых и ожидаемых частот вносят в отдельный файл данных, а затем проводят расчеты в следующем модуле:

#### STATISTICA:

- Модуль “Непараметрическая статистика”
- Раздел “Наблюдаемые частоты в сравнении с ожидаемыми” (“Observed versus expected XI”) (рис. 10.2)

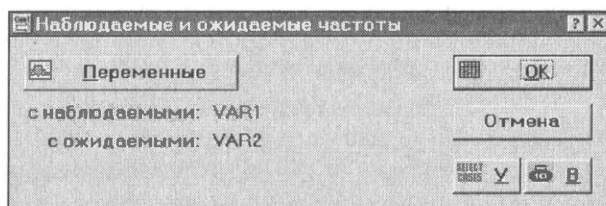


Рис. 10.2. Диалоговое окно процедуры “Наблюдаемые частоты в сравнении с ожидаемыми” модуля “Непараметрическая статистика”.

Нажав на кнопку “Переменные” (“Variables”) в диалоговом окне (см. рис. 10.2), необходимо указать, в котором из столбцов файла данных располагаются наблюдаемые частоты, а в котором — ожидаемые. После выполнения расчетов появится окно результатов (рис. 10.3). Здесь надо обратить внимание на рассчитанное значение  $p$ , располагающееся справа в верхней части окна.

Observed vs. Expected Frequencies (new.sta)				
Continue..		Chi-Square = 22.15000 df = 2 p < .000016		
Case		observed VAR1	expected VAR2	O - E /E
C:	1	54.0000	30.0000	24.0000 19.20000
C:	2	171.0000	180.0000	-9.0000 .45000
C:	3	75.0000	90.0000	-15.0000 2.50000
	Sum	300.0000	300.0000	0.0000 22.15000

Рис. 10.3. Окно результатов процедуры “Наблюдаемые частоты в сравнении с ожидаемыми” модуля “Непараметрическая статистика”.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза о совпадении наблюдаемой и ожидаемой частот не отклоняется.
- Если  $p < 0,05$ , то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о различии между наблюдаемой и ожидаемой частотами.

## 10.2. Сравнение двух групп и более

**Задача:** сравнить две группы и более по качественному признаку.

**Пример:** сравнить распространенность больных с I, II, III стадиями гипертонической болезни (ГБ) в группах больных инфарктом миокарда и инсультом.

**Решение:** задача решается в два этапа:

- 1 — построение таблицы сопряженности;
- 2 — анализ таблицы сопряженности на предмет проверки гипотезы о случайности распределения частот в таблице сопряженности.

Рассмотрим эти этапы последовательно.

**Этап 1-й — построение таблицы сопряженности.** На данном этапе решения задачи необходимо узнать, сколько объектов исследования в каждой группе имеют определенное значение изучаемого качественного признака, т.е. подсчитать абсолютные частоты для каждого из сочетаний значений двух признаков —

группирующего и анализируемого. Таблица, содержащая такие абсолютные частоты, называется таблицей сопряженности, или таблицей кросстабуляции.

**Таблица сопряженности** — это таблица, содержащая абсолютные частоты для всех возможных сочетаний взаимоисключающих значений анализируемых признаков. В принципе таблица сопряженности может содержать различное количество строк и столбцов. При этом строки таблицы соответствуют значениям первого признака, а столбцы — значениям второго признака<sup>1</sup>.

Таблица с двумя строками и двумя столбцами (так называемая таблица  $2 \times 2$ ) является частным случаем таблицы сопряженности и называется *четырёхпольной таблицей* (так как содержит четыре “поля”-ячейки).

Подчеркнем, что в ячейках таблицы сопряженности могут находиться лишь абсолютные частоты, т.е. количества объектов исследования, но не относительные частоты (доли, проценты), средние значения и т.д.

Алгоритм, с помощью которого возможны построение и дальнейший анализ таблицы сопряженности, реализован в ППП **STATISTICA** в следующем модуле.

### **STATISTICA:**

- Модуль “Основные статистики и таблицы”  
 (“Basic statistics /Tables”)
- Подмодуль “Таблицы и заголовки”  
 (“Tables and Banners”) (рис. 10.4)

Для построения двумерной (прямоугольной) таблицы сопряженности в открывшемся первом диалоговом окне (см. рис. 10.4) сначала необходимо активизировать процедуру “Задать таблицы” (“Specify tables”). При этом откроется второе диалоговое окно (рис. 10.5) для задания признаков для строк и столбцов, по которым необходимо провести анализ. Как можно видеть, в ППП предусмотрена возможность указания до 6 признаков для построения таблиц сопряженности, однако рассмотрение таких таб-

<sup>1</sup> Таблица сопряженности может быть не только двумерной (при анализе двух признаков), но и трех-, четырехмерной и т.д. (если анализируются одновременно три, четыре и т.д. признака соответственно). Рассмотрение сложных статистических методов анализа таких таблиц не входит в содержание данного издания.

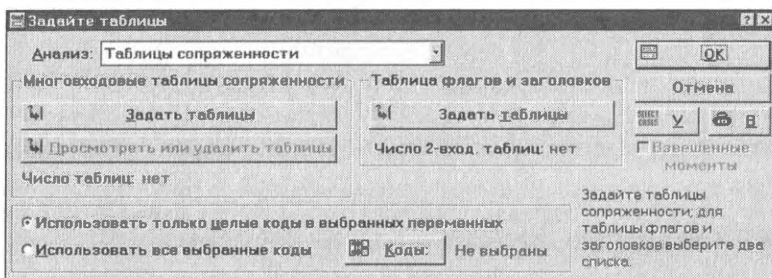


Рис. 10.4. Первое диалоговое окно анализа таблиц сопряженности.

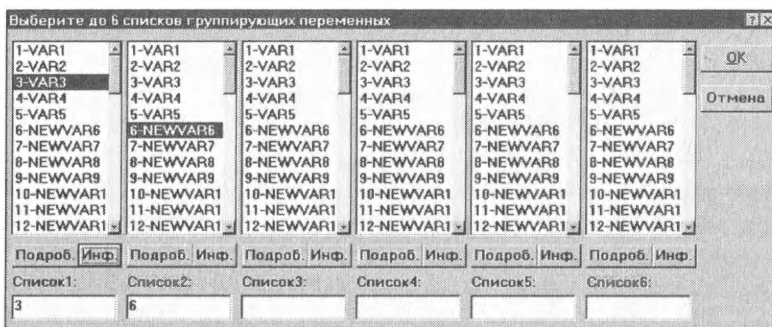


Рис. 10.5. Второе диалоговое окно (окно выбора признаков) анализа таблиц сопряженности.

лиц, как уже отмечалось ранее, не входит в наши задачи. Поэтому далее мы рассматриваем только двумерные таблицы сопряженности, т.е. таблицы, строки которого соответствуют значениям одного признака, а столбцы — значениям другого признака (например, группирующего).

Если необходимо сравнить группы по нескольким признакам, то можно указать сразу несколько признаков в каком-либо из двух списков. Однако таблицы для каждой пары признаков будут выводиться в отдельных окнах по одной.

В нашем примере исследователя интересует распределение числа объектов исследования по группам (первый фактор) и стадиям ГБ (второй фактор). В этом случае первым признаком при построении таблицы сопряженности является код группы, а вторым — стадия ГБ. В другом случае такой парой признаков могут быть, например, метод лечения и исход заболевания.

Указав признаки, необходимо в третьем диалоговом окне (рис. 10.6) обратиться к опции “Просмотреть итоговые таблицы” (“Review summary tables”).

Итоговой таблицей является таблица абсолютных частот — таблица сопряженности (рис. 10.7), строки которой соответствуют значениям первого фактора (признака), а столбцы — значениям второго фактора (признака). В ячейках таблицы находятся значения абсолютных частот (количества объектов иссле-

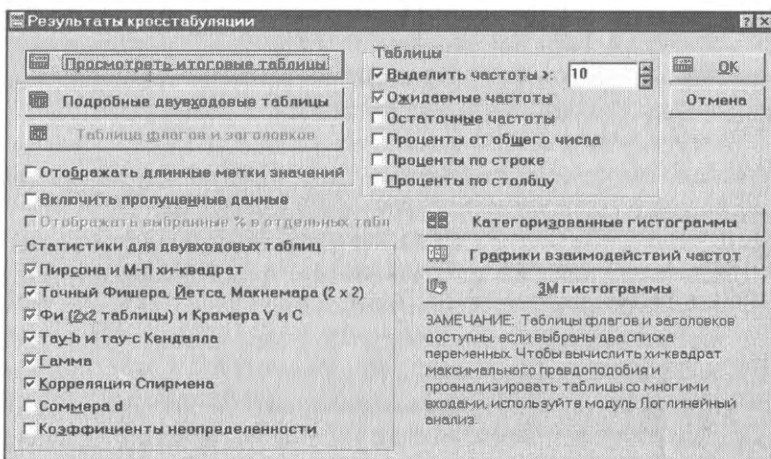


Рис. 10.6. Третье диалоговое окно анализа таблицы сопряженности.

NEWVAR6	NEWVAR7 G_1:1	NEWVAR7 G_2:2	NEWVAR7 G_3:3	Всего по стр.
G_1:0	1	0	0	1
G_2:1	2	2	2	6
G_3:2	5	0	2	7
G_4:3	1	1	1	3
Всего	9	3	5	17

Рис. 10.7. Окно результатов подсчета частот (таблица сопряженности) в процедуре “Таблицы и заголовки” (модуль “Основные статистики и таблицы”).



дования) соответственно для каждого сочетания значений первого и второго факторов.

Вернувшись в третье диалоговое окно (см. рис. 10.6), можно приступить к собственно статистическому анализу таблицы сопряженности, т.е. к проверке статистических гипотез.

**Этап 2-й — анализ таблицы сопряженности.** После построения таблицы сопряженности у исследователя естественно возникает вопрос о том, существует ли статистическая взаимосвязь двух анализируемых признаков, т.е. влияет ли один признак (например, метод лечения) на распределение числа объектов исследования по другому признаку (например, исходам заболевания), и является ли это влияние статистически значимым, т.е. ассоциированы ли эти два признака.

Для решения такой задачи можно сформулировать и проверить нулевую статистическую гипотезу. В данном случае нулевая гипотеза может быть сформулирована по-разному:

- *гипотеза об однородности* — группы (детерминируемые значениями одного из двух анализируемых признаков) происходят из одной и той же популяции, т.е. идентичны в отношении другого признака;
- *гипотеза о независимости признаков* — распределение по одному признаку не влияет на распределение по другому признаку.

Соответственно альтернативными гипотезами являются следующие:

- группы (выборки) получены из разных популяций, т.е. различаются по второму признаку;
- признаки являются зависимыми.

Для проверки любой из нулевых гипотез по таблице сопряженности в ППП **STATISTICA** предлагается использовать следующие методы:

- “Пирсона и МП<sup>1</sup> хи-квадрат” (“Pearson and ML Chi-square”) — применяются для любых двумерных (прямоугольных) таблиц;
- “Точный Фишера<sup>2</sup>, Йетса<sup>3</sup>, МакНемара<sup>4</sup> (таблицы 2×2)” (“Fisher exact, Yates, McNemar (2×2 tables)”) — применя-

---

<sup>1</sup> Максимального правдоподобия.

<sup>2</sup> Точный критерий Фишера.

<sup>3</sup> Хи-квадрат с поправкой Йетса.

<sup>4</sup> Критерий МакНемара.

ются для анализа четырехпольных таблиц, т.е. для анализа двух признаков, каждый из которых является бинарным (подробнее см. главу 11).

Возможно также вычисление следующих мер сопряженности признаков:

- “Фи (таблицы  $2 \times 2$ ) и Крамера  $V$  и  $C$ ”<sup>1</sup> — коэффициенты сопряженности, которые применяются для анализа четырехпольных таблиц, т.е. двух признаков, каждый из которых является бинарным (подробнее см. главу 11).
- “Тау- $b$  и тау- $c$  Кендалла”<sup>2</sup> — коэффициенты сопряженности (непараметрической, или ранговой, корреляции; см. главу 12).
- “Гамма”<sup>3</sup> — коэффициент сопряженности (непараметрической, или ранговой, корреляции; см. главу 12).
- “Корреляция Спирмена”<sup>4</sup> — коэффициент сопряженности (непараметрической, или ранговой, корреляции; см. главу 12).
- “Соммера  $d$ ”<sup>5</sup> — несимметричная мера связи между двумя признаками.
- “Коэффициенты неопределенности”<sup>6</sup> измеряют информационную связь между факторами (строками и столбцами таблицы). Статистика  $S(Y, X)$  является симметричной и измеряет количество информации в признаке  $Y$  относительно признака  $X$  или в признаке  $X$  относительно признака  $Y$ . Эта величина равна нулю, если  $X$  и  $Y$  независимы. Статистики  $S(X|Y)$  и  $S(Y|X)$  выражают направленную зависимость [24].

Мы остановимся в данной главе лишь на первом и наиболее универсальном из перечисленных методов — критерии  $\chi^2$ .

Если таблица сопряженности была получена с использованием подмодуля “Таблицы и заголовки” (“Tables and Banners”), то для проверки гипотезы о сопряженности (зависимости) признаков (т.е. о том, влияет ли распределение по одному признаку на распределение по другому признаку, например, влияет ли метод лечения на исходы заболевания) можно применить критерий  $\chi^2$ . Для этого в диалоговом окне “Результаты кросстабуля-

<sup>1</sup> Phi ( $2 \times 2$  tables) and Cramer's V&C (англ.).

<sup>2</sup> Kendall's tau-b & tau-c (англ.).

<sup>3</sup> Gamma (англ.).

<sup>4</sup> Spearman rank order correlation (англ.).

<sup>5</sup> Sommer's d (англ.).

<sup>6</sup> Uncertainty coefficients (англ.).

ции” (“Crosstabulation tables results”) (см. рис. 10.6) следует отметить процедуру “Пирсона и МП хи-квадрат” (“Pearson and ML Chi-square”).

Необходимо подчеркнуть, что при намерении использовать любой из вариантов критерия  $\chi^2$  необходимо предварительно убедиться в том, что результаты применения этого критерия (в частности, значение  $p$ ) будут правильными (несмещенными). Правильность результатов обеспечивается только в случае, если частоты во всех ячейках таблицы ожидаемых частот больше или равны 5. Следовательно, в каждом случае необходимо строить также и таблицу ожидаемых частот. Для построения такой таблицы следует отметить опцию “Ожидаемые частоты” (“Expected frequencies”) в верхней правой части диалогового окна “Результаты кросстабуляции” (“Crosstabulation Tables Results”) (см. рис. 10.6).

Заметим, что в обсуждаемом диалоговом окне также представлены удобные опции для расчета процентов по строкам, столбцам или тем и другим одновременно.

Итак, отметив все необходимые опции (как минимум опции “Ожидаемые частоты” и “Пирсона и М-П хи-квадрат”) в диалоговом окне (см. рис. 10.6), следует нажать кнопку “Подробные двухвходовые таблицы” (“Detailed two-way tables”). Результаты расчетов выводятся в три отдельных окна:

- окно таблицы наблюдаемых частот — таблицы сопряженности (то же, что и на рис. 10.7);
- окно таблицы ожидаемых частот (рис. 10.8);
- окно собственно результатов применения критерия  $\chi^2$  в двух модификациях — метод Пирсона и метод максимального

2-входовая итоговая: Ожидаемые частоты (new.sta)				
Частоты выделенных ячеек > 10				
ОСНОВНЫЕ СТАТИСТ.	NEWVAR7 G 1:1	NEWVAR7 G 2:2	NEWVAR7 G 3:3	Всего по стр.
NEWVAR6				
G 1:0	.529412	.176471	.294118	1.00000
G 1:1	3.176471	1.058824	1.764706	6.00000
G 1:2	3.705882	1.235294	2.058824	7.00000
G 1:3	1.588235	.529412	.882353	3.00000
Всего	9.000000	3.000000	5.000000	17.00000

Рис. 10.8. Таблица ожидаемых частот, выводимая при анализе таблиц сопряженности.

правдоподобия (рис. 10.9). Эти две модификации дают близкие результаты.

Далее...	хи-квадрат	Ст. св.	p
Хи-квадрат Пирсона	4.533333	df=6	p=.60490
N*П хи-квадрат	5.942363	df=6	p=.42968
Фи	.5163978		
Козфф. сопряженности	.4588315		
Крамера V	.3651484		
Тау b и c Кендалла	b=.0217713	c=.0207612	
Гамма	.0322581		
Корр. Спирмена	.0229862	t=.08905	p=.93022

Рис. 10.9. Окно результатов проверки статистических гипотез в процедуре “Таблицы и заголовки” (модуль “Основные статистики и таблицы”).

**Интерпретация результатов** проводится в два этапа. Сначала анализируется таблица ожидаемых частот. Если во всех ячейках этой таблицы частоты не меньше 5, то в дальнейшем можно с достаточной степенью уверенности интерпретировать результаты применения критерия  $\chi^2$ . Некоторые статистики считают критерий устойчивым, если ожидаемые частоты во всех ячейках хотя бы не нулевые, а доля ячеек, где ожидаемые частоты меньше 5, не превышает 20%. Если и это более мягкое условие не выполняется, рекомендуется превращать таблицу в четырехпольную путем объединения столбцов и/или строк, а затем использовать, например, точный критерий Фишера (см. главу 11). В случае, если это невозможно, следует обязательно при представлении результатов статистического анализа указывать на возможную неустойчивость критерия  $\chi^2$ .

На втором этапе интерпретируются результаты применения критерия  $\chi^2$ , с помощью которого проверяется нулевая статистическая гипотеза в той или иной описанной выше формулировке. Рассчитанные значения  $p$  для каждого из вариантов метода выводятся в соответствующий столбец таблицы результатов:

- при  $p > 0,05$  нулевая гипотеза об однородности (т.е. об отсутствии различий между группами) либо нулевая гипотеза о независимости признаков не отклоняются;

- при  $p < 0,05$  нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о различии групп или об ассоциированности признаков.

**Замечание.** Если результаты расчетов по методу Пирсона и по методу максимального правдоподобия дали величины  $p$ , находящиеся по разные стороны от пороговой величины уровня значимости 0,05, то можно воспользоваться тем, который “лучше” (соответствует меньшему значению  $p$ ). Дальнейший анализ этой ситуации может быть проведен лишь высококвалифицированным статистиком.

**Представление результатов.** Привести следующую информацию:

- таблицу сопряженности;
- название статистического метода и его варианта (Пирсона или МП);
- точное значение  $p$ .

---

## Глава 11. Сравнение групп по качественному бинарному признаку

Бинарные (дихотомические) признаки наиболее часто анализируются в эпидемиологических исследованиях и клинических испытаниях. Обычно они представляют собой альтернативные клинические исходы для объектов исследования, например: улучшение — без улучшения; заболел — не заболел; есть осложнение — нет осложнения.

Именно на анализе истинных клинических исходов построены все современные методики клинических испытаний, соответствующих принципам доказательной медицины. Изучение так называемых суррогатных, или промежуточных, исходов (результатов лабораторных и инструментальных исследований) в настоящее время не может считаться доказательством наличия или отсутствия клинической эффективности препаратов. В клинических испытаниях их рассматривают лишь в качестве косвенных критериев оценки эффективности вмешательства.

В данной главе рассмотрено несколько наиболее широко распространенных задач анализа качественных бинарных данных:

- сравнение выборочной относительной частоты с популяционной (см. раздел 11.1);
- сравнение двух относительных частот внутри одной и в двух группах (см. раздел 11.2);
- сравнение частот бинарного признака в двух несвязанных группах (см. раздел 11.3);
- сравнение частот бинарного признака в двух связанных группах (см. раздел 11.4);
- сравнение частот бинарного признака в трех и более несвязанных группах (см. раздел 11.5).

Особый случай сравнения бинарных признаков, являющихся диагностическими или экспертными заключениями, рассмотрен в разделе 15.5.2.

## 11.1. Сравнение выборочной относительной частоты с популяционной

**Задача:** сравнить относительную частоту  $P$  какого-либо явления, наблюдаемую в изучаемой выборке, с известным (или ожидаемым) популяционным значением относительной частоты этого события  $P_p^1$ .

**Пример:** наблюдаемая относительная частота случаев атопического дерматита у 500 обследованных учеников школы № 7315 г. Н-ска равна 11%. Известно, что в среднем по России относительная частота этого заболевания среди школьников составляет 6% (95% ДИ — [4%; 8%]). Необходимо узнать, действительно ли в этой школе данное заболевание встречается чаще, чем в общей популяции?

**Решение.** Могут использоваться следующие подходы:

- сравнение ДИ для выборочной относительной частоты (пропорции, доли) с популяционным значением (см. раздел 11.1.1);
- методы проверки гипотез о совпадении наблюдаемых и ожидаемых частот;
- $z$ -критерий (см. раздел 11.1.2);
- критерий  $\chi^2$  (см. раздел 11.1.3).

### 11.1.1. Сравнение доверительного интервала для выборочной относительной частоты с популяционной

**Решение.** Построение ДИ для относительной частоты значения бинарного признака описано в разделе 9.3.

В отношении популяционного значения могут рассматриваться две ситуации:

- известна точечная оценка популяционной частоты, т.е. одно число (в нашем примере — 6%);
- известна интервальная оценка популяционной частоты, т.е. ДИ (в нашем примере — [4%; 8%]).

Решение заключается соответственно либо в рассмотрении, попадает ли популяционное значение в границы вычисленного

<sup>1</sup> Prevalence (англ.) — распространенность; в отечественной литературе иногда также используется термин “преваленс”.

ДИ (в первой ситуации), либо в рассмотрении, перекрываются ли доверительные интервалы для выборочного и популяционного значений.

**Интерпретация результатов.** Если точечная оценка популяционного значения находится вне рассчитанного по выборке ДИ, то можно считать, что выборка статистически значимо (с уровнем значимости 0,05 в случае сравнения с 95% ДИ) отличается от популяции по значению изучаемого бинарного признака (пропорции). Если же популяционное значение попадает внутрь ДИ, то выборка не отличается от популяции.

Аналогично в случае наличия ДИ для популяционного значения. Если доверительные интервалы выборочной и популяционной величин перекрываются, то различия статистически незначимы. Если не перекрываются, то можно считать выборку статистически значимо отличающейся от популяции.

### 11.1.2. Проверка гипотез (z-критерий)

Проверка указанной нулевой гипотезы о совпадении наблюдаемой (выборочной) и ожидаемой (популяционной) частот может быть осуществлена следующим способом:

1) необходимо вычислить стандартную ошибку  $m_p$  для предполагаемого (ожидаемого) популяционного значения в данной выборке по формуле:

$$m_p = \sqrt{\frac{P_p(1-P_p)}{n}},$$

где  $n$  — число наблюдений (объектов исследования) в выборке;  $P_p$  — относительная частота (распространенность) изучаемого состояния в популяции;

2) вычислить значение z-статистики критерия (с учетом поправки на аппроксимацию биномиального распределения нормальным) по формуле:

$$z = \frac{|P - P_p| - \frac{1}{2n}}{m_p},$$

где  $P$  — относительная частота, наблюдаемая в выборке;

3) воспользоваться “Вероятностным калькулятором” (рис. 11.1) для вычисления значения  $p$  по значению  $z$  в целях проверки нулевой гипотезы о том, что наблюдаемая в выборке относительная частота и популяционная относительная частота не различаются.



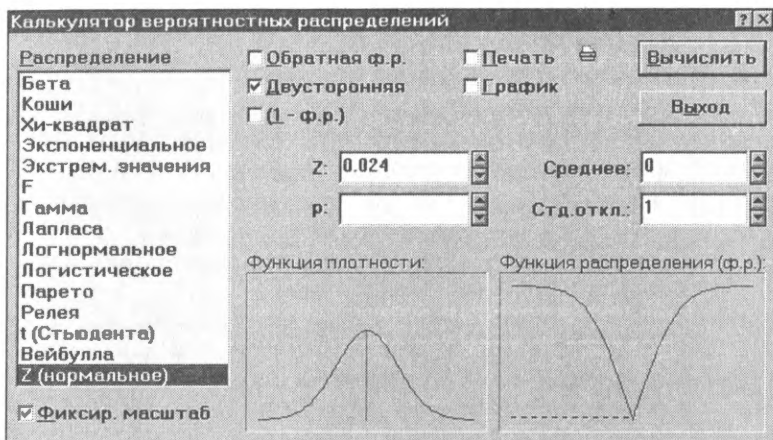


Рис. 11.1. Диалоговое окно “Вероятностного калькулятора” (модуль “Основные статистики и таблицы”).

### STATISTICA:

- Модуль “Основные статистики и таблицы” (“Basic statistics/Tables”)
  - Подмодуль “Вероятностный калькулятор” (“Probability calculator”)
    - “z (нормальное)” (“Normal distribution (z)”, “Compute p from z”)

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза о совпадении относительных частот не отклоняется, т.е. можно считать, что наблюдаемая относительная частота соответствует популяционной (ожидаемой).
- Если  $p < 0,05$ , то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о том, что различия между выборочной и популяционной относительными частотами статистически значимы.

**Представление результатов.** Привести следующую информацию:

- значение популяционной относительной частоты и ее ДИ;

- значение выборочной относительной частоты и ее ДИ;
- точное значение  $p$ , полученное при проверке нулевой гипотезы.

### 11.1.3. Проверка гипотез (критерий $\chi^2$ )

Другой способ проверки гипотез о совпадении наблюдаемой и ожидаемой частот значений бинарного признака — вычисление критерия  $\chi^2$ . Для реализации этого способа сначала следует сформировать следующую таблицу (рис. 11.2). Ожидаемая абсолютная частота события  $P$  является произведением ожидаемой относительной частоты события  $P_p$  (в нашем примере  $6\%=0,06$ ) на объем выборки (в нашем примере 500).

$$P = N \cdot P_p = 0,06 \cdot 500 = 30$$

	Наблюдаемые абсолютные частоты	Ожидаемые абсолютные частоты
Изучаемое событие (наличие изучаемого исхода — заболевания, смерти, осложнения и т.д.)	55	30
Отсутствие изучаемого события	$500 - 55 = 445$	$500 - 30 = 470$

Рис. 11.2. Таблица наблюдаемых и ожидаемых частот бинарного признака (пример).

Данные таблицы наблюдаемых и ожидаемых частот вносят в отдельный файл данных, а затем проводят расчеты в следующем модуле.

#### STATISTICA:

- Модуль “Непараметрическая статистика”
- Раздел “Наблюдаемые частоты в сравнении с ожидаемыми” (“Observed versus expected XI”) (рис. 11.3)

В диалоговом окне (см. рис. 11.3) после нажатия на кнопку “Переменные” (“Variables”) необходимо указать, в котором из столбцов файла данных располагаются наблюдаемые частоты, а в котором — ожидаемые. После выполнения расчетов появится окно результатов (рис. 11.4). Здесь надо обратить внимание на рассчитанное значение  $p$ , располагающееся справа в верхней части окна.

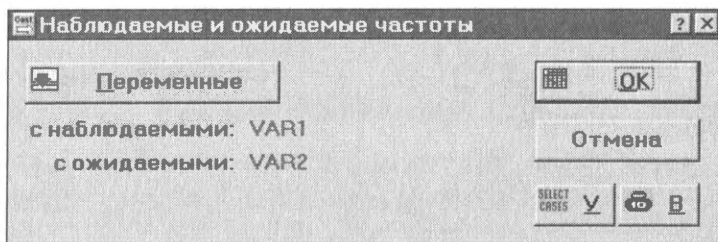


Рис. 11.3. Диалоговое окно процедуры “Наблюдаемые частоты в сравнении с ожидаемыми” модуля “Непараметрическая статистика”.

Наблюдаемые - Ожидаемые Частоты (new...)				
Далее...		хи-квадрат = 22.16312 сс = 1 p < .000003		
Наблюд.	наблюд. VAR1	ожд. VAR2	Н - О	(Н-О) **2 /О
С: 1	55.0000	30.0000	25.0000	20.83333
С: 2	445.0000	470.0000	-25.0000	1.32979
Сумма	500.0000	500.0000	0.0000	22.16312

Рис. 11.4. Окно результатов процедуры “Наблюдаемые частоты в сравнении с ожидаемыми” модуля “Непараметрическая статистика”.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза о совпадении наблюдаемой и ожидаемой частот не отклоняется.
- Если  $p < 0,05$ , то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о различии между наблюдаемой и ожидаемой частотами.

## 11.2. Сравнение относительных частот внутри одной группы и в двух группах

По результатам подсчета абсолютных и относительных частот и визуального анализа таблицы частот для качественного признака в одной или двух группах у исследователя могут возникнуть следующие предположения:

- предположение о том, что доля объектов с каким-либо значением качественного признака больше доли объектов с другим значением этого же признака в этой же группе;
- предположение о том, что доли объектов с каким-либо значением качественного признака в двух группах различаются.

Сравнения относительных частот какого-либо признака “на глаз” недостаточно. Оба таких предположения следует проверять с использованием статистических методов. Соответственно могут быть сформулированы следующие задачи.

**Задача 1:** изучить преобладание какого-либо значения качественного признака (относительной частоты) в выборке.

*Пример 1:* необходимо исследовать, преобладают ли в изучаемой выборке больных с инсультом лица с повышенным уровнем холестерина в сыворотке крови.

**Задача 2:** сравнить две группы по относительной частоте значений качественного (номинального или порядкового) признака.

*Пример 2:* необходимо выяснить, различаются ли больные с ишемическим и геморрагическим инсультом по наличию отягощенной наследственности.

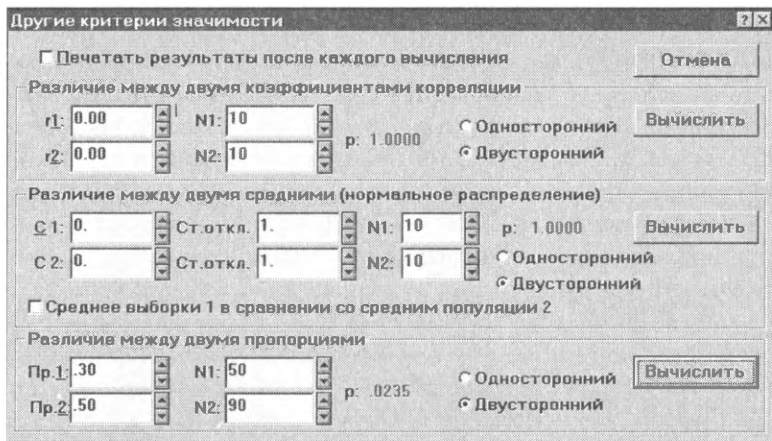
**Решение:** для решения таких задач недостаточно подсчитать долю больных (например, 60%) с повышенным уровнем холестерина (даже если они составляют большинство в выборке) и отметить, что эта доля больше доли больных с нормальным уровнем холестерина (например, 40%). Необходимо исследовать, случаен ли этот результат, т.е. будет ли такой же результат наблюдаться, например, в 95 из 100 других выборок из изучаемой популяции. В данном случае необходимо показать, что 60% статистически значимо больше, чем 40%. Понятно, что если такой результат достигнут в выборке из 100 больных, то он более убедителен, чем если он достигнут в выборке из 10 больных (в последнем случае выше доля случайности). Для заключения о преобладании того или иного значения признака необходимо использовать один из двух следующих способов.

**Способ I** — сравнение ДИ относительных частот. Этот способ очень прост. Для этого необходимо вычислить границы ДИ для каждой из сравниваемых относительных частот (см. раздел 9.3) и сравнить эти ДИ. Если они не перекрываются, то различия частот можно считать статистически значимыми (с уровнем значимости 0,05, если анализировались 95% ДИ). Соответственно, если интервалы перекрываются, то различия статистически незначимы.

**Способ II** — проверка нулевой статистической гипотезы о равенстве относительных частот в двух популяциях. Для реализации этого способа необходимо обратиться к разделу “Вероятностный калькулятор” (“Probability calculator”). В данном случае проверяется нулевая статистическая гипотеза об отсутствии различий между относительными частотами в одной группе (задача 1) либо в двух группах (задача 2). Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий между группами.

**STATISTICA:**

- Модуль “Основные статистики и таблицы” (“Basic statistics/Tables”)
  - Подмодуль “Другие критерии значимости” (“Other significance tests”) (в версиях ППП 5.\*) (рис. 11.5) или подмодуль “Probability calculator” — кнопка “More” (в версиях ППП 4.\*)
    - Процедура “Различие между двумя пропорциями” (“Difference between two percentages”)



**Рис. 11.5.** Диалоговое окно процедуры “Другие критерии значимости” (в модуле “Основные статистики и таблицы”).

В строки окна необходимо ввести относительные частоты объектов исследования с изучаемым значением признака (не в про-

центах, а в десятичных долях единицы) и общее число объектов исследования в группах (принятое за 100%). В задаче 1 число объектов исследования в группе одно и то же для первой и второй строк.

В данном разделе существует возможность выбора одностороннего или двустороннего критерия статистической значимости. Односторонний критерий статистической значимости следует использовать лишь в тех случаях, когда Вам заранее (априори) известно, какая из сопоставляемых подгрупп преобладает по частоте, и лишь необходимо вычислить значение  $p$  для того, чтобы подтвердить статистическую значимость исходного предположения. Однако в этом случае такая априорная гипотеза должна быть обоснована. В подавляющем большинстве ситуаций такое обоснование невозможно, поэтому необходимо пользоваться двусторонним критерием статистической значимости. Данный критерий выдает значение  $p$ , превышающее примерно в 2 раза такое для одностороннего критерия.

**Замечание 1.** Если исследуемый признак изучался не во всей группе объектов исследования, а только у части из них, то за 100% должно быть принято то число больных, у которых этот признак изучен, а не общий объем группы.

**Замечание 2.** Может возникнуть ситуация, при которой изучаемый признак в одной из сравниваемых групп встречался у всех (100%) или ни у одного (0%) объекта исследования. В этой ситуации в диалоговое окно следует вводить не число 1,00 (относительная частота в случае 100%) а число, близкое к 1, т.е. 0,9999999. Аналогично вместо 0 следует вводить число, близкое к 0, т.е. 0,0000001. Такая замена практически не влияет на результат вычислений.

После ввода исходных данных необходимо нажать кнопку “Вычислить”, а затем прочесть вычисленное значение “ $p$ ”.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза об отсутствии различий между относительными частотами значений признака не отклоняется.
- Если  $p < 0,05$ , то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о существовании различий между относительными частотами значений признака.

**Рекомендуемое представление результатов:**

- привести число наблюдений в группе (группах);
- привести относительные частоты;
- привести точное значение  $p$ .

### **11.3. Сравнение частот бинарного признака в двух несвязанных (независимых) группах (анализ таблиц 2x2)**

Бинарным признаком, сравнение групп по которому обычно требуется в медицинских исследованиях, является клинический исход. Исходом может быть улучшение или ухудшение состояния, возникновение осложнений, выздоровление, смерть и т.д. Исходы чаще всего изучаются в клинических исследованиях при исследовании методов лечения и в эпидемиологических исследованиях при исследовании факторов риска.

Исход обычно является бинарным признаком, т.е. имеет два возможных значения — “да” или “нет” (например, наличие заболевания или его отсутствие, наличие осложнений или их отсутствие и т.д.).

**Задача:** сравнить группы по частоте значения какого-либо бинарного признака (симптома, синдрома, исхода).

*Пример:* исследованы две группы, в одной из которых больных ( $n=200$ ) лечили с использованием активного препарата, а в другой ( $n=500$ ) — давали плацебо. Необходимо сопоставить группы по частоте улучшения, которое отмечено в 100 из 500 (20%) случаев в группе плацебо и в 80 из 200 (40%) случаев в группе активного лечения.

**Решение:** для сравнения групп по бинарному признаку необходимо построить таблицу сопряженности, содержащую частоты для взаимоисключающих значений изучаемого бинарного признака в каждой из групп. Таблица сопряженности обычно организуется следующим образом:

— в строках описаны группы, сформированные в зависимости от используемых методов лечения, факторов риска и т.д. При этом подразумевается, что один объект исследования может быть отнесен только к одной из этих групп;

— в столбцах описаны возможные исходы. Подразумевается, что для каждого объекта исследования возможен только один из исходов (исходы являются взаимоисключающими).

Как уже отмечалось в главе 10, в принципе таблица сопряженности может содержать различное количество строк и столбцов. Таблица с 2 строками и 2 столбцами (так называемая таблица  $2 \times 2$ ) является частным случаем таблицы сопряженности. Подчеркнем, что в ячейках таблицы сопряженности должны находиться абсолютные частоты, т.е. количества объектов исследования, но не доли, проценты, средние значения и т.д.

Таблица сопряженности  $2 \times 2$  в общем виде приведена на рис. 11.6.

Исследуемая группа	Болезнь/Исход/ Эффект <i>есть</i>	Болезни/Исхода/ Эффекта <i>нет</i>
Вмешательства	<i>A</i>	<i>B</i>
Контроля	<i>C</i>	<i>D</i>

Рис. 11.6. Таблица сопряженности для сравнения групп по бинарному признаку.

Часто в клинических и эпидемиологических исследованиях при анализе таблиц сопряженности  $2 \times 2$  используются следующие понятия:

1) **абсолютный риск (АР)** — это относительная частота изучаемого события в определенной группе; измеряется в долях единицы или процентах.

$AP_B = A / (A + B)$  — абсолютный риск в группе вмешательства;

$AP_K = C / (C + D)$  — абсолютный риск в группе контроля.

2) разность относительных частот событий (исходов, эффектов и т.п.) в двух группах. В случае исследования факторов риска или эффективности вмешательства эта величина называется **изменением абсолютного риска**<sup>1</sup> (ИАР). В зависимости от того, снижается или повышается АР в группе вмешательства по отношению к группе контроля, говорят о снижении или повышении АР (САР или ПАР соответственно):

$$САР \text{ (или ПАР)} = |A / (A + B) - C / (C + D)|$$

<sup>1</sup> Attributable risk (AR; англ.); иногда переводится как “атрибутивный риск”.



Большинство врачей традиционно считают этот показатель — разность относительных частот исходов в двух группах — основным показателем различия между применяемыми вмешательствами. Часто показатель САР называют “терапевтической пользой”.

3) отношение относительных частот (абсолютных рисков) в двух группах. В случае исследования факторов риска или эффективности методов лечения эта величина называется *относительным риском*<sup>1</sup> (ОР):

$$ОР = \frac{A/(A+B)}{C/(C+D)}$$

4) *снижение* (или *повышение*) *относительного риска* (СОР или ПОР соответственно) — отношение САР к частоте исходов в группе контроля:

$$СОР = \frac{|A/(A+B) - C/(C+D)|}{C/(C+D)}$$

5) *число больных, которых необходимо лечить исследуемым методом в течение определенного времени, чтобы достичь благоприятного эффекта или предотвратить определенный неблагоприятный исход у одного больного*<sup>2</sup> (ЧБНЛ):

$$ЧБНЛ = 1/САР$$

4) *отношение шансов*<sup>3</sup> (ОШ) — отношение шансов события в одной группе к шансам этого же события в другой группе. Вычисляется по формуле:

$$ОШ = \frac{A/B}{C/D}$$

**Важное замечание.** Вышеприведенные формулы относятся только к такой таблице сопряженности, строки и столбцы которой организованы таким образом, как показано на рис. 11.6. Изменение порядка столбцов и строк повлечет за собой и соответствующее изменение формул!

<sup>1</sup> Relative risk (RR; англ.).

<sup>2</sup> Number needed to treat (NNT; англ.).

<sup>3</sup> Odds ratio (OR; англ.).

**Замечание 1.** В случае исследования не неблагоприятных исходов (смерть, развитие заболевания и т.д.), а нейтральных или положительных (например, отсутствие осложнений) вместо термина “риск” более уместно использовать термин “вероятность”. Таким образом, исследуемыми параметрами могут быть “изменение абсолютной вероятности”, “относительная вероятность”.

**Замечание 2.** В условиях “обычной” для врачебной практики частоты событий бывает достаточно сравнивать группы по частоте исходов с использованием абсолютных рисков. Если же в популяции относительная частота события (например, заболевания) высока или низка (т.е. заболевание соответственно широко или мало распространено в популяции), то для более объективной и наглядной оценки эффективности вмешательства возникает потребность сделать поправку на распространенность изучаемого события в отсутствие исследуемого вмешательства. Всегда желательно вычислять и  $AR$ , и  $OR$ . Например, вмешательство, снижающее летальность в 2 раза в условиях низкой распространенности изучаемого заболевания или состояния (например, 2/1000), дает лишь незначительное снижение  $AR$  (1/1000). В этом случае использование для оценки эффективности вмешательства только  $OR$  недостаточно.

Таким образом, при низкой или высокой распространенности изучаемого состояния предпочтительно применение обоих показателей —  $AR$  и  $OR$  (подробнее см. [3]). Если же необходимо сравнить несколько исследований между собой, то поправка на исходную распространенность изучаемого состояния (в контрольной группе) в сравниваемых исследованиях становится совершенно необходимой.

*Пример:* у больных с ангиографически подтвержденным стенозом каротидных артерий более 70% и преходящими нарушениями мозгового кровообращения в анамнезе применение каротидной эндартерэктомии снижает  $OR$  развития тяжелого инсульта в предстоящие 5 лет почти в 2 раза. При этом  $AR$  возникновения инсульта в контрольной группе (консервативного лечения) составляет всего лишь 15%, а в группе хирургического лечения — 8%. Очевидно, что  $OR$  в этом случае не является достаточно информативным параметром, чтобы сделать выводы для клинической практики; необходимо учитывать также  $AR$  в группах.

**Замечание 3.** Понятие ОР может применяться лишь для проспективных и одномоментных исследований, в которых группы формируются на основании наличия или отсутствия фактора риска или медицинского вмешательства. Для ретроспективных исследований это понятие не имеет смысла, так как в них группы формируются на основании уже имеющегося состояния, и, следовательно, ОР не может быть оценен. Однако в ретроспективных исследованиях может быть вычислено ОШ, значение которого приближенно равно значению ОР в случае, если заболевание является относительно редким (т.е.  $A+B \approx B$  и  $C+D \approx D$ ). Подробнее об этом см. [3], с. 83.

Пример вычислений: таблица сопряженности для приведенной в начале раздела примера выглядит так, как показано на рис. 11.7. Для приведенных данных перечисленные характеристики вычисляются следующим образом:

$$CAP = |80/(80+120) - 100/(100+400)| = 0,2$$

$$OP = \frac{80/(80+120)}{100/(100+400)} = 2$$

$$COP = \frac{|80/(80+120) - 100/(100+400)|}{100/(100+400)} = 1$$

$$OШ = \frac{80/120}{100/400} = 8/3 \approx 2,7$$

Вмешательство	Результат	
	улучшение	без улучшения
Прием лекарственного препарата ( $n=200$ )	80	120
Прием плацебо ( $n=500$ )	100	400

Рис. 11.7. Таблица сопряженности (пример).

Далее мы рассмотрим способы анализа различий между группами по бинарному признаку:

- расчет ДИ для разности абсолютных рисков (для несвязанных групп);
- расчет ДИ для отношения относительных рисков (для несвязанных групп);

- расчет ДИ для ОШ (для несвязанных групп);
- проверка статистических гипотез о различии относительных частот (для несвязанных и связанных групп).

К сожалению, в ППП **STATISTICA** отсутствуют алгоритмы для вычисления указанных ДИ, но они есть в некоторых других пакетах, например EpiInfo (CDC, США)<sup>1</sup>. Вместе с тем вычисления границ ДИ крайне просты и могут быть выполнены с помощью калькулятора.

### 11.3.1. Доверительный интервал для разности относительных частот

Как уже отмечалось, разность относительных частот бинарного признака вычисляется по четырехпольной таблице с использованием следующей формулы:

$$\left| A/(A+B) - C/(C+D) \right|$$

ДИ для разности относительных частот бинарного признака в двух группах может быть вычислен, если число наблюдений (объектов исследования) в каждой из групп не менее 10, и каждый исход наблюдался не менее чем в 10 случаях.

Приведем один из способов<sup>2</sup> построения ДИ для разности относительных частот  $\Delta = P_1 - P_2$ :

1. Вычислить стандартную ошибку  $m_\Delta$  для разности относительных частот:

$$m_\Delta = \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}},$$

где  $P_1 = A/(A+B)$  — относительная частота признака в первой группе,  $n_1$  — число наблюдений (объектов исследования) в первой группе,  $P_2 = C/(C+D)$  — относительная частота признака во второй группе,  $n_2$  — число наблюдений (объектов исследования) во второй группе.

2. Вычислить границы ДИ:

$$(P_1 - P_2) - t \times m_\Delta; (P_1 - P_2) + t \times m_\Delta,$$

где  $t$  — значение  $t$ -критерия, соответствующее объему исследуемой выборки и ДК (95, 90 или 99%), обычно принимается приближенно равным 1,96 для ДК 95%.

<sup>1</sup> Пакет EpiInfo можно бесплатно получить в Интернет <http://www.cdc.gov/epo/epi/epiinfo.htm>

<sup>2</sup> Другие способы вычисления границ ДИ описаны, например, в [21 — 25].

Значение  $t$  для конкретного случая можно узнать, воспользовавшись опцией “Вероятностного калькулятора”.

### STATISTICA:

- Модуль “Основные статистики и таблицы”  
 (“Basic statistics/Tables”)
- Подмодуль “Вероятностный калькулятор”  
 (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “ $t$  Стьюдента” (“Student  $t$  test”);
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1-ф.р.”
- задать число степеней свободы “Ст. св.” (“ $df$ ”),  $df=n-1$ ;
- задать значение  $p$  (например, 0,05 для вычисления границ 95% ДИ);
- нажать кнопку “Вычислить” (“Compute”).

Искомое значение  $t$  появится в окне “ $t$ ”.

**Интерпретация результатов** обоснована только в предположении, что исследуемая выборка репрезентативна по отношению к популяции.

Если ДИ не содержит нуля, то можно с уверенностью, соответствующей ДК (например, 95% в случае 95% ДИ) утверждать, что различия относительных частот в двух группах статистически значимы. В обратном случае различия между группами по данному признаку статистически незначимы.

*Пример вычислений ДИ для AP:*

$$P_1 = 80/200 = 0,4$$

$$P_2 = 100/500 = 0,2$$

$$P_1 - P_2 = 0,4 - 0,2 = 0,2$$

$$m_{\Delta} = \sqrt{\frac{0,4(1-0,4)}{200} + \frac{0,2(1-0,2)}{500}} \approx 0,04$$

$$df = 200 + 500 - 1 = 699$$

$$t = 1,96 \text{ (для } df = 699 \text{ и } p = 0,05)$$

$$\text{Границы ДИ: } 0,2 \pm 1,96 \times 0,04 \approx [0,12; 0,28]$$

**Интерпретация результата.** ДИ не включает 0, следовательно, различия между группами по изучаемому признаку являются статистически значимыми.

### 11.3.2. Доверительный интервал для относительного риска

Как было указано выше, отношение относительных частот — относительный риск (ОР) — для бинарного признака вычисляется по четырехпольной таблице с использованием следующей формулы:

$$OP = \frac{A/(A+B)}{C/(C+D)}$$

ДИ для этой величины вычисляется (по методу Katz) следующим образом:

1. Вычисляется стандартная ошибка  $m$  для натурального логарифма ОР —  $\ln(OP)$ :

$$m = \sqrt{\frac{B/A}{B+A} + \frac{D/C}{D+C}}$$

2. Вычисляются границы ДИ для функции  $\ln(OP)$ :

$$L = \ln(OP) - t \times m;$$

$$U = \ln(OP) + t \times m,$$

где  $\ln(OP)$  — натуральный логарифм для величины ОР,  $L$  — нижняя граница ДИ для  $\ln(OP)$ ,  $U$  — верхняя граница ДИ для  $\ln(OP)$ ,  $t$  — значение  $t$ -критерия,  $m$  — стандартная ошибка для ОР.

Точное значение  $t$  можно узнать с использованием описанной в разделе 11.3.1 опции “Вероятностного калькулятора”.

3. Вычисляются границы ДИ для ОР путем возведения числа  $e^L$  в степень  $L$  и  $e^U$  соответственно:

$$e^L ; e^U.$$

**Интерпретация результатов** обоснована только в предположении, что исследуемая выборка репрезентативна по отношению к популяции.

— Если ДИ для ОР включает единицу, то различия между группами по изучаемому бинарному признаку статистически незначимы.

— Если ДИ располагается в области справа от единицы (т.е. все значения ДИ больше 1), то ОР статистически значимо повы-

<sup>1</sup>  $e$  — математическая константа, равная приблизительно 2,72.

шен в группе, соответствующей первой строке четырехпольной таблицы (группе исследуемого вмешательства) по отношению к группе, соответствующей второй строке четырехпольной таблицы (контрольной группе), т.е. имеет место повышение ОР (ПОР).

- Если ДИ располагается левее единицы (все его значения меньше 1), то ОР статистически значимо снижен в группе, соответствующей первой строке четырехпольной таблицы (группе исследуемого вмешательства) по отношению к группе, соответствующей второй строке четырехпольной таблицы (контрольной группе), т.е. имеет место снижение ОР (СОР).

Напомним, что термин “риск” подразумевает изучение каких-либо негативных исходов (т.е. заболевания, осложнения, смерти и т.п.), поэтому если наблюдается статистически значимое СОР (т.е. ОР оказывается меньше 1), то это свидетельствует об эффективности исследуемого вмешательства. Если же изучаются позитивные исходы (например, выздоровление), то термин “риск”, вообще говоря, не очень адекватен. Предпочтительно в этом случае использовать термин “вероятность”. Если все же вычисляется ОР, то при условии эффективности вмешательства он должен повышаться, и, следовательно, ПОР является свидетельством эффективности вмешательства.

*Пример вычислений ДИ для ОР:*

$$OR = \frac{80/(80+120)}{100/(100+400)} = 2$$

$$\ln(2) = 0,69$$

$$m = \sqrt{\frac{120/80}{120+80} + \frac{400/100}{400+100}} \approx 0,12$$

$$t = 1,96 \text{ (для } df = 699 \text{ и } p = 0,05)$$

$$L = 0,69 - 1,96 \times 0,12 \approx 0,45$$

$$U = 0,69 + 1,96 \times 0,12 \approx 0,93$$

$$e^L \approx 1,57$$

$$e^U \approx 2,53$$

$$\text{ДИ для ОР} = [1,57; 2,53]$$

**Интерпретация результата.** ДИ расположен правее единицы, следовательно, ОР статистически значимо выше в группе, соответствующей первой строке четырехпольной таблицы.

### 11.3.3. Доверительный интервал для отношения шансов

Как уже отмечалось ранее, отношение шансов (ОШ) вычисляется для бинарного признака по четырехпольной таблицы с использованием следующей формулы:

$$\text{ОШ} = \frac{A/B}{C/D}$$

ДИ для ОШ вычисляется (по методу Woolf) следующим образом.

1. Вычисляется стандартная ошибка  $m$  для натурального логарифма ОШ:

$$m = \sqrt{\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}}$$

2. Вычисляются границы ДИ для натурального логарифма ОШ:

$$L = \ln(\text{ОШ}) - t \times m;$$

$$U = \ln(\text{ОШ}) + t \times m,$$

где  $\ln(\text{ОШ})$  — натуральный логарифм для величины ОШ,  $L$  — нижняя граница ДИ для  $\ln(\text{ОШ})$ ,  $U$  — верхняя граница ДИ для  $\ln(\text{ОШ})$ ,  $t$  — значение  $t$ -критерия,  $m$  — стандартная ошибка для  $\ln$  ОШ.

Величину  $t$  можно узнать с использованием описанной в разделе 11.3.1 опции “Вероятностного калькулятора”.

3. Вычисляются границы ДИ для ОР путем возведения числа  $e$  в степень  $L$  и  $U$  соответственно:

$$e^L ; e^U$$

**Интерпретация результатов** обоснована только в предположении, что исследуемая выборка репрезентативна по отношению к популяции.

- Если ДИ для ОШ включает единицу, то различия между группами по изучаемому бинарному признаку статистически незначимы.
- Если ДИ располагается в области справа от единицы (все значения ДИ больше 1), то шанс развития изучаемого состояния статистически значимо выше в группе, соответствующей первой строке четырехпольной таблицы.
- Если ДИ располагается левее единицы (меньше 1), то ОШ выше в группе, соответствующей второй строке четырехпольной таблицы.



Пример вычислений:

$$\text{ОШ} = \frac{80/120}{100/400} = 2,7$$

$$m = \sqrt{\frac{1}{80} + \frac{1}{120} + \frac{1}{100} + \frac{1}{400}} \approx 0,18$$

$$\ln(2,7) \approx 0,99$$

$$t \approx 1,96$$

$$L = 0,99 - 1,96 \times 0,18 \approx 0,64$$

$$U = 0,99 + 1,96 \times 0,18 \approx 1,34$$

$$e^L \approx 1,90$$

$$e^U \approx 3,82$$

$$\text{ДИ для ОШ} — [1,90; 3,82]$$

**Интерпретация результата.** ДИ расположен правее единицы, следовательно, шанс развития изучаемого состояния статистически значимо выше в группе, соответствующей первой строке четырехпольной таблицы.

#### 11.3.4. Проверка гипотез (точный критерий Фишера, $\chi^2$ с поправкой Йетса)

Другой способ анализа различия частот в двух независимых группах объектов исследования — проверка нулевой статистической гипотезы об отсутствии различий этих величин. Для этого в ППП *STATISTICA* существуют несколько процедур.

В случае, если вычисление абсолютных частот ранее не было проведено, рекомендуем воспользоваться модулем для построения таблиц сопряженности, описанным в главе 10.

**STATISTICA:**

→ Модуль “Основные статистики и таблицы”

(“Basic statistics and tables”)

→ Процедура “Таблицы и заголовки”

(“Tables and banners”)

Зачастую же материалом для сравнения групп по бинарному признаку является не таблица исходных данных (описание каждого наблюдения, или объекта исследования, по каждому из анализируемых признаков), а уже ранее рассчитанная таблица частот (т.е. известны количества объектов исследования для каждого значения каждого из двух признаков). В этом

случае достаточно легко построить четырехпольную ( $2 \times 2$ ) таблицу (рис. 11.8).

Исследуемые группы	Исход 1	Исход 2
Группа 1	<i>A</i>	<i>B</i>
Группа 2	<i>C</i>	<i>D</i>

**Рис. 11.8. Четырехпольная таблица абсолютных частот.**

*A* — число объектов исследования в группе 1 с исходом 1, *B* — число объектов исследования в группе 1 с исходом 2, *C* — число объектов исследования в группе 2 с исходом 1, *D* — число объектов исследования в группе 2 с исходом 2.

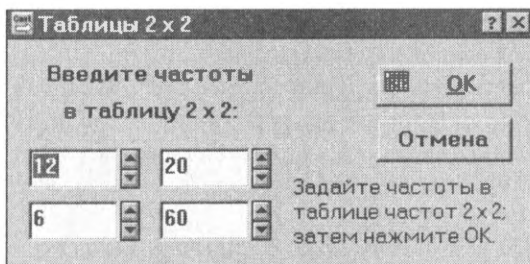
### STATISTICA:

→ Модуль “Непараметрические статистики”

(“Nonparametrics/Distrib.”)

→ Процедура “Таблицы  $2 \times 2$ : хи/ $V$ /фи, Макнемара, точный Фишера” (“ $2 \times 2$  Tables XI/VI/Phil, McNemar, Fisher exact”)

В строки диалогового окна (рис. 11.9) необходимо ввести абсолютные частоты (число объектов исследования) соответственно для каждого из двух исходов в каждой из двух групп.



**Рис. 11.9. Диалоговое окно процедуры “Таблицы  $2 \times 2$ ...” (в модуле “Непараметрические статистики”).**

После ввода исходных данных необходимо нажать кнопку “ОК”. Откроется окно результатов (рис. 11.10). Перечислим используемые при вычислениях статистические методы (за небольшими исключениями все они присутствуют в каждом из двух упомянутых модулей):

НЕПАРАМ. СТАТИСТ.	Столб. 1	Столб. 2	Сумма по строкам
Частоты, строка 1	12	20	32
Процент от общего	12.245%	20.408%	32.653%
Частоты, строка 2	6	60	66
Процент от общего	6.122%	61.224%	67.347%
Сумма по столбцу	18	80	98
Процент от общего	18.367%	81.633%	
Хи-квадрат (ст.св.=1)	11.60	p= .0007	
V-квадрат (ст.св.=1)	11.48	p= .0007	
Поправка Йетса	9.78	p= .0018	
фи коэффициент	.11837		
Фишера p, односторонний		p= .0012	
двусторонний		p= .0015	
МакНемара Хи-квадрат (A/D)	30.68	p= .0000	
Хи-квадрат (B/C)	6.50	p= .0108	

**Рис. 11.10.** Окно результатов процедуры “Таблицы 2x2: хи/V/фи, МакНемара, точный критерий Фишера” (в модуле “Непараметрические статистики”).

- “Хи-квадрат” (“Chi-square”) — классический критерий  $\chi^2$  по Пирсону;
- “V-квадрат” (“V-square”) — скорректированный критерий  $\chi^2$ ;
- “Поправка Йетса” (“Yates corrected Chi-square”) —  $\chi^2$  с поправкой Йетса на непрерывность — используется, если абсолютные частоты в клетках таблицы частот меньше 10;
- “Фи-коэффициент” (“Phi-square”) — мера ассоциации бинарных признаков;
- “Фишера p” — точный критерий Фишера (“Fisher exact p”) — наиболее точный метод из перечисленных. Предлагаются два варианта метода:
  - “односторонний” (“one-tailed”) — односторонний тест;
  - “двусторонний” (“two-tailed”) — двусторонний тест — предпочтителен в большинстве задач (см. раздел 5.6).
- “МакНемара хи-квадрат” (“McNemar Chi-square”) — используется для связанных групп, например при описании результатов исследований до и после некоторого воздействия (см. раздел 11.4.2).

**Замечание.** Напомним, что при использовании любого из вариантов критерия  $\chi^2$  необходимо убедиться в том, что результаты применения этого критерия (в частности, значение  $p$ ) будут правильными (несмещенными). Правильность результатов обеспечивается только в случае, если частоты во всех ячейках таблицы ожидаемых частот больше или равны 5. Следовательно, в каждом случае необходимо строить таблицу ожидаемых частот (см. раздел 10.2). Для построения такой таблицы рекомендуем использовать следующую опцию.

### STATISTICA:

- Модуль “Основные статистики и таблицы”
- Подмодуль “Таблицы и заголовки” (“Tables and banners”)
  - Опция “Ожидаемые частоты” (“Expected frequencies”) второго диалогового окна “Результаты кросстабуляции” (“Crosstabulation Tables Results”) (см. рис. 10.6)

Точный критерий Фишера является методом выбора в случае, если частота хотя бы в одной ячейке таблицы ожидаемых частот меньше или равна 5. Риск возникновения такой ситуации велик всегда, когда выборки не очень велики. Поэтому точный критерий Фишера предпочтителен во всех случаях не очень больших выборок. Лишь при очень большом количестве объектов исследования (несколько тысяч) его расчеты теряют точность.

Другим традиционно применяемым и рекомендуемым методом является критерий  $\chi^2$  с поправкой Йетса.

### Интерпретация результатов любого из тестов:

- если  $p > 0,05$ , то нулевую гипотезу об отсутствии различий между группами по частоте изучаемого признака не отклоняют;
- если  $p < 0,05$ , то следует отклонить нулевую гипотезу и принять альтернативную гипотезу о существовании различий между группами по частоте изучаемого признака.

**Представление результатов.** Привести следующую информацию:

- таблицу сопряженности (таблицу наблюдаемых частот);

- название используемого статистического критерия;
- точное значение  $p$ .

## 11.4. Сравнение частот бинарного признака в двух связанных (зависимых) группах наблюдений (случай парных наблюдений)

Напомним, что связанными группы являются, например, в следующих случаях:

- если состоят из одних и тех же объектов исследования, например больных, обследованных в разные моменты времени — до и после лечения;
- если больные в две группы набирались подобранными парами (например, по возрасту);
- если исследовались группа детей и группа их родителей;
- если объектами исследования являются не люди, а части тела, например, глаза, каждый из которых является отдельным объектом исследования.

Заметим, что связанные группы — это не обязательно пары наблюдений. В эпидемиологических ретроспективных исследованиях часто на 1 случай (больного) подбираются 3—4 контрольных случая (здоровых).

В данном разделе мы рассматриваем только наиболее простой случай: когда количества наблюдений в группах равны.

**Задача:** сравнить связанные группы по частоте какого-либо признака (симптома, синдрома, исхода).

**Пример:** наблюдалась группа больных до и после лечения. Требуется оценить эффект лечения по наличию некоторого симптома (например, наличие боли до и после лечения).

**Решение:** оценка эффективности метода лечения возможна двумя способами:

- построение ДИ для разности относительных частот в связанных группах;
- проверка статистической гипотезы о различии относительных частот в связанных группах.

В обоих случаях требуется построение таблицы абсолютных частот для парных наблюдений. Такая таблица не является таблицей сопряженности! Она организуется иным способом по сравнению с таблицей сопряженности (рис. 11.11).

До лечения	После лечения	
	симптом есть	симптома нет
Симптом есть	A	B
Симптома нет	C	D

Рис. 11.11. Таблица абсолютных частот парных наблюдений.

Главное ее отличие от таблицы сопряженности состоит в том, что изучаемой единицей для такой таблицы является не один человек, животное или т.п., а пара наблюдений, например один и тот же человек до и после лечения.

### 11.4.1. Доверительный интервал для разности относительных частот

Разность относительных частот бинарного признака в двух связанных группах  $\Delta = P_1 - P_2$  вычисляется по четырехпольной таблице (см. рис. 11.11) с использованием следующей формулы:

$$\Delta = \frac{A+B}{N} - \frac{A+C}{N} = \frac{B-C}{N},$$

где  $N$  — общее число наблюдений,  $P_1 = \frac{A+B}{N}$ ,  $P_2 = \frac{A+C}{N}$ .

Чтобы построить ДИ для разности относительных частот  $\Delta$ , необходимо следующее:

1. Вычислить стандартную ошибку  $m_\Delta$  для разности относительных частот:

$$m_\Delta = \frac{1}{N} \sqrt{B+C - \frac{(B-C)^2}{N}}.$$

2. Вычислить границы ДИ:

$$(P_1 - P_2) - t \times m_\Delta; (P_1 - P_2) + t \times m_\Delta,$$

где  $t$  — значение  $t$ -критерия, соответствующее объему исследуемой выборки и ДК (95, 90 или 99%), обычно принимается равным 1,96 для ДК 95%.

Точное значение  $t$  можно узнать, воспользовавшись опцией “Вероятностный калькулятор”.

**STATISTICA:**

- Модуль “Основные статистики и таблицы” (“Basic statistics / Tables”)
- Подмодуль “Вероятностный калькулятор” (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “*t* Стьюдента” (“Student *t* test”);
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1–ф.р.”
- задать число степеней свободы “Ст. св.” (“*df*”),  $df = n - 1$ ;
- задать значение *p* (например, 0,05 для вычисления границ 95% ДИ);
- нажать кнопку “Вычислить” (“compute”).

Искомое значение *t* появится в окне “*t*”.

**Интерпретация результатов.** Если ДИ не содержит нуля, то можно с уверенностью, соответствующей ДК (например, 95% в случае 95%ДИ) утверждать, что различия относительных частот до и после лечения существуют. В обратном случае различия между группами по данному признаку статистически незначимы.

В качестве примера вычислений ДИ рассмотрим таблицу (рис. 11.12).

До лечения	После лечения	
	симптом есть	симптома нет
Симптом есть	4	16
Симптома нет	3	9

Рис. 11.12. Таблица частот парных наблюдений (пример).

$$P_1 = (4 + 16) / 32 \approx 0,63$$

$$P_2 = (4 + 3) / 32 \approx 0,22$$

$$P_1 - P_2 = 0,63 - 0,22 = 0,41$$

$$m_{\Delta} \approx 0,12$$

$$df = 32 - 1 = 31$$

$$t = 2,04 \text{ (для } df = 31 \text{ и } p = 0,05)$$

$$\text{Границы ДИ: } 0,41 \pm 2,04 \cdot 0,12 \approx [0,17; 0,65].$$

**Интерпретация результата:** ДИ не включает нуль, следовательно, различия между группами по изучаемому признаку являются статистически значимыми.

### 11.4.2. Проверка гипотез (критерий МакНемара)

Данные таблицы абсолютных частот для парных наблюдений следует внести в диалоговое окно следующей процедуры.

**STATISTICA:**

- Модуль “Непараметрические статистики” (“Nonparametrics/Distrib.”)
- Процедура “Таблицы 2×2: хи/V/фи, МакНемара, точный критерий Фишера” (“2×2 Tables XI/VI/Phil, McNemar, Fisher exact”)

После выполнения расчетов откроется окно результатов (см. рис. 11.10). Результат применения статистического критерия для анализа частот в связанных группах будет выведен в следующие строки окна результатов:

- МакНемара  $\chi^2 (A/D)$  — используется, если проверяется гипотеза о том, что частоты *A* и *D* совпадают.
- МакНемара  $\chi^2 (B/C)$  — используется, если проверяется гипотеза о том, что частоты *B* и *C* совпадают.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевую гипотезу об отсутствии различий между группами в отношении частоты изучаемого признака не отклоняют.
- Если  $p < 0,05$ , то нулевую гипотезу отклоняют и принимают альтернативную гипотезу о существовании различий между группами по частоте наблюдения изучаемого признака.

**Представление результатов.** Привести следующую информацию:

- таблицу абсолютных частот парных наблюдений;
- название статистического критерия;
- точное значение *p*.

## 11.5. Сравнение трех групп и более по бинарному признаку

При сравнении трех групп и более по бинарному признаку (т.е. по частоте какого-либо события) возможны две ситуации:



- группы неупорядочены (группирующий признак является номинальным);
  - группы упорядочены (группирующий признак является порядковым).
- Рассмотрим эти две ситуации последовательно.

### 11.5.1. Случай неупорядоченных групп

**Задача:** сравнить частоты в нескольких неупорядоченных группах.

*Пример:* исследуемые группы — больные с гайморитом, синуситом, фронтитом (т.е. группирующий признак — номинальный). Исследуемый фактор риска (бинарный признак) — курение.

**Решение.** Рекомендуется следовать следующей процедуре: провести сравнение по критерию  $\chi^2$  (см. главу 10). Если будет установлено, что частоты событий в группах различны (принята альтернативная гипотеза), то рекомендуется (в отсутствие какой-либо априорной гипотезы) осуществить парное сравнение групп или объединение наиболее сходных групп и сравнение объединенных групп (таблицы  $2 \times 2$ ; см. раздел 11.3.4). При этом следует учитывать проблему множественных сравнений (см. раздел 15.1).

### 11.5.2. Случай упорядоченных групп

**Задача:** сравнить относительные частоты в нескольких упорядоченных группах.

*Пример:* исследуемые группы — больные с I, II и III стадиями гипертонической болезни (группирующий признак — порядковый). Исследуемый фактор риска (бинарный признак) — курение.

**Решение.** Рекомендуемые методы:

- регрессионный анализ (см. главу 14);
- критерий  $\chi^2$  для тенденции (не описывается в настоящем издании; см., например, [26]);
- критерий Манна—Уитни в обратном применении: в качестве группирующего признака выступает бинарный, а собственно группирующий признак анализируется как порядковый.

---

## Глава 12. Анализ связи (корреляции, ассоциации) двух признаков

**Задача:** изучить взаимосвязь (установить наличие и силу связи) признаков следующих типов:

- 1) двух количественных признаков;
- 2) количественного и качественного порядкового признаков;
- 3) двух качественных порядковых;

В статистике принято называть взаимосвязи признаков, упомянутых в пп. 1—2, корреляциями, а взаимосвязь признаков, упомянутых в пункте 3 — ассоциацией.

*Пример 1:* изучить взаимосвязь концентрации холестерина в плазме крови и деформируемости эритроцитов.

**Решение:** анализ корреляций.

**Корреляция** — это описание взаимосвязи количественных или порядковых признаков. Мера (коэффициент) корреляции показывает, в какой степени изменение значения одного признака сопровождается изменением значения другого признака в данной выборке, т.е. в определенных интервалах значений каждого из признаков.

Значения коэффициента корреляции изменяются в интервале от  $-1$  до  $1$ . Крайние значения этого интервала указывают на функциональную линейную зависимость признаков, нуль — на отсутствие статистической связи.

Принята (условно) следующая классификация силы корреляции в зависимости от значения коэффициента корреляции  $r$ :

$|r| \leq 0,25$  — слабая корреляция;

$0,25 < |r| < 0,75$  — умеренная корреляция;

$|r| \geq 0,75$  — сильная корреляция.

Один и тот же коэффициент корреляции, полученный в маленькой или большой выборке, может интерпретироваться совершенно по-разному в зависимости от контекста исследования. Так, малое значение коэффициента корреляции, полученное в эпидемиологиче-

ском исследовании, может быть весьма эпидемиологически значимо, а такое же значение коэффициента корреляции в клиническом исследовании может оказаться клинически незначимым.

Одним из подходов к интерпретации корреляции является вычисление доли объясняемой дисперсии, т.е. доли вариабельности одного признака, зависящего от вариабельности второго признака. Эта мера вычисляется по формуле:  $r^2 \times 100$  (%).

Например, для коэффициента корреляции  $r=0,5$  доля объясняемой дисперсии равна  $100 \times 0,5^2 = 25\%$ .

Знак (плюс или минус) при коэффициенте корреляции указывает направление связи. При отрицательном значении коэффициента корреляции признаки обнаруживают обратную корреляцию (чем больше значение одного признака, тем меньше значение второго признака). И наоборот, при положительном значении коэффициента корреляции связь прямая.

Существование и сила предполагаемой корреляции между признаками могут быть установлены путем проверки нулевой статистической гипотезы о равенстве нулю коэффициента корреляции (т.е. об отсутствии связи признаков). Альтернативная гипотеза: коэффициент корреляции отличен от нуля.

При анализе взаимосвязи качественных (порядковых) признаков корректнее использовать термин “ассоциация” вместо термина “корреляция”.

Помимо парной корреляции в статистике известна и так называемая частная, или парциальная, корреляция. Парциальный коэффициент корреляции отражает силу корреляции между двумя признаками при постоянных значениях всех остальных признаков. Несомненный интерес представляет сравнение между собой по величине и знаку парных и парциальных коэффициентов корреляции для одной и той же пары признаков. В настоящем издании мы не останавливаемся на описании частной корреляции.

**Важное замечание.** Следует особо подчеркнуть, что наличие корреляции двух признаков (любой силы) не может интерпретироваться как доказательство причинно-следственной связи этих признаков. В случае обнаружения корреляции возможны следующие варианты:

- признак 1 влияет на признак 2;
- признак 2 влияет на признак 1;
- оба признака находятся под влиянием третьих факторов.

Судить о том, который из вариантов в действительности имеет место, корреляционный анализ не позволяет; он устанавливает наличие и силу лишь *статистической* связи. Поэтому клиническая и эпидемиологическая интерпретация корреляции весьма трудна.

Отсутствие линейной корреляции не означает, что анализируемые признаки независимы, так как их зависимость может быть нелинейной.

Анализ взаимосвязей (в том числе анализ корреляций) признаков предполагает, что для каждого наблюдения существует измеренное значение каждого из признаков. Наблюдения, для которых информация по какому-либо признаку отсутствует, исключаются из анализа.

Наиболее часто используемыми методами исследования корреляции признаков являются следующие:

- 1) параметрический корреляционный анализ Пирсона (см. раздел 12.1) — для исследования взаимосвязи нормально распределенных количественных признаков. Полученные пирсоновские коэффициенты корреляции можно затем сравнить с использованием специального метода (см. раздел 12.3);
- 2) непараметрические методы корреляционного анализа Спирмена, Кендалла, гамма (см. раздел 12.2) — для исследования взаимосвязи:
  - количественных признаков независимо от вида их распределения;
  - количественного и качественного порядкового признака;
  - двух порядковых признаков.

*Пример 2:* изучить взаимосвязь степени тяжести заболевания с полом.

**Решение:** необходимо провести анализ ассоциаций.

**Ассоциация** — это описание взаимосвязи качественных признаков. Мера ассоциации между качественными признаками является количественным индикатором силы этой связи.

Наличие ассоциации, так же как и наличие корреляции, не является доказательством причинно-следственной связи признаков. Даже сильно ассоциированные признаки могут в действительности зависеть от какого-либо третьего признака.

Анализ ассоциации двух качественных признаков сводится к описанному ранее методу проверки гипотез о независимости

признаков (см. раздел 10.2). После построения таблицы сопряженности, т.е. таблицы, в ячейках которой располагаются абсолютные частоты объектов исследования, ее анализируют и вычисляют меры ассоциации (сопряженности). К таким мерам, в частности, относятся и коэффициенты ранговой корреляции. Их вычисление можно проводить в том же модуле, где и строилась таблица сопряженности (см. раздел 10.2), так и с помощью процедуры, описанной ниже в разделе 12.2.

**Замечание.** *Особый случай оценки взаимосвязи (согласия) признаков, являющихся диагностическими или экспертными заключениями, рассмотрен в разделе 15.5.2.*

## 12.1. Параметрический метод (метод Пирсона)

**Задача:** исследовать линейную связь двух количественных признаков, каждый из которых является нормально распределенным.

**Пример:** анализ связи уровня холестерина в крови и возраста.

**Решение:** для анализа взаимосвязи количественных нормально распределенных признаков используется параметрический корреляционный метод Пирсона. С помощью данного метода проверяется нулевая статистическая гипотеза об отсутствии связи признаков, т.е. о том, что коэффициент корреляции равен нулю. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о том, что коэффициент корреляции не равен нулю.

**Важное замечание.** *Перед проведением данного анализа необходимо проверить гипотезу о нормальности распределения обоих признаков в конкретной исследуемой группе наблюдений (о способах проверки соответствия распределения нормальному закону см. в разделе 7.1). Только в случае принятия гипотезы о нормальности распределения для каждого из этих двух признаков можно применять метод анализа корреляций по Пирсону.*

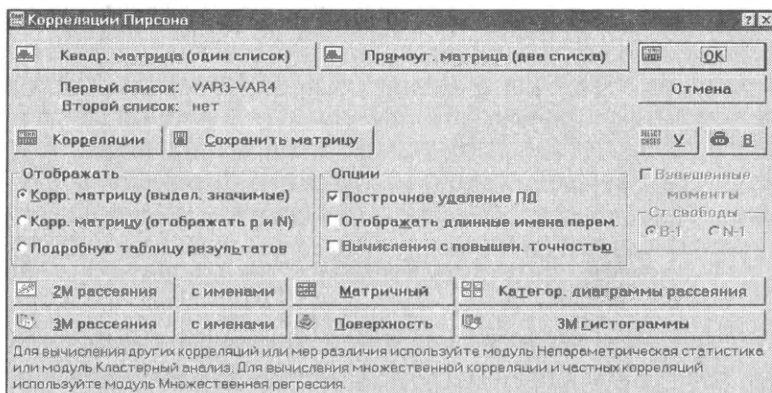
До вычисления коэффициента корреляции рекомендуется построить график с распределением объектов исследования в координатах двух исследуемых признаков с тем, чтобы визуально оценить линейность (либо криволинейность) исследуемой зависимости. Весьма важно оценку коэффициентов корреляции производить только в однородных группах.

Нередко вычисленный по всему массиву данных коэффициент корреляции оказывается статистически незначим, т.е. следует сделать вывод, что корреляция отсутствует. Однако при выполнении этой процедуры в отдельных подгруппах (подобранных по полу, по диагнозу, по стадиям заболевания и т.п.) обнаруживается статистически значимая сильная корреляция. В силу того что в разных подгруппах корреляция может иметь разную направленность (положительную и отрицательную), в общем массиве данных эти корреляции взаимно “гасятся”. Заметим однако, что при проведении корреляционного анализа в подгруппах часто возникает так называемая проблема множественных сравнений (подробнее см. раздел 15.1), чреватая возникновением ложноположительных результатов, т.е. завышением статистической значимости рассчитанных коэффициентов корреляции.

### STATISTICA:

→ Модуль “Основные статистики и таблицы”  
 (“Basic statistics / Tables”)

→ Процедура “Корреляционные матрицы”  
 (“Correlation matrices”) (рис. 12.1)



**Рис. 12.1.** Диалоговое окно процедуры “Корреляции Пирсона” в подмодуле “Корреляционные матрицы” (модуль “Основные статистики и таблицы”).

В диалоговом окне необходимо указать:

- анализируемые признаки. Существует возможность указывать признаки;

- в одном списке: кнопка “Квадратная матрица” (“One variable list (square matrix)”) — для получения результатов корреляционного анализа каждого признака с каждым;
- в двух списках: кнопка “Прямоугольная матрица” (“Two lists (rectang. matrix)”) — для получения результатов корреляционного анализа каждого признака из первого списка с каждым признаком второго списка;
- рекомендуемый режим вывода результатов — “Отображать: Корр. матрицу (отображать  $p$  и  $n$ )” (“Display: Corr. Matrix (display  $p$  and  $n$ )”).

После выполнения анализа по указанным признакам на экран выводится таблица, содержащая значения пирсоновского коэффициента корреляции  $r$ , число пар анализируемых признаков, точное значение  $p$  для каждой пары признаков (рис. 12.2).

Перемен.	NEWVAR13	NEWVAR14	NEWVAR15
NEWVAR13	1.0000	.4485	-.5133
	$p$ = ---	$p$ =.071	$p$ =.035
NEWVAR14	.4485	1.0000	-.5063
	$p$ =.071	$p$ = ---	$p$ =.038
NEWVAR15	-.5133	-.5063	1.0000
	$p$ =.035	$p$ =.038	$p$ = ---

Рис. 12.2. Окно результатов пирсоновского корреляционного анализа.

**Интерпретация результатов.** При анализе связи признаков проверяется нулевая статистическая гипотеза об отсутствии линейной связи признаков (т.е.  $r=0$ ), и вычисляется значение  $p$ :

- если  $p > 0,05$  (или другого принятого уровня значимости), то независимо от значения коэффициента корреляции  $r$  нулевая гипотеза о том, что его значение равно нулю, не отклоняется;
- при  $p < 0,05$  нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о том, что значение коэффициента корреляции  $r$  не равно нулю. В этом случае его значение можно подвергнуть интерпретации.

Если коэффициент корреляции оказался статистически значимым и, кроме того, свидетельствующим о сильной корреляции,

то имеются основания попробовать интерпретировать стоящие за ней биологические закономерности. Подчеркнем еще раз, что наличие корреляции между двумя признаками вовсе не означает обязательного наличия прямой причинно-следственной связи между ними. Вполне возможно, что согласованные изменения обоих признаков есть следствие вариации некоторого третьего признака.

### **Рекомендуемое представление результатов:**

- приведите значение коэффициента корреляции  $r$  (с точностью до двух значащих цифр);
- укажите название примененного метода корреляционного анализа;
- приведите точное значение  $p$ ;
- укажите число анализируемых пар признаков (для каждой пары признаков это могут быть разные числа, желательно привести их все);
- приведите 95% ДИ для наиболее важных коэффициентов корреляции (о расчете ДИ см. раздел 12.3) с целью интерпретации биологической или клинической значимости связи;
- для наиболее важных результатов для наглядности желательно представить график рассеяния объектов исследования в координатах, соответствующих двум исследуемым признакам (построение графиков иницируется кнопкой “2М рассеяния” (“2D scatterp.”) в диалоговом окне (см. рис. 12.1)). Чем сильнее связь, тем более вытянуто облако наблюдений вдоль линии тренда (рис. 12.3);
- если изучались связи нескольких признаков, удобно представить результаты в виде корреляционной матрицы с обязательным указанием значений  $p$ .

## **12.2. Непараметрические методы (методы Спирмена, Кендалла, гамма)**

**Задача:** исследовать связь двух признаков в любом из следующих случаев:

- анализируются два количественных признака, и нет сведений о виде их распределений;
- ранее установлено, что по крайней мере одно из распределений анализируемых количественных признаков не является нормальным;



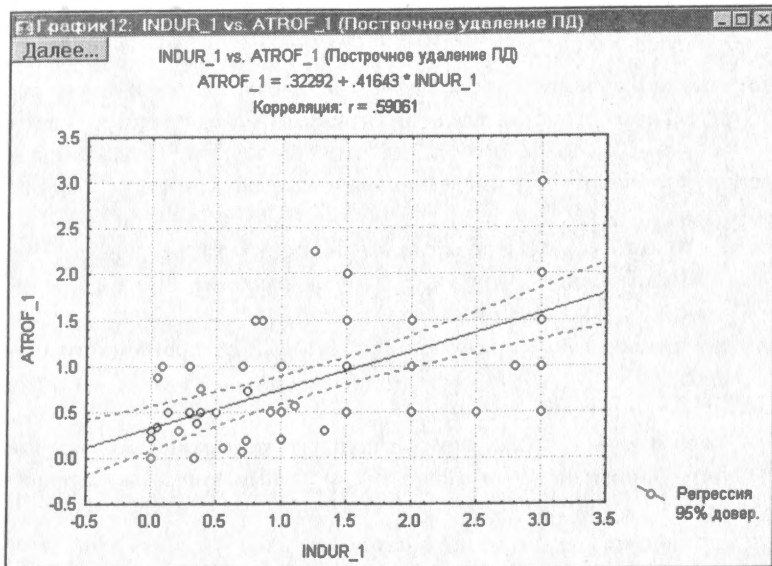


Рис. 12.3. График рассеяния при анализе корреляционной связи.

- анализируются количественный и качественный порядковый признаки;
- анализируются два качественных порядковых признака.

*Пример:* анализ связи концентрации холестерина в плазме крови и степени атеросклероза сонных артерий.

**Решение:** для анализа данных такого типа необходимо применять метод непараметрического корреляционного анализа. В ППП **STATISTICA** представлены следующие непараметрические методы корреляционного анализа:

- ранговая корреляция по Спирмену ( $\rho$ ); метод рекомендуется использовать для оценки взаимосвязи количественных (независимо от вида распределения) и/или порядковых признаков;
- ранговая корреляция по Кендаллу ( $\tau$ ); метод рекомендуется использовать для оценки взаимосвязи двух порядковых признаков или порядкового и количественного признаков;
- гамма-корреляция; рекомендуется использовать, когда в данных имеется много совпадающих значений.

С помощью этих методов проверяют нулевую статистическую гипотезу об отсутствии связи признаков, т.е. о том, что коэффи-

коэффициент корреляции равен нулю. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о том, что коэффициент корреляции не равен нулю.

### STATISTICA:

- Модуль “Непараметрические статистики” (“Nonparametrics/Distrib.”)
  - Процедура “Корреляции (Спирмена, тау Кендалла, гамма)” (“Correlations (Spearman, Kendall tau, gamma)”)

В диалоговом окне (рис. 12.4) необходимо указать следующее:

- анализируемые переменные в двух списках для получения результатов корреляционного анализа каждого признака из первого списка с каждым признаком второго списка;
- метод корреляционного анализа “Корреляция: Спирмена R”, “...тау Кендалла” или “...гамма”;
- рекомендуемый режим “Вычислить: Подробный отчет” (“Compute: Detailed report”).

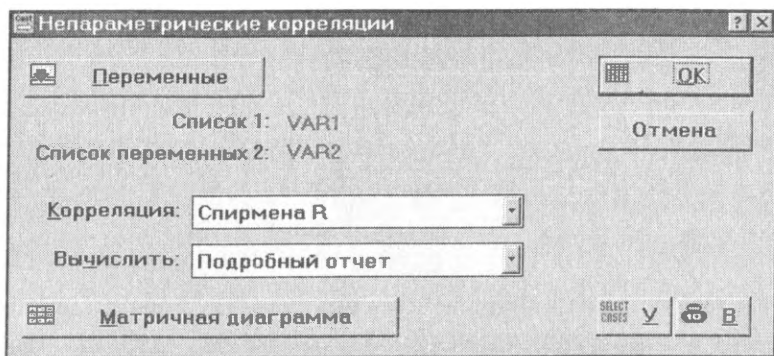


Рис. 12.4. Диалоговое окно “Непараметрические корреляции” в разделе “Корреляции Спирмена, тау Кендалла, гамма” (модуль “Непараметрические статистики”).

После выполнения анализа по указанным признакам на экран выводится таблица, содержащая число пар анализируемых признаков, значения коэффициента корреляции, значения  $r$ .

Вызов программы “Матричная диаграмма” (“Matrix plot”) инициирует построение и выведение на экран графиков рассеяния.

**Интерпретация результатов** анализа и рекомендуемое их представление аналогичны таковым для пирсоновской корреляции (см. раздел 12.1).

### 12.3. Доверительный интервал для коэффициента корреляции

Не приводя математического обоснования расчетов, опишем способ построения ДИ для коэффициента корреляции (к сожалению, этот алгоритм не реализован в ППП *STATISTICA*). Данный способ может применяться для расчета ДИ как для пирсоновского коэффициента корреляции, так и для коэффициентов корреляции, рассчитанных непараметрическими методами в тех случаях, если объем выборки превышает 10.

1. Вычисляется значение функции  $z$ :

$$z = 0,5 \cdot \ln \frac{1+r}{1-r}$$

2. Вычисляется стандартная ошибка  $m_z$  для  $z$ :

$$m_z = \frac{1}{\sqrt{n-3}},$$

где  $n$  — число объектов исследования, для которых рассчитывалась  $r$ .

3. Вычисляются нижняя граница  $z_1$  и верхняя граница  $z_2$  ДИ для величины  $z$ :

$$\begin{aligned} z_1 &= z - t \times m_z \\ z_2 &= z + t \times m_z \end{aligned}$$

где  $t$  — значение  $t$ -критерия для данного числа степеней свободы  $df = n - 1$  и заданного уровня значимости. Точное значение  $t$  можно узнать, воспользовавшись следующей процедурой.

#### **STATISTICA:**

→ Модуль “Основные статистики и таблицы”  
 (“Basic statistics / Tables”)

→ Подмодуль “Вероятностный калькулятор”  
 (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “*t* Стьюдента” (“Student *t* test”);
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1–ф.р.”;
- задать число степеней свободы “Ст. св.” (“*df*”),  $df = n - 1$ ;
- задать значение *p* (например, 0,05 для вычисления границ 95% ДИ);
- нажать кнопку “Вычислить” (“compute”).

Искомое значение *t* появится в окне “*t*”.

4. Вычисляются нижняя и верхняя границы ДИ для коэффициента корреляции *r*:

$$\left[ \frac{e^{2z_1} - 1}{e^{2z_1} + 1}; \frac{e^{2z_2} - 1}{e^{2z_2} + 1} \right]$$

### Интерпретация ДИ для коэффициента корреляции.

Можно считать с вероятностью, соответствующей доверительному коэффициенту (с вероятностью 95% при построении 95% ДИ), что истинное значение коэффициента корреляции находится в рассчитанных границах.

## 12.4. Сравнение двух коэффициентов корреляции

Весьма полезно производить проверку статистических гипотез о равенстве коэффициентов корреляции (одного знака), полученных в разных группах объектов исследования. Особенно важно сравнивать их между группами “больных” и “здоровых”.

**Задача:** сравнить два коэффициента корреляции, т.е. установить, случайны ли их различия.

**Пример:** в одной группе наблюдений (больные с ишемическим инсультом) коэффициент корреляции тяжести инсульта (сумма баллов по Скандинавской шкале степени неврологического дефицита) и уровня артериального давления составил 0,45 ( $p=0,011$ ), в другой группе наблюдений (больные с геморрагическим инсультом) коэффициент корреляции оказался равным 0,56 ( $p=0,023$ ). Требуется установить, случайны ли различия между двумя коэффициентами корреляции.

**Решение:** могут использоваться два подхода:

- 1) вычисление ДИ для каждого из коэффициентов корреляции и последующее их сравнение. Если ДИ пересекаются, то коэффициенты корреляции статистически значимо не различаются (их различие носит случайный характер);
- 2) проверка гипотезы о равенстве коэффициентов корреляции. При применении статистического теста происходит проверка нулевой статистической гипотезы об отсутствии различий между коэффициентами корреляции. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании различий между коэффициентами корреляции. Первый подход очевидно ясен из раздела 12.3. Рассмотрим второй подход подробнее.

### **STATISTICA:**

→ Модуль “Основные статистики и таблицы”  
 (“Basic statistics/Tables”)

→ Подмодуль “Другие критерии значимости”  
 (“Probability calculator” — кнопка “More”  
 в версии **STATISTICA 4.\***)

→ Процедура “Различие между двумя  
 коэффициентами корреляции” (“Difference  
 between two correlation coefficients”)  
 (рис. 12.5)

### **Опции:**

- двусторонний тест (если нет теоретического обоснования того, что один из коэффициентов корреляции заведомо больше другого);
- односторонний тест (если есть такое теоретическое обоснование, и требуется лишь установить, являются ли различия статистически значимыми).

После введения значений коэффициентов корреляции и соответствующих объемов двух выборок необходимо нажать кнопку “Вычислить” (“Compute”). Будет вычислено значение  $p$  при проверке гипотезы об отсутствии различий между указанными коэффициентами корреляции.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза об отсутствии различий между коэффициентами корреляции не отклоняется.

Другие критерии значимости

Печатать результаты после каждого вычисления Отмена

Различие между двумя коэффициентами корреляции

$r_1$ : .40  N1: 100   $r$ : .0306  Односторонний  Двусторонний Вычислить

$r_2$ : .60  N2: 200

Различие между двумя средними (нормальное распределение)

$\bar{X}_1$ : 0.  Ст.откл. 1.  N1: 10   $p$ : 1.0000 Вычислить

$\bar{X}_2$ : 0.  Ст.откл. 1.  N2: 10   Односторонний  Двусторонний

Среднее выборки 1 в сравнении со средним популяции 2

Различие между двумя пропорциями

Пр.1: .50  N1: 10   $p$ : .5000  Односторонний  Двусторонний Вычислить

Пр.2: .50  N2: 10

Рис. 12.5. Диалоговое окно для сравнения двух коэффициентов корреляции (модуль “Основные статистики и таблицы”, подмодуль “Другие критерии значимости”, процедура “Различие между двумя коэффициентами корреляции”).

- Если  $p < 0,05$  (или другого принятого уровня значимости), то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о том, что коэффициенты корреляции различны.

**Рекомендуемое представление результатов:**

- привести значения коэффициентов корреляции;
- привести объемы выборок, на которых эти значения получены (для каждой пары признаков это могут быть разные количества пар);
- указать точное значение  $p$ , полученное при сравнении коэффициентов корреляции.

## Глава 13. Многофакторный анализ данных

В предыдущих главах мы рассматривали задачи одномерного и двумерного анализа данных:

- сравнение групп по одному признаку или группе признаков, при этом каждый признак анализировался отдельно от других (одномерный анализ);
- анализ взаимосвязи (корреляции, ассоциации) двух признаков (двумерный анализ).

Помимо перечисленных методом двумерного анализа является также однофакторный регрессионный анализ — частный случай многофакторного регрессионного анализа (будет рассмотрен в главе 14).

Кроме одно- и двумерного анализа существует также большое разнообразие методов многофакторного анализа данных, которые позволяют анализировать одновременно три признака и более. К самым популярным методам многофакторного анализа данных относятся, в частности, следующие:

- регрессионный анализ (см. главу 14);
- многофакторный дисперсионный анализ (см. раздел 13.1);
- дискриминантный анализ (см. раздел 13.2);
- кластерный анализ (см. раздел 13.3);
- анализ главных компонент и факторный анализ (см. раздел 13.4).

В настоящем издании мы считаем целесообразным более подробно остановиться лишь на первом из них, поскольку для корректного использования методов многофакторного анализа требуется достаточно глубокое знание статистики, в связи с чем мы настоятельно рекомендуем проводить такую работу совместно со специалистом по прикладной статистике. Однако в данной главе нам все же хотелось бы дать информацию читателю-врачу о тех задачах, которые могут быть решены с помощью этих методов.

**Замечание.** *Выполнение многофакторного анализа данных и особенно интерпретация результатов любого из этих методов — достаточно сложная и ответственная задача!*

В двух последних разделах данной главы мы остановимся на общих проблемах многофакторного анализа данных:

- анализ наблюдений с пропущенными значениями (см. раздел 13.5);
- проверка работоспособности (валидизация) многомерных моделей (см. раздел 13.6).

## 13.1. Многофакторный параметрический дисперсионный анализ

Ранее (см. раздел 8.4.1) мы рассмотрели наиболее простой вариант дисперсионного анализа — анализ нескольких групп по одному количественному признаку. В данном разделе мы рассмотрим более сложную ситуацию.

**Задача:** исследовать, зависит ли значение количественного нормально распределенного признака от значений двух и более количественных или качественных признаков (факторов) одновременно (для оценки одновременных эффектов на зависимый признак).

**Пример:** изучить, зависит ли артериальное давление от пола и возрастной группы.

**Решение:** многофакторный параметрический дисперсионный анализ оценивает влияние двух и более качественных или количественных независимых признаков (факторов) на один количественный зависимый признак.

### Условия применимости метода:

- нормальность распределения количественного признака во всех сравниваемых подгруппах;
- равенство дисперсий количественного признака во всех сравниваемых подгруппах.

Проверка выполнимости этих требований была описана выше (см. разделы 7.1 и 8.4.1.1 соответственно). В случае, если данные не отвечают перечисленным условиям применимости метода, результаты применения многофакторного дисперсионного анализа оказываются смещенными (искаженными). Подробно последствия нарушения этих требований описаны в литературе [27, 28].



Для проведения многофакторного дисперсионного анализа следует обратиться к следующей процедуре ППП.

### STATISTICA:

→ Модуль “Дисперсионный анализ (ANOVA/MANOVA)”

В первом диалоговом окне (см. рис. 13.1) необходимо указать:

- независимые признаки (факторы);
- зависимый признак.

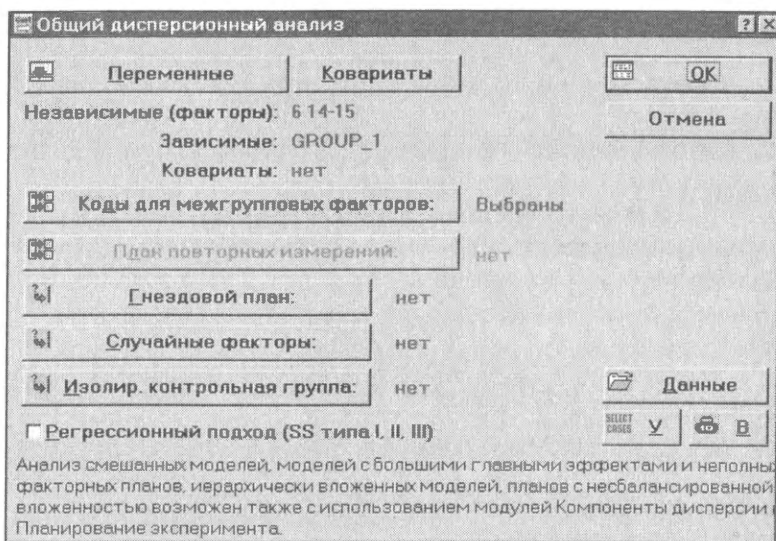


Рис. 13.1. Диалоговое окно многофакторного дисперсионного анализа (модуль ANOVA/MANOVA).

Анализ проводится в два этапа.

**1. Собственно анализ вариаций.** На данном этапе вычисляется значение  $F$ -критерия и соответствующее значение  $p$ . Процедура анализа активизируется нажатием кнопки “Все эффекты” во втором диалоговом окне (рис. 13.2). Результаты анализа вариаций представляются в окне результатов (рис. 13.3).

**2. Парные сравнения.** Выполняются путем активизации процедур в окне (рис. 13.4), вызываемом нажатием кнопки “Апо-

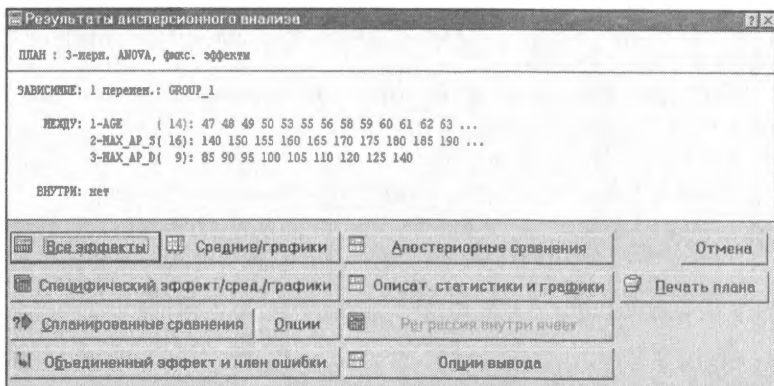


Рис. 13.2. Второе диалоговое окно многофакторного дисперсионного анализа (модуль ANOVA/MANOVA).

Эффект	Ст. св. Эффекта	MS Эффекта	Ст. св. Ошибки	MS Ошибки	F	p-уров.
1	15	.507720	42	.157341	3.226873	.001391

Рис. 13.3. Окно результатов анализа вариаций в многофакторном дисперсионном анализе (модуль ANOVA/MANOVA).

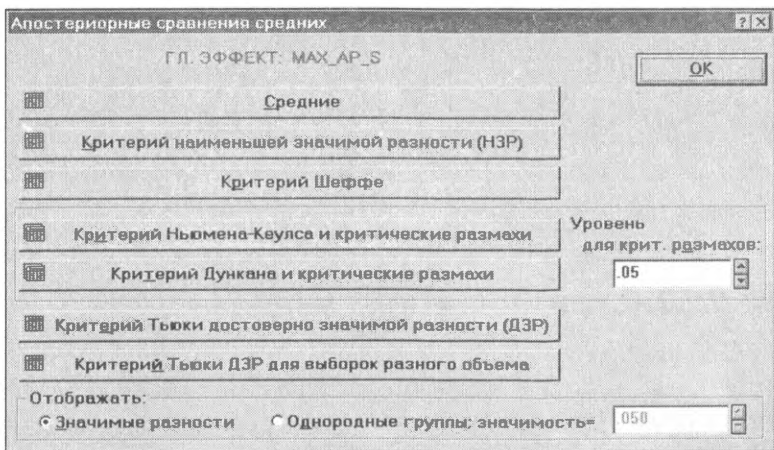


Рис. 13.4. Диалоговое окно процедур апостериорных сравнений в многофакторном дисперсионном анализе (модуль ANOVA/MANOVA).

стериорные сравнения” (“Post-hoc comparisons”) во втором диалоговом окне (см. рис. 13.2).

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p > 0,05$ , то нулевая гипотеза об отсутствии влияния изучаемых факторов на зависимый признак не отклоняется.
- Если  $p < 0,05$ , то нулевая гипотеза отклоняется, и следует принять альтернативную гипотезу о существовании влияния изучаемых факторов на зависимый признак, описываемого моделью с параметрами, оценки которых получены в этой процедуре.

Об интерпретации результатов апостериорных сравнений см. раздел 8.4.1.3.

**Рекомендуемое описание результатов.**

- Сначала опишите исследуемые признаки с помощью описательной статистики.
- Подтвердите, что условия применимости дисперсионного анализа соблюдаются.
- Приведите результаты дисперсионного анализа в виде таблицы (желательно).
- Укажите, проверялись ли независимые признаки на взаимосвязь и каким образом.
- Приведите точные значения  $p$  для каждого коэффициента модели.

**Замечание.** Статистически значимые результаты регрессионного анализа (впрочем, как и других видов многофакторного анализа) получить трудно в силу обычно небольшого числа наблюдений в медицинских исследованиях и наличия пропусков в данных (см. раздел 13.5). Также большую трудность всегда представляет интерпретация результатов.

## 13.2. Дискриминантный анализ

**Задача 1:** разработать формальное правило (иногда его называют еще решающим правилом) отнесения нового объекта исследования к той или иной ранее изученной группе — в задачах диагностики или прогноза.

**Пример 1:** изучались две группы больных — с хорошей и плохой динамикой восстановления после черепно-мозговой травмы. Требуется построить правило прогноза динамики (хорошая

либо плохая) у вновь поступающего пациента на основе нескольких количественно измеренных симптомов.

**Задача 2:** выявить наиболее информативные (в смысле дифференциации двух групп объектов и более) признаки из некоторого набора признаков.

**Пример 2:** найти особенности липидного спектра крови у больных с разными локализациями преимущественного поражения сосудистой стенки при атеросклерозе, например с преимущественным поражением коронарных или сонных артерий.

**Решение:** для решения таких задач можно воспользоваться методом дискриминантного анализа — методом исследования связи одного качественного признака (в примере 1 — тип динамики, в примере 2 — локализация преимущественного поражения сосудистой стенки) и набора количественных нормально распределенных признаков (в примере 1 — симптомов, в примере 2 — концентрации липидов). При использовании этого метода алгоритм выбирает такие признаки, значения которых наиболее близки у объектов исследования одной группы и наиболее различны у объектов исследования из разных групп.

При этом признак, детерминирующий группы объектов исследования (группирующий признак), должен быть качественным и иметь 2 значения и более.

**Замечание.** *Дискриминантный анализ — один из методов решения так называемой задачи классификации, или распознавания образов — задачи отнесения изучаемого объекта (например, больного) к одной из нескольких групп на основе измерения у него некоторого числа признаков (например, симптомов).*

## STATISTICA:

→ Модуль “Дискриминантный анализ”  
 (“Discriminant analysis”) (рис. 13.5)

В первом диалоговом окне (см. рис. 13.5) следует выбрать группирующий признак и независимые признаки, нажав на кнопку “Переменные”. Во втором диалоговом окне (рис. 13.6) необходимо выбрать метод анализа — стандартный (standard), прямой пошаговый (forward stepwise), обратный пошаговый (backward stepwise).

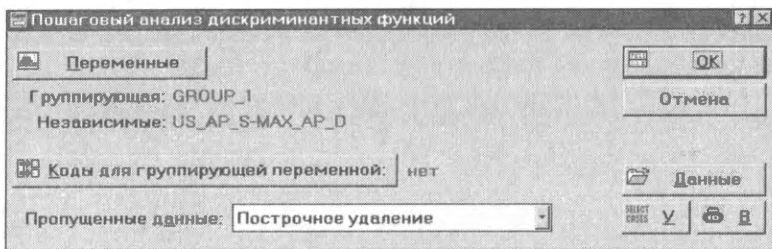


Рис. 13.5. Первое диалоговое окно дискриминантного анализа.

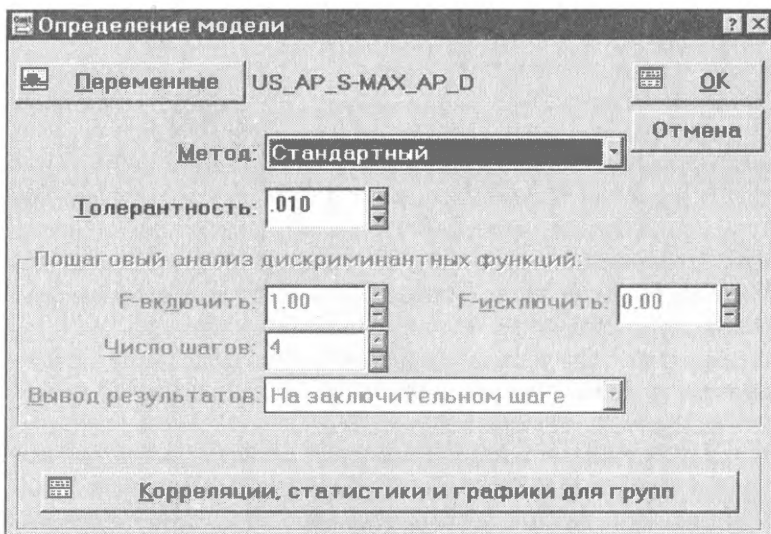


Рис. 13.6. Второе диалоговое окно дискриминантного анализа.

Результатом дискриминантного анализа является оценка коэффициентов математической модели — линейной дискриминантной функции (ЛДФ). ЛДФ — это линейный многочлен следующего вида:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n + C,$$

где  $x_i$  — наиболее информативные из анализируемых признаков,  $a_i$  — коэффициенты,  $C$  — константа.

В окно результатов (рис. 13.7) выводится значение  $p$  для построенной математической модели (в верхней части окна). В нижней части окна доступны различные опции для более под-

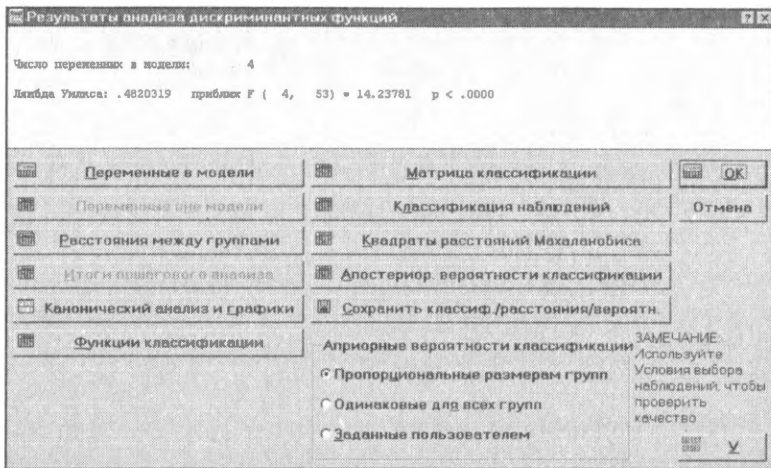


Рис. 13.7. Первое окно результатов дискриминантного анализа.

робного представления результатов дискриминантного анализа, из которых наиболее важными можно считать следующие:

- “Переменные в модели” (“Variables in the model”) — значения  $p$  для каждого из включенных в модель независимых признаков. Отбор признаков осуществляется в соответствии с определенными статистическими критериями. В результате в ЛДФ включаются те признаки, которые в соответствии с этими критериями являются наиболее информативными и при этом наименее зависимыми друг от друга.
- “Функции классификации” (“Classification functions”) — коэффициенты линейных дискриминантных функций для каждого из значений зависимого признака;
- “Матрица классификации” (“Classification matrix”) — один из главных результатов дискриминантного анализа. Пример такой матрицы приведен на рис. 13.8. Из приведенного примера видно, что 80 наблюдений (объектов исследования) из 100, относящихся к 1-й группе, правильно отнесены в 1-ю группу с помощью ЛДФ (качество распознавания 80%); 70 наблюдений (объектов исследования) 2-й группы правильно отнесены в свою группу (качество распознавания 70%). Среднее качество распознавания составило 75%.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

	1-я группа (прогнозируемое состояние)	2-я группа (прогнозируемое состояние)	Всего
1-я группа (истинное состояние)	80 (80%)	20	100
2-я группа (истинное состояние)	30	70 (70%)	100
Всего	110	90	200

**Рис. 13.8.** Матрица классификации как результат применения линейной дискриминантной функции (число и доля объектов исследования).

- При  $p < 0,05$  можно считать, что построенная модель достаточно хороша.
- Если качество распознавания достаточно высокое (более 90% в каждой из исследуемых групп), то можно считать ЛДФ достаточно эффективной, и стоит проверить работоспособность модели (валидизировать модель)<sup>1</sup>.
- При использовании пошаговой (stepwise) процедуры анализ порядка включения признаков может быть полезен для интерпретации относительной информативности признаков и их связей.

**Проверка работоспособности модели.** Результатом дискриминантного анализа является создание статистической модели (уравнений линейной дискриминантной функции). Как у всякой математической модели, у этой модели следует проверять работоспособность (валидизировать модель). Процедура проверки работоспособности модели заключается в применении ЛДФ к новым объектам исследования, не вошедшим в обучающую выборку — экзаменационной выборке — с целью оценки эффективности ЛДФ и устойчивости результата распознавания (подробнее см. раздел 13.5).

**Представление результатов.** Привести следующую информацию:

- число объектов исследования в каждой из групп;

<sup>1</sup> Высокие требования к качеству распознавания (не менее 90% правильного распознавания) обусловлены тем, что оно должно быть заведомо выше того, которое может быть достигнуто без применения статистического моделирования.

- процент правильной классификации в каждой из групп;
- таблицу с коэффициентами ЛДФ;
- точное значение  $p$  для уравнения ЛДФ.

**Замечание.** Для проведения дискриминантного анализа мы рекомендуем обращаться к специалистам в области прикладной статистики или к специальной литературе, например [24].

### 13.3. Кластерный анализ

Кластерный анализ является методом поиска закономерностей группирования как объектов исследования, так и признаков в отдельные локальные подмножества (кластеры).

Основными типами кластерного анализа являются следующие:

- группировка объектов исследования в кластеры;
- группировка признаков в кластеры;
- одновременная группировка объектов исследования и признаков.

**Группировка объектов исследования в кластеры** применяется в тех случаях, когда предполагается, что имеющаяся выборка гетерогенна, но причина гетерогенности при этом неизвестна. Результатом применения процедуры кластеризации может быть формирование нескольких подгрупп (кластеров) объектов исследования, в каждой из которых содержатся сходные наблюдения. Дальнейший анализ подгрупп может выявить некоторые объективные признаки, по которым эти подгруппы различаются.

**Группировка признаков в кластеры** применяется на достаточно однородной (в отношении наблюдений, или объектов исследования) выборке с целью поиска неизвестных закономерностей связи признаков (или групп признаков). Результатом может быть формирование нескольких групп признаков, в каждой из которых содержатся признаки, обнаружившие статистически значимые взаимосвязи.

Проведение кластерного анализа и интерпретация его результатов достаточно сложны, поэтому мы рекомендуем обращаться за помощью к специалистам в области прикладной статистики или к специальной литературе, например [24, 29].



## **13.4. Анализ главных компонент и факторный анализ**

В типичном клиническом исследовании число исследованных признаков бывает достаточно велико. Однако большое количество признаков (большая размерность признакового пространства) не только приводит к увеличению полноты и глубины исследования проблемы, но и затушевывает имеющиеся закономерности. Для решения этой проблемы могут применяться обсуждаемые методы.

Основными их целями являются:

— снижение размерности признакового пространства путем выявления статистически значимых линейных комбинаций нескольких признаков (анализ главных компонент);

— изучение взаимосвязи признаков путем выявления факторов, определяющих дисперсию признаков (факторный анализ).

Проведение статистического анализа этих видов является достаточно сложным, поэтому мы рекомендуем обращаться за помощью к специалистам в области прикладной статистики или к специальной литературе [24, 30—32].

## **13.5. Многофакторный анализ объектов исследования с пропущенными значениями**

Если в исследовании ставится задача, для решения которой требуется использование методов многомерного анализа данных, рекомендуется оценить имеющиеся данные с двух точек зрения.

Во-первых, общей рекомендацией при построении статистических моделей является следующая: число объектов исследования должно существенно (хотя бы в 3—5 раз для каждой из групп) превышать число признаков, включаемых в многофакторный анализ. Если данное условие не выполняется, почти наверняка можно сказать, что полученная модель окажется неустойчивой, т.е. нероботоспособной на других наборах данных.

Во-вторых, при многофакторном анализе данных (одновременном анализе нескольких признаков) в большинстве исследований возникает следующая проблема: у значительной доли объектов исследования отсутствуют данные по какому-либо

одному или даже нескольким признакам. Такая ситуация достаточно типична для клинических исследований, где бывает весьма трудоемко (или дорого) провести одинаково подробное обследование всех больных, включенных в выборку. Это особенно касается не проспективных исследований, проводящихся в большинстве случаев по заранее разработанному протоколу, а ретроспективных исследований, в которых анализируются архивные данные.

В случае одномерного анализа объект исследования, для которого информация о каком-либо признаке отсутствует, не участвует в анализе именно по этому признаку, а участвует по тем признакам, которые у него имеются.

При двумерном анализе (как мы уже описывали в главе 12) из анализа исключаются объекты исследования, у которых отсутствуют значения хотя бы по одному из пары анализируемых признаков.

При многофакторном анализе объект исследования, у которого отсутствуют данные хотя бы по одному признаку из числа включенных в анализ (регрессионный, дискриминантный и др.), также исключается из анализа<sup>1</sup>. Обычно оказывается, что доля объектов исследования, у которых отсутствуют данные хотя бы по одному из признаков, достаточно велика. Таким образом, число объектов исследования, которые могут участвовать в многофакторном анализе, оказывается гораздо меньше общего числа объектов исследования. Иногда их оказывается просто мало. Так, при числе объектов исследования меньше 20 в каждой из групп многомерный анализ проводить вообще бесперспективно.

Существует возможность преодоления этой проблемы. В ряде ППП существует возможность заполнения пропусков в данных несколькими способами, например:

- пропуски заменяются средними по группе (предлагается в ППП **STATISTICA**);
- пропуски заменяются значениями регрессионной функции.

Такие алгоритмы хотя и привлекательны, но дают слишком оптимистические результаты, которые обычно потом не подтверждаются при проверке работоспособности математической модели.

---

<sup>1</sup> Анализ обычно проводится в режиме *casewise deletion*, т.е. при наличии пропуска в данных хотя бы по одному признаку объект исследования исключается из анализа.

## 13.6. Проверка работоспособности многофакторных моделей

Все модели многофакторного анализа необходимо проверять на работоспособность, т.е. оценивать их устойчивость и эффективность на разных наборах данных.

Выборка, на которой происходило построение модели (например, оценка коэффициентов линейной дискриминантной функции в дискриминантном анализе), называется *обучающей*. Другая выборка, на которой оценивается работоспособность модели, называется *экзаменационной*.

Существуют несколько основных методов оценки работоспособности модели:

- если обучающая выборка велика, можно разделить ее на две, соотносящиеся по числу объектов исследования, например как 7:3, затем построить 2 модели на каждой из выборок, а затем модели сравнить;
- применение процедуры “скользящего экзамена” (“jack-knife”) — временное поочередное удаление объектов исследования из набора данных и пересчет параметров модели, а затем сопоставление полученных моделей. Этот метод применяется при существенных трудностях со сбором данных для экзаменационной выборки. Такой способ может использоваться главным образом при наличии соответствующих программных процедур. К сожалению, в ППП *STATISTICA* такие процедуры не предусмотрены;
- оценка работоспособности модели на дополнительной, вновь набранной экзаменационной выборке;
- создание другой модели на другом наборе сходных данных и анализ идентичности моделей.

---

## Глава 14. Анализ вида зависимости одного признака от одного или нескольких признаков (регрессионный анализ)

**Задача:** исследовать вид зависимости одного признака (количественного или качественного) от одного или нескольких других признаков (количественных или качественных) с целью прогнозирования значения одного признака по значению одного или нескольких признаков.

**Пример:** изучить возможность прогнозирования возраста, в котором начнет развиваться некое наследственное заболевание у пациента, по активности фермента в плазме его крови и возрасту, в котором это заболевание началось у родителя пациента.

**Решение:** регрессионный анализ — один из методов статистического моделирования. Моделью в данном случае является уравнение регрессии, параметры (коэффициенты) которого и рассчитываются в ходе регрессионного анализа.

Регрессионный анализ тесно связан с другими статистическими методами — методами корреляционного и дисперсионного анализа. В отличие от корреляционного анализа, который изучает направление и силу статистической связи признаков, регрессионный анализ изучает *вид* зависимости признаков, т.е. параметры функции зависимости одного признака (зависимого<sup>1</sup>, объясняемого<sup>2</sup>, исхода<sup>3</sup>, доли больных, ответивших на то или иное вмешательство<sup>4</sup>) от одного или нескольких других при-

---

<sup>1</sup> Dependent (англ.).

<sup>2</sup> Explained (англ.).

<sup>3</sup> Outcome (англ.).

<sup>4</sup> Response (англ.); в отечественной литературе иногда встречается термин “отклик”.

знаков (независимого<sup>1</sup>, объясняющего<sup>2</sup>, прогностического<sup>3</sup>, фактора<sup>4</sup>, ковариаты<sup>5</sup>).

В отличие от дисперсионного анализа, с помощью которого исследуется зависимость количественного признака от одного или нескольких *качественных* признаков, в регрессионном анализе обычно исследуется зависимость (количественного или качественного) признака от одного или нескольких *количественных* признаков. Кроме того, ANOVA является одним из этапов процедуры регрессионного анализа.

**Замечание.** Регрессионный анализ (как и описываемый в разделе 13.1 дискриминантный анализ, а также не описываемые подробно в настоящем издании факторный анализ, кластерный анализ и т.д.) относится к методам математического моделирования и является достаточно сложным для проведения и интерпретации без соответствующей математической подготовки. В связи с этим при описании перечисленных методов мы останавливаемся лишь на их возможностях и не описываем подробно все опции и параметры, которые задействованы в процедурах, реализующих указанные методы.

### Классификация методов регрессионного анализа

1. По количеству независимых признаков:
  - однофакторный, или простой (один независимый признак);
  - многофакторный (два независимых признака и более).
2. По типу математической зависимости:
  - линейный (полиномиальная функция первой степени; см. раздел 14.1);
  - нелинейный:
    - логистический (степенная функция; см. раздел 14.2);
    - пробит-регрессия<sup>6</sup>;
    - экспоненциальная регрессия<sup>7</sup>;

---

<sup>1</sup> Independent (англ.).

<sup>2</sup> Explanatory (англ.).

<sup>3</sup> Predictor (англ.).

<sup>4</sup> Factor (англ.).

<sup>5</sup> Covariate (англ.) — сопутствующая случайная величина; вспомогательная переменная.

<sup>6</sup> Probit regression (англ.).

<sup>7</sup> Exponential growth regression (англ.).

- кусочно-линейная регрессия<sup>1</sup>;
- регрессия пропорциональных рисков по Коксу (см. раздел 15.4).

Подчеркнем, что однофакторная регрессионная модель является методом анализа двух признаков — независимого и зависимого. Таким образом, однофакторный регрессионный анализ является методом двумерного анализа данных, как и корреляционный анализ.

Из методов нелинейной регрессии мы рассмотрим лишь множественный логистический регрессионный анализ как наиболее актуальный метод для медицинских исследований.

**Замечание.** *Классический регрессионный анализ не предусматривает использование количественных дискретных и качественных признаков. Специальные приемы использования таких признаков в регрессионном анализе описаны, например, в [33, 34].*

Этапы построения регрессионной модели следующие:

- 1) анализ ассоциаций зависимого признака с каждым из независимых путем оценки корреляций и построения двумерных графиков;
- 2) отбор наиболее сильных ассоциаций;
- 3) построение регрессионного уравнения.

Цель регрессионного анализа — поиск таких комбинаций независимых признаков, которые “лучше” (в определенном статистическом смысле — более высокого значения коэффициента детерминации  $R^2$ , меньших значений дисперсии ошибок и т.д.) прогнозируют значение зависимого признака.

В ходе выполнения регрессионного анализа происходит проверка нулевой статистической гипотезы об отсутствии связи признаков. Если нулевая гипотеза отклоняется, то следует принять альтернативную гипотезу о существовании связи признаков.

## 14.1. Линейный регрессионный анализ

**Задача:** рассчитать значение (выполнить прогнозирование) одного количественного (объясняемого) признака по значениям одного или нескольких других количественных (объясняющих) признаков.

<sup>1</sup> Piecewise linear regression (англ.).

*Пример:* необходимо найти правило прогнозирования длительности острой фазы инфекционного заболевания по уровню антител в крови и температуре тела.

**Решение:** линейный регрессионный анализ. При этом происходит построение уравнения следующего вида:

$$Y = a + bX \text{ (в случае простой линейной регрессии)}$$

или

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \text{ (в случае множественной линейной регрессии),}$$

где  $X_1, X_2, \dots, X_n$  — независимые (объясняющие) признаки (прогностические признаки),  $Y$  — зависимый (объясняемый) признак,  $a$  — константа,  $b_1, b_2, \dots, b_n$  — коэффициенты регрессии (в случае простого регрессионного анализа коэффициент  $b$  является тангенсом угла наклона регрессионной прямой к оси  $X$ ).

#### **Условия применимости метода:**

- число объектов исследования должно быть в несколько раз больше числа прогностических (объясняющих) признаков;
- все анализируемые признаки (и независимые, и зависимый) должны быть количественными и нормально распределенными;
- зависимый признак  $Y$  должен иметь нормальные распределения с равными дисперсиями для каждого значения независимого признака (прогностического признака)  $X_i$ ;
- независимые признаки  $X_i$  могут быть количественными и/или качественными;
- взаимосвязи между каждым из независимых признаков  $X_i$  и зависимым признаком  $Y$  линейны в интервале изученных значений;
- каждое значение  $Y$  независимо от другого для каждого значения каждого  $X_i$ ;
- в случае множественного регрессионного анализа требуется отсутствие сильных линейных корреляций независимых признаков, которые предполагается включать в регрессионный анализ. Если какие-либо из независимых признаков сильно коррелированы, то необходимо включать в регрессионный анализ тот из них, который имеет более сильную корреляцию с зависимым признаком. При применении пошаговых процедур коррелированность прогностических признаков не так критична, поскольку из пары признаков, имеющих сильную корреляцию, алгоритм включает только один признак;

- величина  $E_i$  — разница между фактическим и прогнозируемым значением зависимой переменной, есть случайная величина с нормальным распределением и нулевым математическим ожиданием;
- все значения  $E_i$  не коррелированы между собой и имеют одинаковую дисперсию  $s^2$ .

**Замечание.** Степень устойчивости оценок коэффициентов регрессии в случае коррелированности (мультиколлинеарности) признаков во многом определяется алгоритмом включения прогностических признаков в уравнение. Если задано принудительное включение всех признаков, то при наличии мультиколлинеарности искажения будут более выражены, тогда как при использовании пошаговых процедур искажения будут менее выражены. Важно также учитывать возможность оценки уравнений разных типов, однако это специальный вопрос, который здесь не обсуждается.

Мы не будем также обсуждать способы проверки соответствия данных вышеперечисленным условиям. Заметим только, что существуют как формальные (например, проверка гипотез), так и неформальные (например, изучение графика остатков) способы проверки этих допущений. Иногда в данные, нарушающие эти предположения, могут быть внесены поправки (например, может быть выполнено преобразование данных) для достижения соответствия условиям применимости метода. Если такие преобразования сделаны, они должны быть описаны в публикации.

Как простой, так и множественный линейный регрессионный анализ в ППП **STATISTICA** выполняется в модуле “Множественная регрессия”.

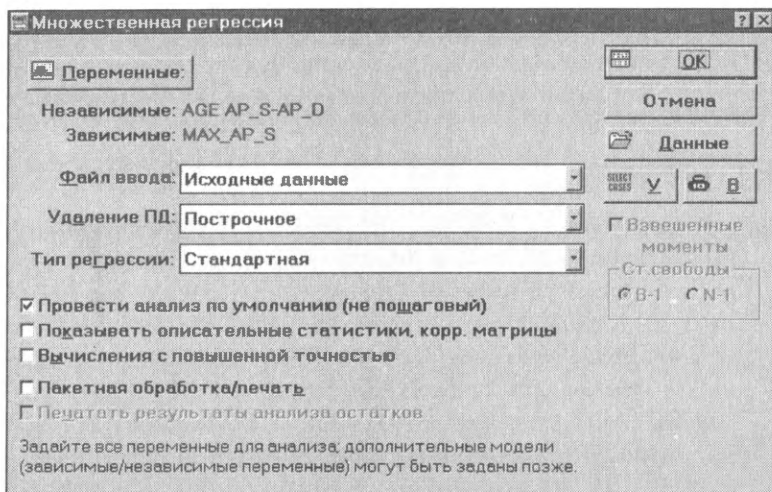
### **STATISTICA:**

→ Модуль “Множественная регрессия” (“Linear regression”)

Нажав на кнопку “Переменные” (рис. 14.1), следует в диалоговом окне указать независимые и зависимый признаки. Можно указать несколько зависимых признаков; в этом случае анализ будет выполняться последовательно для каждого зависимого признака (в режиме пакетной обработки<sup>1</sup>).

<sup>1</sup> Batch processing (англ.).





**Рис. 14.1.** Диалоговое окно линейного регрессионного анализа (модуль “Множественная регрессия”).

Необходимо также (в случае множественного регрессионного анализа) выбрать одну из процедур отбора признаков. В ППП **STATISTICA** реализованы три такие процедуры:

- 1) стандартная (Standard);
- 2) прямая пошаговая (Forward stepwise);
- 3) обратная пошаговая (Backward stepwise).

При выполнении стандартной процедуры множественного регрессионного анализа признаки оцениваются независимо друг от друга и включаются в уравнение все, независимо от силы их относительного влияния на зависимый признак.

При использовании алгоритма прямой пошаговой процедуры вначале в уравнении отсутствуют прогностические признаки. Затем на первом шаге в уравнение регрессии включается тот прогностический признак, который имеет наибольший по модулю частный (парциальный) коэффициент корреляции с зависимой переменной. Далее с использованием уравнения регрессии с этим прогностическим признаком оценивают значения зависимой переменной. После этого из фактических значений зависимой переменной вычитают значения, прогнозируемые по этому уравнению регрессии. Затем оценивают парциальные (частные) коэффициенты корреляции полученных остатков и прогностических

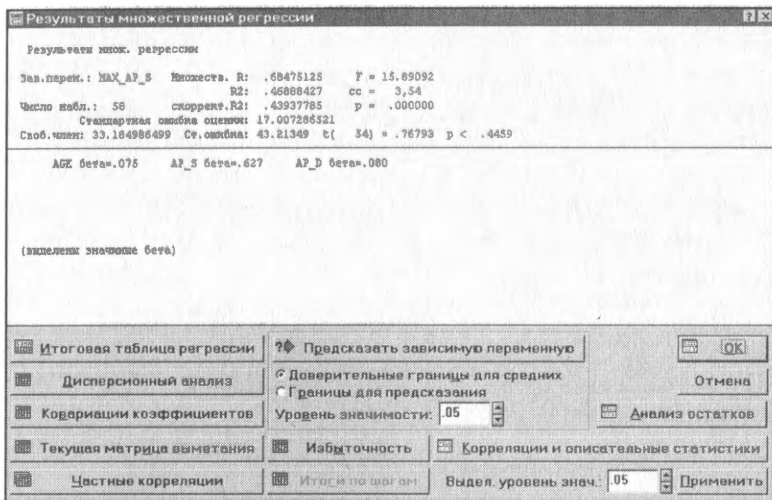
признаков, еще не вошедших в уравнение. На каждом из всех последующих шагов один прогностический признак либо включается в уравнение как наиболее информативный из оставшегося подмножества прогностических признаков, либо исключается из уже вошедших в уравнение как неинформативный признак. Таким образом, при прямой пошаговой процедуре на каждом шаге включается тот независимый признак, который в наибольшей (по сравнению с другими не включенными в модель признаками) степени (при уже включенных на более ранних шагах процедуры признаках) влияет на зависимый признак.

При использовании обратного пошаговой процедуры в уравнение регрессии вначале включаются (автоматически) принудительно все прогностические признаки, затем на первом шаге удаляется тот прогностический признак, который имеет минимальный частный коэффициент корреляции с зависимым признаком, и т.д.

В пошаговых процедурах обычно существуют следующие возможности:

- установить пороговый уровень статистики  $F$ , начиная с которого признаки будут включаться в модель. Это позволяет включать в модель только сильные прогностические (независимые) признаки;
- указать максимальное число шагов процедуры, на каждом из которых в модель включается один признак. Таким образом, можно ограничить число прогностических признаков в модели.

**Замечание.** При использовании пошаговых процедур включения прогностических признаков в уравнение регрессии обязательно просматривайте порядок включения и исключения прогностических признаков в уравнение. Далее, изменяйте значения “F-включить” (“F-enter”) и “F-исключить” (“F-remove”) таким образом, чтобы в конечном уравнении все вошедшие прогностические признаки имели приемлемое для Вас значение  $r$ , например в интервале от 0,05 до 0,10. При использовании алгоритма “Прямая пошаговая процедура” (“Forward stepwise”) целесообразно начинать процесс оценки регрессионного уравнения со значений “F-включить” (“F-enter”) = 2,0 и “F-исключить” (“F-remove”) = 1,9. Далее, обязательно сравнивайте уравнения, полученные с использованием прямой и обратной пошаговой процедуры.



**Рис. 14.2.** Второе диалоговое окно процедуры линейного регрессионного анализа.

Результаты построения модели выводятся во второе диалоговое окно (рис. 14.2).

В данном окне представлены основные расчетные характеристики модели:

- значение  $p$  для модели в целом (третья строка в первом окне результатов);
- стандартная ошибка оценки (standard error of estimate);
- коэффициенты  $beta$  для каждого из независимых (объясняющих) признаков — стандартизованные эквиваленты коэффициентов  $b$ ;
- значение константы  $a$  (свободный член, intercept);
- коэффициент детерминации  $r^2$  (в простом регрессионном анализе) или коэффициент множественной детерминации  $R^2$  (в множественном регрессионном анализе).

Из первого окна результатов могут быть запрошены следующие процедуры:

1. “Итоговая таблица регрессии” (“Regression summary”) (рис. 14.3). В окне результатов этой процедуры можно увидеть:
  - значения стандартных ошибок (второй столбец таблицы результатов) для каждого регрессионного коэффициента (первый столбец таблицы результатов);

Итоги регрессии для зависимой переменной: MAX_AP_S						
МНОЖЕСТВ РЕГРЕС.	R= .68475125 R2= .46888427 Скооррект. R2= .43937785 F(3,54)=15.891 p<.00000 Станд. ошибка оценки : 17.007					
N=58	БЕТА	Ст. ош. БЕТА	B	Ст. ош. B	t(54)	p-уров.
Св. член			33.18499	43.21349	.767931	.445873
AGE	.074707	.107489	.39172	.56361	.695022	.490022
AP_S	.627186	.151197	.66429	.16014	4.148128	.000119
AP_D	.079555	.158462	.15266	.30408	.502043	.617678

Рис. 14.3. Итоговая таблица регрессии.

— точные значения  $p$  для каждого из регрессионных коэффициентов (последний столбец таблицы результатов);

2. “Анализ остатков” (“Residual analysis”), где представлены многочисленные способы визуализации результатов регрессионного анализа:

- построение графиков остатков с помощью процедуры “График остатков” раздела “Гистограммы” (“Graph of residuals”). По графикам остатков можно качественно оценить выполнимость условий применимости метода (нормальности распределения остатков);
- построение графиков рассеяния (для простого регрессионного анализа) с использованием процедуры “Две переменные” (“Bivariate correlation”). В левой части окна необходимо указать независимый (объясняющий) признак, в правой части окна — зависимый (объясняемый) признак. Будут построены диаграмма рассеяния, линия регрессии и доверительная область для линии регрессии. При этом выводятся нестандартизованные коэффициенты  $b$  и значение константы  $a$ .

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- если значение  $p$  для уравнения регрессии (выводимое в первом окне результатов и рассчитанное для  $F$ -критерия Фишера) меньше 0,05, то можно считать, что полученная модель достаточно адекватно описывает взаимосвязь признаков;
- если  $p > 0,05$ , то регрессионная модель неудовлетворительно описывает взаимосвязь признаков;
- стандартная ошибка оценки модели (Standard error of estimate) показывает степень рассеяния разности между фактическими и прогнозируемыми значениями зависимой переменной;

- коэффициент детерминации  $r^2$  (в случае простого регрессионного анализа) является квадратом коэффициента корреляции и показывает, какая часть дисперсии зависимого признака может быть объяснена дисперсией независимого признака. Он имеет значения от 0 (нет связи признаков) до 1 (детерминирующая связь признаков);
- коэффициент множественной детерминации  $R^2$  (в случае множественного регрессионного анализа) показывает, в какой степени вариация зависимого признака определяется вариацией независимых признаков, включенных в модель. Чем выше  $R^2$ , тем лучше регрессионная модель;
- весовые коэффициенты  $b$  (*beta* — их стандартизованный вариант) являются коэффициентами уравнения регрессии. Они показывают, насколько сильно среднее значение зависимого признака  $Y$  варьирует с каждой единицей изменения независимого признака  $X$ . В случае простой линейной регрессии регрессионный коэффициент (наклон регрессионной прямой) является мерой связи между анализируемыми признаками  $X$  и  $Y$ . Нулевой наклон (горизонтальная линия на графике) означает отсутствие линейной связи между признаками. Поэтому в ходе регрессионного анализа проверяются гипотезы о равенстве нулю этих коэффициентов;
- если для какого-либо коэффициента  $b_i$ ,  $p_i < 0,05$ , то соответствующий признак  $X_i$  является значимым прогностическим для признака  $Y$  (в окне результатов выделен красным цветом).
- если  $p > 0,05$ , то соответствующий признак  $X_i$  не является значимым прогностическим для признака  $Y$  (в окне результатов записан синим цветом).

**Замечание.** Помните, что результаты регрессионного анализа применимы только к тому интервалу значений данных, на которых они получены. Например, если анализировалась возможность прогнозирования возраста, в котором начнется заболевание, по данным биохимических тестов для лиц 20—40 лет, и была получена статистически значимая регрессионная модель, то для лиц старше 40 лет использовать эту модель некорректно.

**Проверка работоспособности модели.** Результатом регрессионного анализа является создание статистической модели. Как у всякой математической модели, у регрессионной модели следует проверить работоспособность (подробнее см. раздел 13.6).

**Представление результатов.** Привести следующую информацию:

- число объектов исследования;
- описательную статистику (хотя бы  $M$  и  $s$ ) для зависимого и каждого независимого признаков;
- желательно привести результаты предварительного парного корреляционного анализа каждого из независимых признаков с зависимым признаком (с указанием значений  $r$  для каждого коэффициента корреляции), а также независимых признаков между собой;
- уравнение регрессии. Если в записи уравнения используются значения нестандартизованных регрессионных коэффициентов  $b$ , то необходимо включать в уравнение константу  $a$ . Если же используются стандартизованные регрессионные коэффициенты  $beta$ , то константа  $a$  не включается;
- список безразмерных коэффициентов  $beta$  в порядке убывания их значений, поскольку этот список позволяет судить о порядке прогностических признаков (независимых признаков) по силе их влияния на зависимый признак;
- точные значения  $r$  и стандартные ошибки для каждого из регрессионных коэффициентов;
- диаграмму рассеяния фактических и рассчитанных (по регрессионной модели) значений зависимого признака и график регрессионной прямой (в случае простого регрессионного анализа). Желательно также отразить линии, соответствующие границам ДИ вокруг линии регрессии.

## 14.2. Логистическая регрессия как один из методов нелинейного регрессионного анализа

**Задача:** исследовать связь бинарного признака с одним или несколькими количественными и/или качественными признаками (прогностическими признаками, факторами).

**Пример:** необходимо прогнозировать возникновение инфаркта миокарда по данным о курении, возрасте, наличии артериальной гипертензии.

**Решение:** логистический регрессионный анализ.

Логистический регрессионный анализ позволяет строить статистическую модель для прогнозирования вероятности наступле-

ния события по имеющимся данным (например, факторам риска). При этом имеющиеся данные по независимым признакам (в нашем примере — курение, возраст, наличие артериальной гипертонии) называются независимыми (объясняющими) признаками. Зависимым признаком  $Y$ , в отличие от линейной множественной регрессии, является не само значение зависимого (объясняемого) признака  $P$ , а его логит-преобразование:

$$\text{logit}(P) = \ln(P/(1-P)).$$

Логит-преобразование позволяет вместо значений  $P$ , определенных на интервале  $[0; 1]$  перейти к интервалу  $[-\infty; +\infty]$ . Вследствие этого возрастает чувствительность данной функции к изменению комбинаций, сочетаний разнообразных прогностических признаков.

Результатом логистического регрессионного анализа является расчет оценок регрессионных коэффициентов  $b_0, b_1, b_2, \dots, b_i$  уравнения

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_i X_i,$$

где  $X_1, \dots, X_i$  — независимые признаки (факторы риска),

$$Y = \ln(P/(1-P)).$$

#### Условия применимости метода:

- зависимый признак — качественный бинарный (имеет только два возможных значения: 1 (например, наличие заболевания) и 0 (отсутствие заболевания));
- независимые (объясняющие) признаки — любые (количественные и/или качественные);
- независимые (объясняющие) признаки должны быть независимы также друг от друга, т.е. не должны быть коррелированы или ассоциированы (рекомендуется предварительно провести соответствующий анализ).

**Замечание.** В некоторых других пакетах статистических программ предусмотрена возможность использования зависимых переменных с числом значений более двух, а также с большим набором алгоритмов оценки уравнения регрессии.

#### STATISTICA:

→ Модуль “Нелинейное оценивание”  
 (“Nonlinear estimation”)

→ Раздел “Логит-регрессия”<sup>1</sup>

(“Logistic regression”) (рис. 14.4)

<sup>1</sup> Имеется в виду логистическая регрессия, или логистический регрессионный анализ.

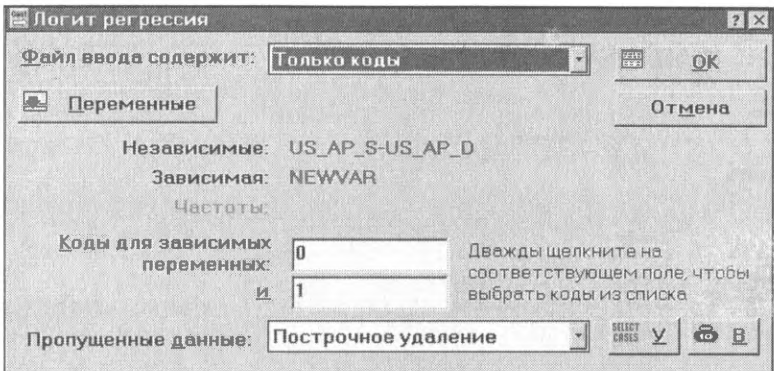


Рис. 14.4. Первое диалоговое окно для проведения логистического регрессионного анализа (модуль “Нелинейное оценивание”, процедура “Логит-регрессия”).

Этапы анализа визуализируются последовательно в отдельных окнах (рис. 14.5—14.7). В последнем из этих окон следует прежде всего активизировать процедуру “Оценки параметров” (“Parameters estimates”). Результаты оценки коэффициентов уравнения ( $b_0, b_1, \dots, b_i$ ) будут выведены в отдельное окно (рис. 14.8).

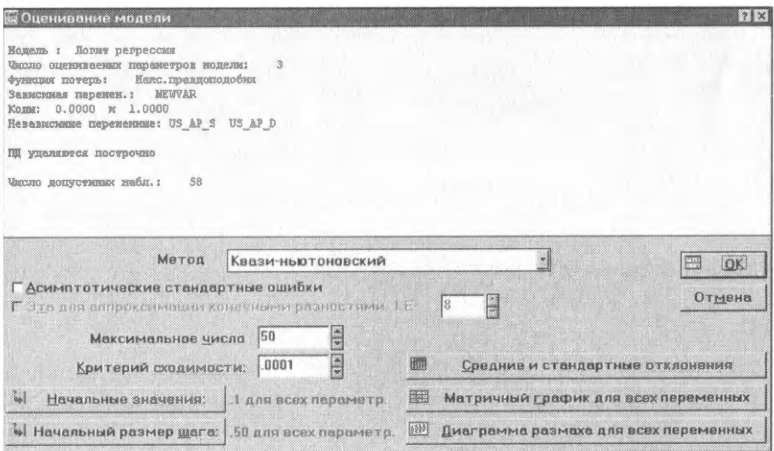


Рис. 14.5. Первое окно промежуточных результатов логистического регрессионного анализа.



Итерация	Потери	Параметры		
к 13	39.1623	-1.4021	.030474	-.03416
к 14	39.1623	-1.4021	.030474	-.03416
к 15	39.1541	-1.7618	.031815	-.03239
к 16	39.1540	-1.8071	.031852	-.03196
к 17	39.1540	-1.8073	.031852	-.03196
к 18	39.1540	-1.8073	.031852	-.03196
к 19	39.1540	-1.8073	.031852	-.03196

Процесс оценивания параметров сошелся

Рис. 14.6. Второе окно промежуточных результатов логистического регрессионного анализа.

Модель : Логит регрессия  
 Число 0 : 29.00000 (50.00000%)  
 Число 1 : 29.00000 (50.00000%)  
 Зависимая перемен.: NEWVAR  
 Независимые переменные: 2  
 Функция потерь: Макс. правдоподобия  
 Оконч. знач. 1 39.154025544  
 -2\*log(Правдопод.) для этой модели= 78.30905 со св.ч. только: 80.40508  
 хи-квадрат = 2.097022, ос = 2, р = .3504713

Оценки параметров	<input checked="" type="checkbox"/> Подогнанная 2М функция и набл. значения	OK
Ковариации/корр. параметры	<input checked="" type="checkbox"/> Подогнанная 3М функция и набл. значения	Отмена
<input checked="" type="checkbox"/> Масштаб средних ошибки 1	Довер. интервал: 95.0 %	<input checked="" type="checkbox"/> Распределение остатков
<input checked="" type="checkbox"/> Остатки	<input checked="" type="checkbox"/> Предсказанные значения	<input checked="" type="checkbox"/> Нормальный график остатков
<input checked="" type="checkbox"/> Наблюдаемые значения	<input checked="" type="checkbox"/> Полунормальный график остатков	
<input checked="" type="checkbox"/> Классификация и отношение несогласия	<input checked="" type="checkbox"/> Предсказанные и наблюдаемые значения	
<input checked="" type="checkbox"/> Средние и стандартные отклонения	<input checked="" type="checkbox"/> Предсказанные значения и остатки	
<input checked="" type="checkbox"/> Отличие (от предыдущей модели)	<input checked="" type="checkbox"/> Матричный график для всех переменных	
<input checked="" type="checkbox"/> Сохранить предсказанные и остатки	<input checked="" type="checkbox"/> Диаграмма размаха для всех переменных	

Рис. 14.7. Третье окно промежуточных результатов логистического регрессионного анализа.

Модель. Логит регрессия			
Далее...	Зав. перем.: NEWVAR Потери: Макс. правд. Ок. потери: 39.154025544 Хи2 ( 2)=2.0970 р=.35047		
N=58	Конст. В0	US_AP_S	US_AP_D
Оценка	-1.80725	.03185	-.031956
Отн. несогл. -ед. изм.	.16410	1.03236	.968549
Отн. несогл. -размах		12.78362	.278526

Рис. 14.8. Окно результатов логистического регрессионного анализа: оценка коэффициентов регрессионного уравнения.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Значения коэффициентов  $b_i$  представляют собой натуральные логарифмы отношения шансов (ОШ) для каждого из соответствующих признаков  $X_i$ . Для  $i$ -го признака ОШ может быть вычислено по формуле:  $ОШ_i = e^{b_i}$ , где  $e$  — математическая константа, приблизительно равная 2,72.
- Значение  $Y$  в уравнении регрессии — это натуральный логарифм ОШ для изучаемого события. Вероятность этого события для каждого конкретного объекта исследования может быть вычислена по формуле:  $P = e^y / (1 + e^y)$ .
- Если для какого-либо из коэффициентов  $p > 0,05$ , то нулевая гипотеза о том, что  $ОШ_i = 1$ , т.е. исследуемый фактор (объясняющий признак) не ассоциирован с изучаемым событием, не отклоняется.
- Если  $p < 0,05$  для какого-либо из коэффициентов, то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о том, что исследуемый фактор (объясняющий признак) ассоциирован с изучаемым событием (бинарным признаком).

Остановимся несколько подробнее на интерпретации ОШ в логистической регрессионной модели. Эта интерпретация такова: при увеличении на единицу значения  $i$ -го признака шанс возникновения изучаемого события увеличивается в число раз, равное величине ОШ. Для бинарного независимого признака ОШ — это отношение шанса того, что событие произойдет в одной группе, к шансу того, что оно произойдет в другой группе.

*Пример:* под наблюдением находились две группы участников исследования: курильщики и некурящие. Подсчитывалось число случаев возникновения инфаркта миокарда в обеих группах. Оба признака — бинарные, данные могут быть представлены в четырехпольной таблице (рис. 14.9):

Исследуемые группы	Инфаркт возник	Инфаркта не было
Курильщики	140	220
Некурящие	50	330

Рис. 14.9. Таблица сопряженности (пример).

Для курильщиков шансы развития инфаркта миокарда составляют  $140/220 = 0,636$ , а для некурящих —  $50/330 = 0,152$ ;

$OШ=0,636/0,152=4,2$ , т.е. шансы развития инфаркта миокарда у курильщиков в 4,2 раза больше, чем у некурящих.

Если бы  $OШ$  равнялось 1, то это означало бы, что обе группы имеют одинаковые шансы на возникновение инфаркта миокарда. Чем больше  $OШ$ , тем больше вероятность возникновения события, которое случится в группе, описанной в числителе  $OШ$ .

**Представление результатов.** Привести следующую информацию:

- описательную статистику для каждого из анализируемых признаков;
- число объектов исследования;
- точное значение  $p$  для модели в целом;
- регрессионные коэффициенты  $b_i$  для независимых признаков  $X_i$  и точные значения  $p$  для каждого из них.

**Проверка работоспособности модели.** Результатом регрессионного анализа является создание статистической модели. Как у всякой математической модели, у регрессионной модели следует проверить работоспособность (подробнее см. раздел 13.6).

---

# Глава 15. Некоторые общие проблемы и частные задачи статистического анализа

## 15.1. Проблема множественных сравнений

При проведении анализа данных часто возникает так называемая проблема множественных сравнений (ПМС), заключающаяся в следующем: чем больше статистических гипотез проверяется на одних и тех же данных, тем более вероятна ошибка первого рода — заключение о наличии различий между группами, в то время как на самом деле верна нулевая гипотеза об отсутствии различий. Так, если за уровень значимости принято значение  $p_0 = 0,05$ , то 5 из 100 вычисленных значений  $p$  в силу случайности (по теории вероятности) окажется меньше 0,05 (хотя на самом деле верна нулевая гипотеза об отсутствии различий). На практике принято считать, что учет ПМС следует начинать в тех случаях, когда число рассчитываемых значений  $p$  (и соответственно публикуемых в статьях и диссертациях) превышает 10.

ПМС возникает в следующих ситуациях.

**1. При вторичном (post hoc) анализе данных.** Как ранее упоминалось, статистический анализ данных можно разделить на два типа: первичный и вторичный. Первичный анализ данных проводится для проверки предположений исследователя, которые возникли у него до начала сбора данных. Вторичный анализ проводится для поиска неизвестных заранее закономерностей в данных. Практически всегда у исследователя возникает соблазн проанализировать с большим трудом полученные данные всеми возможными способами. Однако необходимо понимать,

что такой анализ обычно носит исследовательский характер и не отвечает на вопросы (не проверяет гипотезы) исследователя, а ставит их (выдвигает гипотезы). Интерпретация результатов вторичного анализа данных должна быть очень осторожной и прежде всего опираться на биологическое правдоподобие найденных закономерностей. “Если рыболов вылавливает ботинок, ботинок нужно выбросить, а не объявлять, что рыболов ловил именно ботинки”. Вторичный анализ данных обычно проводится путем анализа всех возможных взаимосвязей в поисках статистически значимых закономерностей. Его результаты не могут считаться вполне статистически значимыми (даже если получены статистически значимые закономерности), так как структура исследования не предусматривала исследование этих закономерностей, и следовательно, выборки могут быть неадекватными (смещенными). Например, выборки могут быть исходно не сбалансированы по некоторым факторам при исследовании вмешательства или могут быть слишком малочисленными для получения надежных статистических выводов. Наиболее часто такой вторичный анализ данных бывает связан с анализом большого числа клинических подгрупп.

**2. При множественных парных сравнениях групп и подгрупп** (по демографическим и клиническим характеристикам, исходам, временным точкам и т.д.). Такой анализ обоснован, если он запланирован в начале исследования, до начала сбора данных, т.е. является первичным. В этом случае подгруппы скорее всего будут достаточны по объему, а их выделение будет иметь биологическое обоснование. В общем случае при планировании исследования, предусматривающего большое количество сравнений групп и подгрупп, следует использовать специальные программы оценки мощности (чувствительности) статистических тестов, а также производить оценку необходимого объема выборок для проверки конкретных гипотез. Если же анализ в подгруппах носит поисковый характер, его результаты должны быть описаны лишь как предварительные. В этой ситуации необходимо пользоваться процедурами группового сравнения, например ANOVA, а в дальнейшем специальными процедурами множественных парных сравнений (см. раздел 8.4.1.3). Наиболее частой ошибкой при множественных парных сравнениях является многократное применение *t*-критерия Стьюдента без необходимой в этих случаях переустановки критического уровня значимости

(обычно принимаемого равным 0,05). Нередко также вычисляется множество (иногда сотни) коэффициентов корреляции, выполняется повторное тестирование на значимость различий по ходу пополнения набора сравниваемых групп.

**3. При установлении эквивалентности групп в начале нерандомизированного исследования вмешательства.** В этом случае бывает необходимо обнаружить дисбаланс между группами, который может возникнуть за счет систематической ошибки при формировании групп. В рандомизированных же исследованиях, если рандомизация проведена успешно, неэквивалентность групп может быть только случайной, и сравнение групп не требуется, хотя описательную статистику групп по основным признакам на начальном этапе исследования необходимо приводить в отчете.

**4. При промежуточном анализе данных, полученных в испытаниях тех или иных вмешательств.** Такой анализ часто требуется для того, чтобы не подвергать участников исследования дальнейшему необоснованному риску. Этот анализ связан с изначальным установлением “критериев остановки” исследования. Если промежуточный анализ показывает, что вмешательство очень эффективно или, наоборот, очевидно неэффективно либо вредно, исследователи могут прервать исследование. Результаты предварительного анализа рекомендуется использовать только для принятия решения о завершении/продолжении исследования, но не для информирования участников исследования или широкого опубликования, так как в этом случае дальнейший ход исследования может быть нарушен (смещен). Однако промежуточный анализ увеличивает число проводимых тестов и является таким образом источником ПМС.

Для того чтобы определить максимально допустимое число сравнений в данном исследовании, можно воспользоваться следующей формулой [26]:

$$k(k-1)/2,$$

где  $k$  — число групп для сравнения.

Существуют следующие **подходы к преодолению ПМС:**

- установление более жесткого, чем 0,05, уровня значимости, например 0,01; 0,005; 0,001 и т.д.;
- применение поправки Бонферрони — перерасчет уровня значимости  $p$  для множественных парных сравнений по формуле  $p_0/n$ , где  $p_0$  — исходно заданный уровень статистической

значимости (например, 0,05),  $n$  — количество парных сравнений. Обычно такой уровень статистической значимости бывает трудно достичь, если число сравнений больше 10. Другой способ применения поправки Бонферрони — умножение вычисленных значений  $p$  на  $n$  и установление этого произведения в качестве уровня значимости;

- придание большего значения результатам первичного анализа данных (оригинальным априорным гипотезам) и меньшего значения результатам вторичного анализа при интерпретации результатов.

## **15.2. Определение интервала нормы (референтного интервала) для количественного признака**

Понятие нормы в медицине всегда было предметом дискуссий. В России приняты определенные методики определения границ референтных интервалов для лабораторных показателей (например, [35]). Однако в современной медицине существуют разные подходы к определению нормы:

- клинические (диагностическое, терапевтическое определение);
- статистические (гауссово, процентильное);
- эпидемиологические (по фактору риска, социальное).

Наиболее обоснованными в настоящее время считаются терапевтические подходы.

**Диагностическое определение нормы.** Норма — это интервал значений количественного признака, которому соответствуют случаи без заболевания, а значениям вне которого соответствуют случаи заболевания. Это определение нормы должно базироваться на доказательстве присутствия или отсутствия заболевания в интервалах нормы и ненормы соответственно. Диагностическое определение показывает вероятность заболевания для каждого значения теста и определяется как интервал значений признака в здоровой популяции, не перекрывающийся с интервалом значений признака в популяции больных.

**Терапевтическое определение нормы** определяет интервал значений признака, при которых лечение не назначается (из-за неэффективности или даже опасности) и вне которого лече-

ние полезно. Такое определение, будучи основано на доказательстве эффективности лечения, является клинически целесообразным. Оно показывает полезность лечения для каждого конкретного результата теста. Положительный результат теста изменяет ведение больного, например, является основанием для назначения какого-либо препарата, который, как показано в исследованиях, эффективен только при значениях теста выше некоторой пороговой величины.

Давно используются и широко распространены **статистические подходы** к определению нормы. Они соответствуют пониманию нормы как наиболее распространенного, частого варианта. Эти подходы следующие:

— **Гауссово определение нормы** основано на измерениях, предпринятых на здоровой популяции. В случае если данные подчиняются закону нормального распределения, интервал нормы определяется как интервал измерений, включающий по два средних квадратических отклонения выше и ниже среднего значения. Таким образом, интервал включает центральные 95% всех измерений. Однако верхние и нижние 2,5% объектов (не норма) могут не иметь клинических проявлений, это просто необычные значения. Большинство лабораторных тестов крови определяют норму именно таким образом. Данное определение строится на предположении, что результаты тестов нормально распределены, но, к сожалению, в реальности результаты тестов редко являются нормально распределенными. Иногда, однако, данные могут быть математически преобразованы в целях получения распределения, более приближенного к нормальному. Значения интервала нормы затем преобразовываются обратно в исходные единицы в целях клинического использования. Недостатком данного подхода к определению нормы является то, что необычные результаты не обязательно означают заболевание, а нормальные не обязательно означают его отсутствие.

— **Процентильное определение нормы** выражает интервал нормы как интервал значений между установленными произвольно нижним или верхним процентилями общего диапазона. Например, интервал, включающий 95% всех нижних результатов теста, может быть определен как нормальный, и только 5% выше могут быть определены как ненорма. В таком определении также нет клинического смысла, а только стати-



стический. Могут быть использованы и 90-, и 99-процентильные интервалы.

**Замечание.** Часто в клинических исследованиях авторы, не набирая контрольную группу, полагаются в качестве контроля на те границы нормы, которые приняты в том или ином учреждении. Такой подход чреват существенными систематическими ошибками. Так, границы нормы могут быть определены на контингенте людей, не сопоставимых с исследуемыми по возрасту и полу. Не обсуждая здесь заведомо более низкое качество неконтролируемого исследования (исследования без группы контроля) по сравнению с контролируемым, отметим, что для проведения статистического анализа — сравнения изучаемой группы с нормой — необходимо знать, как именно рассчитан интервал нормы для того или иного лабораторного признака. Ведь он может быть 90-, 95-, 99-процентильным интервалом либо интервалом, границы которого определены каким-либо другим способом. В каждом из этих случаев результаты статистического сравнения могут оказаться различными. Проконсультироваться по вопросу о том, каково происхождение границ нормы, следует, например, у врача-лаборанта. Иногда можно обнаружить, что эти “нормы” заложены в программу компьютера, управляющего биохимическим анализатором, в таком случае происхождение “норм” часто оказывается неизвестным.

**Эпидемиологические определения нормы** менее полезны для принятия клинических решений, хотя они в настоящее время имеют много сторонников. К эпидемиологическим определениям нормы относятся следующие.

— **Определение нормы по фактору риска.** При этом норма включает интервал значений признака, в котором риск развития заболевания не повышен и ниже которого риск снижен. Это определение строится на предположении, что изменение фактора риска изменяет действительный риск заболевания. Например, высокий уровень холестерина не опасен сам по себе, но так как повышенный риск развития ишемической болезни сердца ассоциирован с высоким уровнем холестерина, это делает повышенный уровень холестерина “ненормальным”. Поскольку обычно зависимость риска развития заболевания от фактора риска возрастает постепенно, установление “границы высокого риска” абсолютно условно.

— **Социальное определение нормы** основано на распространенных представлениях о норме и ненорме. Например, желательная масса тела или “правильное” время прорезывания зубов у ребенка часто имеют социальный смысл, который может иметь, но может и не иметь клинического значения.

На рис. 15.1 проиллюстрированы некоторые подходы к определению нормы.

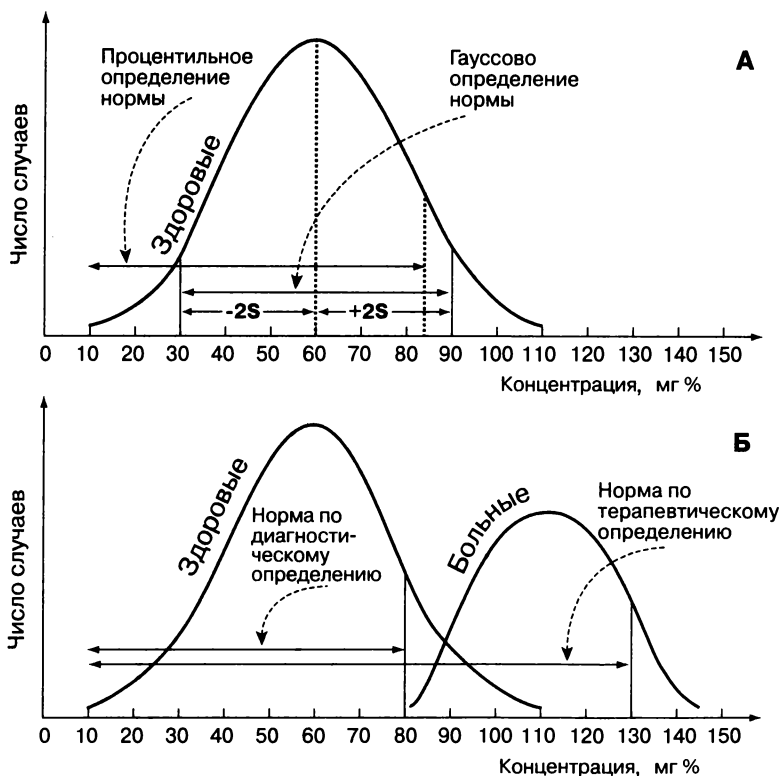


Рис. 15.1. Иллюстрации нескольких определений нормы. А — статистические (гауссово, процентильное); Б — клинические (диагностическое — интервал до начала распределения значений теста в популяции больных, терапевтическое — не связано жестко с самим распределением, а связано с результатами других исследований по эффективности вмешательств) [35].

### 15.3. Определение необходимых объемов выборок при планировании исследования. Статистическая мощность (чувствительность) исследования

Определение объемов выборок необходимо при планировании контролируемых исследований, т.е. исследований, в которых сопоставляются группы (выборки) независимых наблюдений. Это могут быть клинические испытания, одномоментные исследования. Для некоторых типов исследований требуются более сложные вычисления, чем приведенные ниже. В этих случаях мы рекомендуем обращаться к статистике.

Основной смысл определения объема выборок заключается в том, чтобы иметь в результате исследования высокую вероятность выявления реальных различий между выборками как статистически значимых, т.е. обеспечить необходимую статистическую мощность (чувствительность) исследования.

При этом необходимо помнить о различиях между статистической и клинической значимостью различий между выборками (подробно см. раздел 5.4). Например, можно заранее поставить следующую задачу при испытании нового метода лечения: для того чтобы изучаемый метод лечения был принят в клинической практике (т.е. быть клинически значимым), он должен превосходить старый метод по исходам, например на 30%.

Подчеркнем: в процедуре определения объемов выборок предполагается, что заранее известно, по какому признаку мы надеемся обнаружить различия между выборками.

**Статистическая мощность (чувствительность) исследования** — это вероятность того, что при проверке какой-либо статистической гипотезы (например, о равенстве средних, дисперсий, коэффициентов корреляции, долей и т.д.) исследование с данными объемами выборок выявит как статистически значимое реально существующее различие между выборками. Иными словами, статистическая мощность (чувствительность) — это вероятность правильного отклонения нулевой гипотезы в случае, когда на самом деле верна альтернативная гипотеза.

Статистическая мощность исследования равна  $1 - \beta$ , где  $\beta$  — вероятность ошибки второго рода (т.е. вероятность ошибочного

принятия нулевой гипотезы об отсутствии различий между группами), допускаемая в исследовании. Если  $\beta$  принимается равной 0,2 (20%), то в этом случае статистическая мощность исследования равна  $1-0,2=0,8$  (80%).

Чем больше статистическая мощность исследования, тем больше уверенность в том, что имеющиеся в реальности различия между выборками выявлены в ходе исследования. Однако для обеспечения большой статистической мощности исследования требуются очень большие выборки.

Для нахождения необходимого общего объема двух выборок используется сложная формула, однако можно также использовать номограмму (рис. 15.2), входными параметрами для которой являются следующие:

- 1) необходимая исследователю статистическая мощность исследования; ее значение обычно устанавливается на уровне 80—90% (0,8—0,9);
- 2) уровень статистической значимости различий между выборками, например 0,05 или 0,01;
- 3) стандартизованное различие (standardized difference) исследуемого статистического параметра в сравниваемых группах.

Минимальное стандартизованное различие между выборками вычисляется разными способами для данных разных типов (о типах данных см. подробнее в главе 2).

Для **количественных данных** минимальное стандартизованное различие определяется как отношение  $d/s$ , где  $d$  — абсолютное значение минимального среднего различия между группами, имеющего клиническую значимость,  $s$  — известное априори, из ранее проводившихся исследований СКО изучаемого признака у данной категории пациентов.

Если априорная оценка СКО невозможна, то его значение рассчитывают в ходе начального этапа исследования по уже полученным наблюдениям. Существуют и другие способы, но они более сложные, и при необходимости их использовать рекомендуется обращаться к статистикам.

Для динамических исследований (исследований одних и тех же объектов в разные моменты времени) расчет стандартизованного различия производится по формуле  $2d/s$ . В данном случае  $s$  — это СКО средней разности признака.

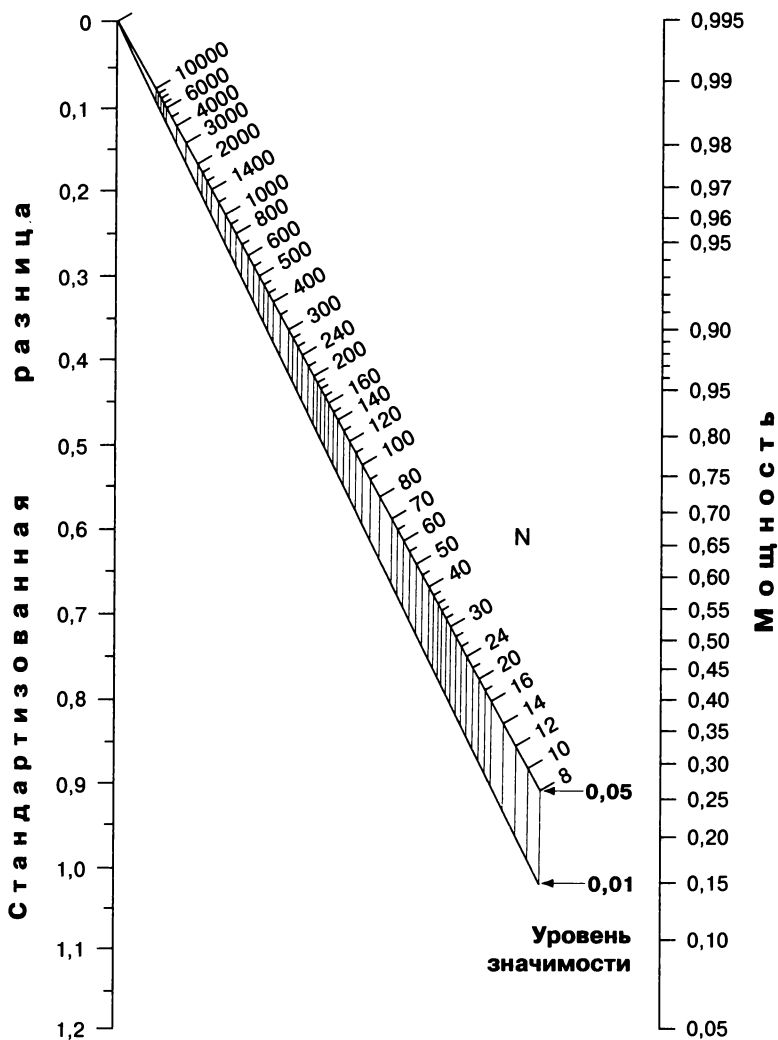


Рис. 15.2. Номограмма для вычисления объемов выборок при планировании исследования.

Для **качественных (бинарных) данных** (при исследовании бинарных исходов лечения, например “улучшение” — “отсутствие улучшения”) стандартизованное различие определяется как

$$\frac{P_1 - P_2}{\sqrt{\frac{P_1 + P_2}{2} \left(1 - \frac{P_1 + P_2}{2}\right)}}$$

где  $P_1$  и  $P_2$  — относительные частоты (доли) желательных исходов соответственно в 1-й и 2-й группах.

Полученная по номограмме величина объема выборки должна быть разделена на 2, если планируется изучать две равные по объему группы. Если же соотношение объемов групп отлично от 1, то общий объем групп  $N_1$  следует рассчитывать по следующей формуле [26]:

$N_1 = N(1+k)^2/4k$ , где  $N$  — полученный по номограмме общий объем выборок,  $k$  — необходимое соотношение объемов групп  $n_1/n_2$ .

Объемы же выборок будут определяться соответственно формулами:

$$\begin{aligned} n_1 &= N_1 / (1+k) \\ n_2 &= kN_1 / (1+k) \end{aligned}$$

**Замечание 1.** Если необходимо вычислить статистическую мощность закончившегося исследования, можно воспользоваться этими же формулами и номограммой для решения обратной задачи. При этом входными параметрами будут стандартизованное различие, уровень значимости и объемы выборок.

**Замечание 2.** Бесплатные программы для вычисления статистической мощности исследования можно получить на следующих страницах Интернета:

<http://members.aol.com/jobnp71/javasta2.html>

<http://www.mp1-pwrc.usgs.gov/powcase/powlinks.html>

Можно также воспользоваться программами на интерактивных страницах Интернета:

<http://www.math.yorku.ca/SCS/Demos/power>

<http://ebook.stat.ucla.edu/calculators/powercalc>

**Замечание 3.** В настоящее время считается, что если в ходе планируемого исследования невозможно обеспечить сбор необходимого числа наблюдений, то такое исследование должно быть запрещено по этическим соображениям. Это прежде всего касается испытаний лекарственных средств. Для преодоления проблемы набора большого числа наблюдений организуются многоцентровые исследования, в которых участвуют несколько учреждений, иногда из разных государств.

## 15.4. Анализ времени до наступления исхода

**Задача:** оценить вероятность наступления исхода (например, возникновения рецидива, наступления выздоровления, смерти) в разные периоды заболевания в зависимости от примененного метода лечения.

**Пример:** оценить вероятность выживания при комбинированном и хирургическом методах лечения рака молочной железы.

**Решение:** применение методов анализа времени до наступления исхода (анализ выживаемости).

Особые подходы к решению такой задачи анализа данных связаны с особенностями самих данных в исследованиях такого рода. Исследуемые исходы наступают в исследуемой когорте больных на протяжении длительного периода времени, обычно значительно превышающего возможные сроки исследования. В связи с этим дождаться, пока исследуемый исход наступит у всех больных (например, все больные умрут), обычно не представляется возможным. Поэтому к моменту окончания исследования имеются три типа наблюдений:

1) наблюдения, в которых изучаемый исход наступил, — **завершенные, полные** (так называемые **нецензурированные**) наблюдения. При этом интервал времени между включением больного в исследование и наступлением исхода известен;

2) наблюдения, в которых изучаемый исход не наступил на момент окончания исследования — **незавершенные, неполные** (так называемые **цензурированные**) наблюдения. При этом интервал времени между включением больного в исследование и изучаемым исходом неизвестен;

3) наблюдения, о которых не известно, наступил изучаемый исход или нет, т.е. выбывшие из исследования по причинам, не связанным с изучаемым исходом (в том числе по причине смерти от других причин, например несчастного случая), — также называются **неполными (цензурированными)** наблюдениями. Интервал времени между включением больного в исследование и изучаемым исходом также неизвестен, зато известен интервал времени между включением и выбыванием участника из исследования.

Как отмечалось уже выше, исследования такого рода могут быть проведены корректно только на синхронизированной ко-

горте — выборке больных, каждый из которых включен в исследование в строго определенный момент своего заболевания. Такими моментами могут быть появление симптомов, постановка диагноза, госпитализация, начало лечения, окончание послеоперационного периода и др. Важно лишь, чтобы для всех больных изучаемой когорты это был один и тот же критерий момента включения. Именно время от момента включения (относительное время) и подлежит анализу. Зачастую больные включаются в исследование на протяжении нескольких лет. Исследования, в которые участники включаются на разных этапах заболевания, могут быть несопоставимы вследствие смещения, вызванного, например, тем, что пациенты, у которых заболевание выявлено на более раннем этапе, обычно характеризуются большим числом лет дожития не в связи с лучшим методом лечения, а только потому, что имеющееся у них заболевание раньше диагностировано, и соответственно раньше было начато лечение. В разделе 3.1.6 описаны особенности подготовки данных к анализу такого рода.

Основными требованиями для обеспечения корректности анализа времени до наступления исхода являются следующие:

- выборка случайна (случайно выбирается из популяции) или по крайней мере репрезентативна по отношению к популяции;
- наблюдения независимы;
- в период исследования не происходило изменений в диагностике, лечении или процедурах наблюдения, т.е. критерии включения не изменялись;
- вероятность наступления изучаемого исхода не изменялась в течение исследования для всех исследуемых больных;
- случаи смерти, выбывания из исследования и включения больных происходили равномерно на всем периоде наблюдений.

Задачами анализа данных в таких случаях могут являться следующие:

1. Анализ вероятности наступления изучаемого исхода в определенный период времени (выживания; см. раздел 15.4.1).
2. Исследование влияния одного фактора на время до наступления изучаемого исхода (см. раздел 15.4.2).
3. Исследование влияния нескольких факторов на время до наступления изучаемого исхода (см. раздел 15.4.3).



**Важное замечание.** *Выживание не всегда является адекватной характеристикой качества метода лечения в медицине. Например, некоторые реанимационные технологии не столько продлевают жизнь, сколько увеличивают срок биологического существования больного. Осмысленное использование медицинских технологий требует оценки их влияния не только на выживание, но на выживание при сохранении определенного качества жизни.*

### **15.4.1. Анализ вероятности наступления изучаемого исхода в определенный период времени (выживания)**

В анализ включаются значения двух столбцов таблицы данных: зависимым признаком является время до наступления изучаемого исхода, а независимым — показатель законченности наблюдения (индикатор цензурирования).

Анализ выживания можно осуществлять двумя методами:

- 1) с помощью таблиц дожития (актуарный метод). Данный метод используется в случае, если выборка достаточно велика (обычно при эпидемиологических исследованиях). При этом дату наступления изучаемого исхода записывают лишь с точностью до определенного периода времени (месяца, года);
- 2) методом Каплана—Мейера. Метод используют в случае, если выборка не очень велика. При этом анализируют точные интервалы до наступления исхода в каждом наблюдении. Этот метод рекомендуется применять в клинических исследованиях.

#### **STATISTICA:**

→ Модуль “Анализ выживаемости”  
 (“Survival Analysis”) (рис. 15.3)

→ Подмодуль “Таблицы и распределения времен жизни” (“Life Tables”)

→ Процедура “Метод множительных оценок Каплана—Мейера”  
 (“Kaplan & Meier product-limit method”) (рис. 15.4)

В обоих случаях в диалоговом окне необходимо указать столбцы, содержащие индикатор цензурирования (с кодами завершенных и незавершенных наблюдений) и анализируемое время до наступления исхода (время жизни).

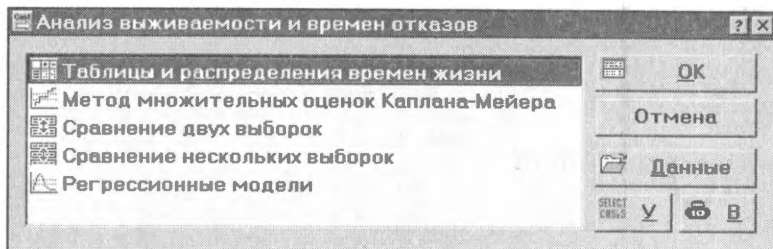


Рис. 15.3. Процедуры анализа выживаемости.

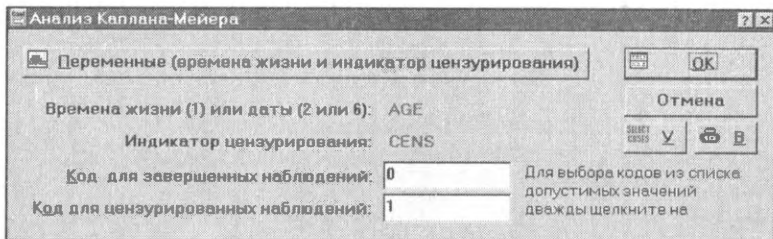


Рис. 15.4. Первое диалоговое окно процедуры “Метод множительных оценок Каплана—Мейера”.

После выполнения первого этапа анализа по методу Каплана—Мейера выводится промежуточное диалоговое окно (рис. 15.5). Основные результаты могут быть получены при обращении к оп-

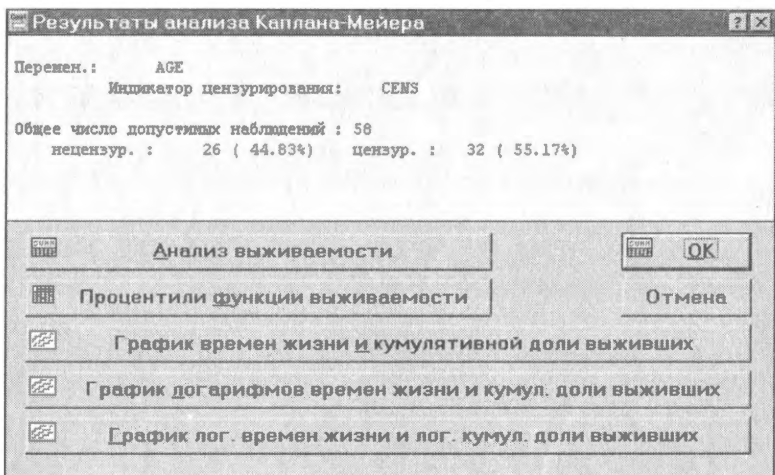


Рис. 15.5. Промежуточное диалоговое окно анализа по методу Каплана—Мейера.

Анализ множит. оценок Каплана-Мейера

Замечание: цензурированные отмечены +

Номер Имя	Номер Набл.	Времена	Кумул. Времена	Станд. Ошибка
Karnauro	2+	0.00000	--	--
Starovoy	1+	0.00000	--	--
Demiynov	53	47.00000	.982143	.017697
Frolov A	4+	48.00000	--	--
Bogomolo	33	49.00000	.963955	.025027
Kuharev	45+	49.00000	--	--
Sheluhin	23	50.00000	.945417	.030652
Rogozin	55	50.00000	.926880	.035213
Kozevnik	37+	50.00000	--	--
Usevich	57	53.00000	.907964	.039248
Barinov	10+	53.00000	--	--
Kuvarkov	29	55.00000	.888646	.042904
Necretin	41+	55.00000	--	--
Obruchni	12	56.00000	.868898	.046273
Golubcov	19+	56.00000	--	--
Inshakov	5+	56.00000	--	--
Marchenk	46+	56.00000	--	--
Gudkov B	13	58.00000	.847705	.049761
Trakov T	25	58.00000	.826512	.052928

Рис. 15.6. Результаты анализа выживаемости по методу Каплана—Мейера.

Таблица времен жизни

Логарифм правдоподобия : -54.9134

Интервал	Кум. доля выживш.	Плот. вероятн.	Интенс. отказов	Стд. ош. кум. выж.	Стд. ош. плот. вер	Стд. ош. ИнтОтказ	Медиана ож. врем.	Стд. ош. ож. врем.
Инт. N1	1.000000	.001508	.001514	0.000000	.002123	.002142	61.43970	.944621
Инт. N2	.991228	.001521	.001541	.012351	.002142	.002180	55.68407	.944658
Инт. N3	.982378	.001508	.001541	.017467	.002123	.002180	49.92901	.936224
Инт. N4	.973607	.001494	.001541	.021264	.002104	.002180	44.17338	.927864
Инт. N5	.964914	.001481	.001541	.024370	.002085	.002180	38.41719	.919580
Инт. N6	.956298	.001468	.001541	.027027	.002067	.002180	32.66045	.911369
Инт. N7	.947760	.001454	.001541	.029360	.002048	.002180	26.90317	.903232
Инт. N8	.939298	.001441	.001541	.031442	.002030	.002180	21.14533	.895168
Инт. N9	.930911	.011743	.013095	.033323	.005668	.006543	15.38696	.899301
Инт. N10	.862588	.023062	.028991	.045109	.008100	.010919	10.05605	.917048
Инт. N11	.728407	.070109	.133681	.060191	.013713	.032915	5.19481	1.038961
Инт. N12	.320499	--	--	.077011	--	--	--	--

Рис. 15.7. Результаты анализа выживаемости с помощью таблиц дожития.

циям “Анализ выживаемости” (“Product-limit survival analysis”) и “Процентили функции выживаемости” (“Percentiles of survivorship function”). Могут быть также построены графики различного типа, основным из которых является график зависимости доли (%) выживших от времени; этот график вызывается кнопкой “График времен жизни и кумулятивной доли выживших” (“Graph of survival times vs. cum. proportion surviving”). Этот убывающий график имеет вид ступенеобразной линии.

После активизации процедуры “Анализ выживаемости” (в методе Каплана—Мейера) или процедуры “Таблица времен жизни” (в методе “Таблицы и распределения времен жизни”) результаты будут выведены в отдельные окна (рис. 15.6 и 15.7).

**Представление результатов.** В качестве результатов должны быть представлены:

- 1) выживаемость (доля участников исследования, у которых изучаемый исход пока не наступил) для интересующего срока наблюдения:
  - см. столбец “Кум. доля выживш.” (“Cum. Prop. Surviving”) в методе таблиц дожития;
  - см. столбец “Кумул. времена” (“Cumulative Survival”) в методе Каплана—Мейера;
- 2) стандартная ошибка:
  - см. столбец “Стд. ош. кум. выж.” (“std. err. cum. surv.”) в методе таблиц дожития;
  - см. столбец “Стандартная ошибка” (“standard error”) в методе Каплана—Мейера;
- 3) число наблюдений для интересующего срока наблюдения;
- 4) медиана времени выживания (median survival time) — период времени, в течение которого исследуемый исход произойдет у 50% участников исследования, а также 25-й и 75-й процентили кривой дожития (в методе Каплана—Мейера);
- 5) желательно привести график — кривую дожития;
- 6) 95% ДИ для  $P$  (доля выживших)  $= P \pm 1,96 \cdot P \cdot \sqrt{\frac{1-P}{N}}$ .

#### 15.4.2. Исследование влияния одного фактора на время до наступления изучаемого исхода

**Задача:** исследовать, влияет ли какой-либо фактор на время до наступления изучаемого исхода.

*Пример:* исследовать, различаются ли группы больных по выживаемости, если в этих группах применялись разные методы лечения (например, хирургический или комбинированный).

**Решение:** такого типа задача сводится к сопоставлению групп больных по времени до наступления изучаемого исхода.

На предварительном этапе сравнения групп по выживаемости (или по времени до наступления другого исхода) должна быть продемонстрирована сопоставимость групп по основным демографическим и клиническим характеристикам, а также результатам лабораторного и инструментального исследования, позволяющим объективно оценить стадию заболевания, и т.п.

Две и более кривых дожития могут быть сопоставлены с помощью проверки нулевой статистической гипотезы об отсутствии различий кривых.

## STATISTICA:

→ Модуль “Анализ выживаемости” (Survival Analysis)

→ Процедуры:

- “Сравнение двух выборок” (“Comparing two samples”);
- “Сравнение нескольких выборок” (“Comparing multiple samples”)

В стартовом диалоговом окне для сравнения двух выборок (рис. 15.8) необходимо указать следующее:

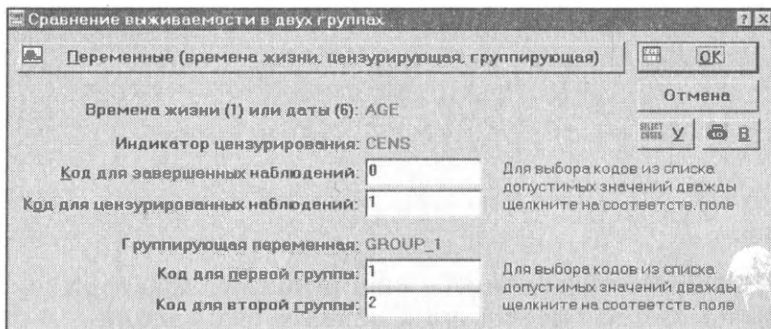


Рис. 15.8. Диалоговое окно сравнения выживаемости в двух группах.

- столбец, содержащий данные о времени до наступления изучаемого исхода или выбытия из исследования;
- столбец, содержащий индикатор цензурирования;
- столбец, содержащий группирующий признак (код группы).

В следующем диалоговом окне будет доступен вызов пяти тестов для сопоставления групп (рис. 15.9).

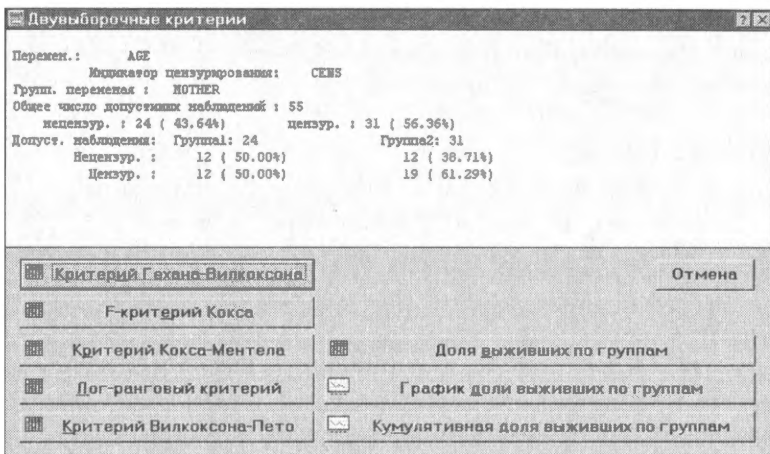


Рис. 15.9. Диалоговое окно для сопоставления времени до наступления изучаемого исхода в двух группах.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- При  $p > 0,05$  нулевая гипотеза о том, что кривые выживания идентичны, т.е. различий между группами по времени до наступления изучаемого исхода не существует, не отклоняется.
- При  $p < 0,05$  нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о различии между группами по времени до наступления изучаемого исхода.

**Представление результатов.** Привести следующую информацию:

- медианы времени до наступления изучаемого исхода в сопоставляемых группах;
- название использованного теста для сопоставления групп;
- точное значение  $p$ .

### 15.4.3. Анализ влияния нескольких факторов на время до наступления изучаемого исхода

**Задача:** определить, какие факторы влияют на время до наступления изучаемого исхода.

**Пример:** изучить, какие факторы, помимо метода лечения, влияют на выживаемость больных.

**Решение:** фактически это наиболее общий случай сравнения групп, рассмотренного в разделе 15.4.2. В данном случае исследуется не один фактор, а несколько.

#### STATISTICA:

→ Модуль “Анализ выживаемости” (“Survival Analysis”)

→ Подмодуль “Регрессионные модели”  
 (“Regression models”)

Регрессионный анализ заключается (как и в других случаях) в построении математической модели зависимости между признаками тем или иным методом. В ППП **STATISTICA** доступны четыре метода построения регрессионной модели выживаемости (выбор из списка моделей осуществляется в стартовом диалоговом окне “Регрессионные модели для цензурированных данных”):

- регрессионная модель Кокса;
- экспоненциальная регрессия;
- логнормальная регрессия;
- нормальная регрессия.

Зависимым (прогнозируемым) признаком в таких моделях является время до наступления изучаемого исхода, а независимыми — изучаемые факторы. После выполнения итерационной процедуры построения модели в диалоговое окно выводится основной результат — рассчитанное значение  $p$ .

Оценки параметров регрессионного уравнения и значения  $p$  для каждого из них могут быть вызваны обращением к функции “Оценки параметров” (“Parameter estimates”).

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $p < 0,05$  для какого-либо исследуемого признака, то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза о том, что этот признак ассоциирован с временем до наступления изучаемого исхода.

- Если  $p > 0,05$ , то нулевая гипотеза об отсутствии связи изучаемого признака с временем до наступления изучаемого исхода не отклоняется.

**Представление результатов.** Описание полученной модели удобно представлять в виде таблицы, содержащей:

- в строках — исследуемые факторы (независимые признаки);
- в столбцах:
  - коэффициенты  $beta$  (веса) факторов в регрессионном уравнении,
  - стандартные ошибки для коэффициентов  $beta$ ,
  - точные значения  $p$  для коэффициентов  $beta$ .

## 15.5. Анализ точности диагностического метода

В данном разделе рассматриваются способы оценки точности диагностического метода путем расчета его операционных характеристик (см. раздел 15.5.1) и способы анализа согласованности независимых диагностических заключений (см. раздел 15.5.2).

### 15.5.1. Операционные характеристики диагностического метода

**Задача:** оценить точность нового диагностического метода при диагностике определенного состояния.

**Пример:** оценить точность компьютерной томографии мозга в выявлении лакунарных инфарктов.

**Решение:** точность диагностического метода может быть оценена лишь в контролируемом исследовании следующей структуры:

- группа наблюдений (больных) подвергается двум диагностическим исследованиям — новому (изучаемому) и тому, который является на момент исследования “золотым стандартом” диагностики (его результаты считаются наиболее правильными, т.е. соответствующими действительности);
- два указанных диагностических метода используются независимо, т.е. результаты применения одного метода неизвестны в момент применения другого метода;
- каждый из методов дает положительное или отрицательное заключение (соответственно о наличии или отсутствии диагностируемого состояния).



**Замечание 1.** “Золотым стандартом” диагностики является морфологический метод. Его рекомендуется использовать в этом качестве всегда, когда это только возможно. Однако в большинстве случаев прижизненное морфологическое исследование сопряжено со значительными трудностями (например, необходимостью хирургического вмешательства).

**Замечание 2.** В качестве стандартного метода не может использоваться клинический диагноз!

Поскольку каждый из диагностических методов (в наиболее общем случае) может давать лишь один из двух результатов — положительный либо отрицательный, — то по окончании обследования больных может быть построена следующая таблица (рис. 15.10).

		"Золотой стандарт" (истина)	
		болен	здоров
Новый (исследуемый) метод	болен (положительный результат теста)	совпадение заключений о наличии болезни (истинно- положительный результат) <i>a</i>	гипердиагностика (ложно- положительный результат, ошибка первого рода, $\alpha$ -ошибка) <i>b</i>
	здоров (отрицательный результат теста)	гиподиагностика (ложно- отрицательный результат, ошибка второго рода, $\beta$ -ошибка) <i>c</i>	совпадение заключений об отсутствии болезни (истинно- отрицательный результат) <i>d</i>

Рис. 15.10. Таблица результатов исследования метода диагностики.

**Замечание 3.** Поскольку в исследовании такого типа анализу подвергаются два признака, каждый из которых имеет два возможных значения, такая задача является частным случаем задачи об ассоциации качественных признаков (см. главы 10—12).

Каждое из наблюдений может быть отнесено в одну из клеток этой четырехпольной таблицы. Таким образом, каждая из этих клеток будет содержать число наблюдений, в нее попавших —  $a, b, c, d$ . На основании этих чисел могут быть вычислены так называемые операционные характеристики диагностического теста:

- диагностическая чувствительность (ДЧ);
- диагностическая специфичность (ДС);
- диагностическая эффективность (ДЭ);
- прогностическая ценность положительного результата (ПЦПР);
- прогностическая ценность отрицательного результата (ПЦОР).

$ДЧ = a/(a+c)$ , в % — доля лиц с положительным результатом теста среди лиц с изучаемым заболеванием.

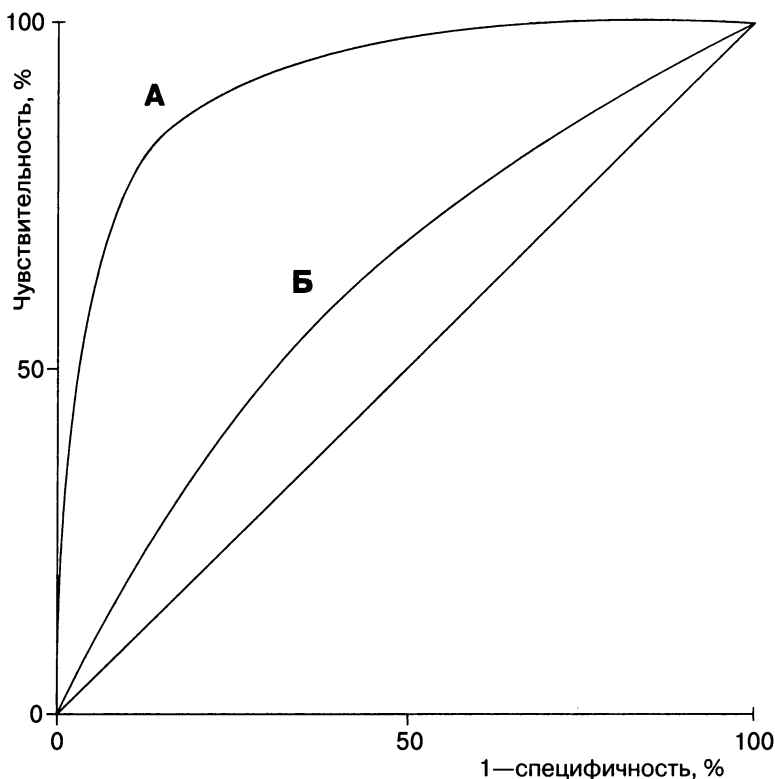
$ДС = d/(d+b)$ , в % — доля лиц с отрицательным результатом теста среди лиц без изучаемого заболевания.

$ДЭ = (ДЧ + ДС)/2$  — среднее между ДЧ и ДС.

В тех случаях, когда результат диагностического теста является количественным признаком (например, концентрация какого-либо вещества в лабораторном тесте), то целью исследования обычно является поиск оптимальной точки разделения между нормой и патологией. Естественно, что оптимальным является сочетание максимальных значений чувствительности и специфичности. Однако большему значению ДС соответствует меньшее значение ДЧ, и наоборот, так как эти величины антагонистичны. Поэтому в исследовании бывает необходимо построить несколько четырехпольных таблиц для разных точек разделения, по каждой из которых можно рассчитать значения ДС и ДЧ. После получения этих результатов обычно строится график в координатах  $(1-ДС)$  (в %) и ДЧ (в %) — характеристическая кривая<sup>1</sup> (рис. 15.11). Точки наносятся на плоскость и соединяются плавной линией. Точка, наиболее близкая к перегибу графика, в общем случае (если нет специальных соображений по поводу оптимизации либо ДЧ, либо ДС) может считаться соответствующей оптимальному соотношению ДЧ и ДС.

**Замечание.** Выбор точки разделения не является проблемой статистического анализа. Обычно этот выбор делается с учетом соотношения “цены” ошибок первого и второго рода.

<sup>1</sup> ROC-кривая (от англ. receiver-operator curve).



**Рис. 15.11.** Характеристические кривые для высокоинформативного (А) и низкоинформативного (Б) методов диагностики.

Характеристиками теста, которыми оперирует врач после получения результата теста у данного больного, являются следующие:

- ПЦПР =  $a / (a + b)$  — вероятность наличия заболевания при положительном результате теста;
- ПЦОР =  $c / (c + d)$  — вероятность отсутствия заболевания при отрицательном результате теста.

Значения ПЦПР и ПЦОР (при одних и тех же значениях ДЧ и ДС) зависят от распространенности заболевания в исследуемой популяции  $P = (a + c) / (a + b + c + d)$  по следующим формулам:

$$\text{ПЦПР} = (\text{ДЧ} \times P) / (\text{ДЧ} \times P + (1 - \text{ДС})(1 - P))$$

$$\text{ПЦОР} = (\text{ДС} \times (1 - P)) / ((1 - \text{ДЧ}) \times P + \text{ДС} \times (1 - P))$$

При разработке скрининговых диагностических тестов обычно учитывают низкую распространенность заболевания в популяции, поэтому скрининговые тесты должны иметь высокую ДС.

Собственно диагностические клинические исследования проходят в условиях высокой распространенности заболевания в обследуемой группе, поэтому для таких тестов особенно важна высокая ДЧ.

Чем больше ДЧ, тем больше ПЦОР (т.е. отрицательные результаты отвергают наличие заболевания).

Чем больше ДС, тем больше ПЦПР (т.е. положительные результаты подтверждают наличие заболевания).

Чем ниже  $P$  (даже при высокой ДС), тем больше ложноположительных результатов.

Чем выше  $P$  (даже при высокой ДЧ), тем больше ложноотрицательных результатов.

Чем ниже  $P$  (стремится к нулю), тем ниже ПЦПР (стремится к нулю).

Чем выше  $P$  (стремится к 100%), тем ниже ПЦОР (стремится к нулю).

Для оценки ПЦПР более существенна  $P$  (колеблющаяся от долей процента до десятков процентов, т.е. в 100 раз и более), чем ДЧ и ДС (колеблющиеся обычно в интервале 60—99%).

*Пример:* положительные результаты ревматической пробы при скрининговом обследовании в популяции с распространенностью ревматизма 10% подтверждаются в 15% случаев, т.е. ПЦПР=15%. В группе больных с наличием симптоматики ревматизма с  $P=26%$  тест подтверждается в 40% случаев, т.е. ПЦПР=40%.

Таким образом, исходя из изложенного ясно, что для анализа точности диагностического теста не требуется применение статистических методов. Однако мы считали необходимым изложить данный материал для обеспечения общности при освещении проблем анализа результатов исследований. Более сложной является задача, когда результат теста включает помимо положительного и отрицательного ответа также третий (например, неопределенный). В этих случаях мы рекомендуем обращаться к статистике.

### **15.5.2. Анализ согласованности независимых диагностических заключений**

Часто возникает необходимость оценить, насколько совпадают заключения, которые дают разные врачи или которые получе-

ны с помощью разных диагностических методов. Для этого могут применяться коэффициент каппа, внутриклассовый коэффициент корреляции и другие статистические оценки. Выбор метода оценки согласия должен проводиться с учетом типа признака, являющегося результатом диагностического метода — бинарного, номинального, порядкового, количественного. Наиболее простым является случай бинарных признаков, решение для которого мы приводим ниже. В других случаях следует использовать более сложные методы, для чего необходимо обращаться к статистике.

**Замечание.** Для оценки согласия количественных результатов тестов часто применяются методы корреляционного анализа (Пирсона, Спирмена и др.), что не вполне корректно, так как результаты оценки согласия могут оказаться завышенными в связи с возникновением аддитивных и мультипликативных ошибок [37].

**Задача:** оценить согласованность результатов, полученных с помощью двух независимых методов диагностики.

**Пример:** оценка совпадения диагностических заключений “болезнь есть — болезни нет” у двух врачей, работающих независимо друг от друга.

**Решение:** к сожалению, в ППП *STATISTICA* нет алгоритмов для решения такой задачи. Предлагаем в таком случае пользоваться следующим достаточно простым способом. Вероятность случайного совпадения (согласия) можно рассчитать по формуле:

$P_c = P_1 \times P_2 + (1 - P_1)(1 - P_2)$ , где  $P_1$  — вероятность положительных заключений у первого врача,  $P_2$  — то же у второго врача. При наличии чистой случайности в вынесении решений искомая вероятность равна 50%. При неслучайном вынесении решений искомая вероятность должна быть больше (на практике бывает больше лишь на 10—20%).

Мерой превышения согласия над случайным согласием может служить коэффициент каппа:

$K = (P_o - P_c) / (1 - P_c)$ , где  $P_o$  — вероятность согласия.

Стандартная ошибка для  $K$  может быть рассчитана по формуле:

$$m_K = \sqrt{\frac{P_o(1 - P_o)}{n(1 - P_c)^2}},$$

а границы ДИ соответственно равны  $K \pm t \times m_K$ .

Значение  $t$  принимается обычно приближенно равным 1,96 для 95% ДИ. Точное значение  $t$  для конкретного случая можно узнать, воспользовавшись опцией “Вероятностного калькулятора”.

### STATISTICA:

- Модуль “Основные статистики и таблицы” (“Basic statistics/Tables”)
- Подмодуль “Вероятностный калькулятор” (“Probability calculator”)

В диалоговом окне (см. рис. 7.9) необходимо выполнить следующее:

- выбрать вид распределения “ $t$  Стьюдента” (“Student  $t$  test”);
- выбрать опции “Обратная ф.р.”, “Двусторонняя” (“Two-sided”) и “1–ф.р.”;
- задать число степеней свободы “Ст. св.” (“ $df$ ”),  $df = n - 1$ ;
- задать значение  $p$  (например, 0,05 для вычисления границ 95% ДИ);
- нажать кнопку “Вычислить” (“compute”).

Искомое значение  $t$  появится в окне “ $t$ ”.

**Интерпретация результатов.** Полученные результаты можно интерпретировать следующим образом.

- Если  $K > 0$ , то заключения врачей (или методов диагностики) совпадают чаще, чем случайно.
- Если  $K = 0$ , то совпадение заключений случайно.
- Если  $K < 0$ , то заключения совпадают реже, чем случайно.

## **Глава 16. Объединение результатов нескольких исследований (мета-анализ)**

В данной главе мы не останавливаемся на технике выполнения мета-анализа, так как она требует определенной подготовки в области прикладной статистики. Однако нам хотелось бы дать читателю общие сведения о данном типе исследований, который в значительной степени основан на использовании статистических процедур.

В соответствии с концепцией доказательной медицины научно обоснованными признаются результаты лишь тех клинических исследований, которые проведены на основе принципов клинической эпидемиологии, позволяющих свести к минимуму как систематические ошибки (путем проведения исследования, структура которого адекватна поставленным задачам), так и случайные ошибки (с помощью корректного статистического анализа полученных в исследовании данных). Наиболее обоснованные результаты обычно могут быть получены при проведении рандомизированных контролируемых испытаний лечебных и профилактических вмешательств, так как в таких случаях организация (т.е. структура) и проведение исследования наиболее близки к эксперименту в общенаучном понимании этого термина.

Достаточно часто результаты исследований, в которых оценивается эффективность одного и того же лечебного или профилактического вмешательства или диагностического метода при одном и том же заболевании, различаются. В связи с этим возникает необходимость относительной оценки результатов разных исследований и интеграции их результатов с целью получения обобщающего вывода. Эту задачу начало решать возникшее в 1993 г. Кокрановское сотрудничество, названное в честь английского эпидемиолога А. Кокрана, который впервые призвал своих коллег оценивать эффективность лечебных вмешательств путем анализа результатов всех когда-либо проведенных клинических испыта-

ний. Кокрановское сотрудничество объединяет исследователей из десятков стран (его Российское отделение имеет сайт в Интернете: [www.cochrane.ru](http://www.cochrane.ru)). Цель Кокрановского сотрудничества — искать и обобщать путем подготовки систематических обзоров результаты всех когда-либо проводившихся клинических испытаний лечебных и профилактических вмешательств. Кокрановское сотрудничество ведет регистр рандомизированных клинических испытаний в форме базы данных.

**Систематический обзор** представляет собой разновидность научного исследования, в котором объектом изучения являются результаты ряда оригинальных исследований по определенной проблеме. В отличие от традиционных обзоров систематические обзоры не тенденциозны и выполняются в соответствии с четким алгоритмом, обеспечивающим максимально возможную надежность выводов обзора.

Систематические обзоры классифицируются как **качественные** и **количественные**. В последних используются статистические методы обобщения результатов (**мета-анализ**) различных исследований с целью получения единого вывода об эффективности лечебного или профилактического вмешательства либо об информативности диагностического метода. При этом существенно увеличивается выборка, что повышает статистическую мощность анализа и точность оценки эффекта лечения. Часто систематический обзор, выполненный с использованием статистических методов, также называют мета-анализом. *Цель мета-анализа* — выявление, изучение и объяснение различий (вследствие наличия статистической неоднородности, или гетерогенности) в результатах исследований, а также более точная оценка изучаемого эффекта [38].

Существуют два основных подхода к выполнению мета-анализа.

*Первый* из них заключается в статистическом повторном анализе отдельных исследований путем сбора первичных данных о включенных в оригинальные исследования наблюдениях. Очевидно, что далеко не всегда это выполнимо, но необходимо упомянуть о том, что в настоящее время за рубежом уже начали создаваться банки первичных данных по определенным медицинским научным проблемам.

*Второй* (и основной) подход заключается в обобщении опубликованных результатов исследований, посвященных одной про-



блеме. Такой мета-анализ выполняется обычно в несколько этапов [39], среди которых важнейшими являются:

- выработка критериев включения оригинальных исследований в мета-анализ;
- оценка гетерогенности (статистической неоднородности) результатов оригинальных исследований;
- проведение собственно мета-анализа (получение обобщенной оценки величины эффекта);
- анализ чувствительности выводов.

Необходимо отметить, что этап определения круга включаемых в мета-анализ исследований (как в аспекте полноты выявления выполненных работ, так и в отношении формальных критериев оценки их методологического качества) часто становится источником систематических ошибок мета-анализа. Качество мета-анализа существенно зависит от качества включенных в него исходных исследований и статей.

К основным проблемам при включении исследований в мета-анализ относятся такие, как различия исследований по критериям включения и исключения, структуре исследования, контролю качества.

Существует также смещение, связанное с преимущественным опубликованием положительных результатов исследования (исследования, в которых получены статистически значимые результаты, чаще публикуются, чем те, в которых такие результаты не получены). Поскольку мета-анализ основан главным образом на опубликованных данных, следует обращать особое внимание на недостаточную репрезентативность отрицательных результатов в литературе. Включение в мета-анализ неопубликованных результатов также представляет значительную проблему, так как их качество неизвестно в связи с тем, что они не проходили рецензирование.

Собственно мета-анализ может осуществляться с использованием ППП ReviewManager (RevMan, Cochrane Collaboration)<sup>1</sup>, содержащим основные необходимые программные средства для оформления обзора и статистические процедуры для выполнения собственно мета-анализа. Выбор метода определяется типом анализируемых данных (бинарные или непрерывные) и типом мо-

---

<sup>1</sup> Пакет прикладных программ RevMan доступен бесплатно по адресу <http://www.cochrane.de/cochrane/revman.htm>

дела (фиксированных эффектов, случайных эффектов). Остановимся на этом подробнее.

**Бинарные данные** обычно анализируются путем вычисления отношения шансов (ОШ), относительного риска (ОР) или разности рисков в сопоставляемых выборках. Все перечисленные показатели характеризуют эффект вмешательств. Представление бинарных данных в виде ОШ удобно использовать при статистическом анализе, но этот показатель достаточно трудно интерпретировать клинически.

**Неперывными данными** обычно являются диапазоны значений изучаемых признаков или нестандартизованная разница взвешенных средних в группах сравнения, если исходы оценивались во всех исследованиях одинаковым образом. Если же исходы оценивались по-разному (например, по разным шкалам), то используется стандартизованная разница средних (так называемая величина эффекта) в сравниваемых группах.

Одним из первых этапов мета-анализа является оценка **гетерогенности (статистической неоднородности)** результатов эффекта вмешательства в разных исследованиях. Для оценки гетерогенности часто используют критерий  $\chi^2$  с нулевой гипотезой о равном эффекте во всех исследованиях и с уровнем значимости 0,1 для повышения статистической мощности (чувствительности) теста.

Источниками гетерогенности результатов разных исследований принято считать дисперсию внутри исследований (обусловленную случайными отклонениями результатов разных исследований от единого истинного фиксированного значения эффекта), а также дисперсию между исследованиями (обусловленную различиями между изучаемыми выборками по характеристикам больных, заболеваний, вмешательств, приводящими к несколько разным значениям эффекта — случайными эффектами).

Если предполагается, что дисперсия между исследованиями близка к нулю, то каждому из исследований приписывается вес, величина которого обратно пропорциональна дисперсии результата данного исследования. Дисперсия внутри исследований в свою очередь определяется объемом выборки и рядом других характеристик исследования.

При нулевой дисперсии между исследованиями можно использовать модель фиксированных (постоянных) эффектов. В этом случае предполагается, что изучаемое вмешательство во всех ис-

следованиях имеет одну и ту же эффективность, а выявляемые различия между исследованиями обусловлены только дисперсией внутри исследований. В этой модели пользуются методами Mantel—Haenszel и Peto [40, 41].

Модель случайных эффектов предполагает, что эффективность изучаемого вмешательства в разных исследованиях может быть разной. Данная модель учитывает дисперсию не только внутри одного исследования, но и между разными исследованиями. В этом случае суммируются дисперсии внутри исследований и дисперсия между исследованиями. В модели случайных эффектов чаще всего применяют метод DerSimonian и Laird [42], позволяющий оценить дисперсию между исследованиями.

Целью мета-анализа непрерывных данных обычно является представление точечных и интервальных (95% ДИ) оценок обобщенного эффекта вмешательства.

Существует также ряд других подходов к выполнению мета-анализа: байесовский мета-анализ, кумулятивный мета-анализ, многофакторный мета-анализ, мета-анализ выживаемости (перечисленные методы пока не представлены в ППП RevMan версии 3.0).

**Байесовский мета-анализ** позволяет рассчитать априорные вероятности эффективности вмешательства с учетом косвенных данных. Такой подход особенно эффективен при малом числе анализируемых исследований. Он обеспечивает более точную оценку эффективности вмешательства в модели случайных эффектов за счет объяснения дисперсии между разными исследованиями.

**Кумулятивный мета-анализ** — частный случай байесовского мета-анализа — пошаговая процедура включения результатов исследований в мета-анализ по одному в соответствии с каким-либо принципом (в хронологической последовательности, по мере убывания методологического качества исследования и т.д.). Он позволяет рассчитывать претестовые (априорные) и послетестовые (апостериорные) вероятности в итерационном режиме по мере включения исследований в анализ.

**Регрессионный мета-анализ** (логистическая регрессия, регрессия взвешенных наименьших квадратов, модель Кокса и др.) используется при существенной гетерогенности результатов исследований. Он позволяет учесть влияние нескольких характеристик исследования (например, размера выборки, дозы препарата, способа его введения, характеристик больных и др.) на ре-

зультаты испытаний вмешательства. Результаты регрессионного мета-анализа обычно представляют в виде коэффициента наклона с указанием ДИ.

Следует заметить, что мета-анализ может выполняться для обобщения результатов не только контролируемых испытаний медицинских вмешательств, но и когортных исследований (например, исследований факторов риска). Однако при этом следует учитывать высокую вероятность возникновения систематических ошибок.

Особый вид мета-анализа — обобщение оценок информативности диагностических методов, полученных в разных исследованиях. Цель такого мета-анализа — построение характеристической кривой взаимной зависимости чувствительности и специфичности тестов (ROC-кривой) с использованием взвешенной линейной регрессии.

После получения обобщенной оценки величины эффекта возникает необходимость определить ее *устойчивость*. Для этого выполняется так называемый анализ чувствительности. Одним из способов его проведения является сопоставление результатов, получаемых в двух моделях — фиксированных и случайных эффектов. Во второй модели результаты обычно бывают статистически менее значимыми. Другой способ анализа чувствительности — исключение того или иного исследования из анализа и пересчет результатов с последующей оценкой гетерогенности результатов по критерию  $\chi^2$ .

Существуют также способы *оценки полноты выявления включенных в мета-анализ исследований*. Обычно неполнота выявления связана с возникновением систематической ошибки, связанной с преимущественным опубликованием положительных результатов исследований (результатов, описывающих статистически значимые различия групп). Для качественной оценки наличия такой систематической ошибки мета-анализа обычно прибегают к построению воронкообразной диаграммы рассеяния результатов отдельных исследований в координатах (величина эффекта, размер выборки). При полном выявлении исследований эта диаграмма должна быть симметричной. Вместе с тем существуют и формальные методы оценки существующей асимметрии.

Результаты мета-анализа обычно представляются графически (точечные и интервальные оценки величин эффектов каждого из включенных в мета-анализ исследований; пример на рис. 16.1) и

в виде таблиц с соответствующими статистиками (средние, СКО, ДИ).

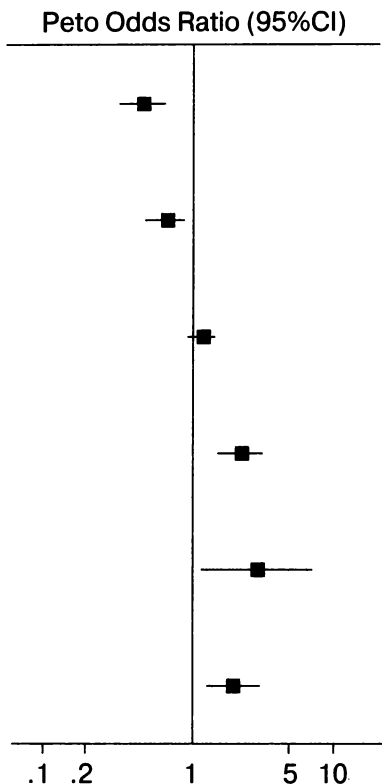


Рис. 16.1. Пример представления результатов мета-анализа [43].

В заключение необходимо подчеркнуть, что мета-анализ является достаточно новой областью применения статистических методов в медицинских исследованиях, при выполнении и интерпретации результатов которого существует много нерешенных проблем. Необходимость исследований в этом направлении определяется бурным развитием концепции и практики доказательной медицины в современном цивилизованном обществе.

---

## **Глава 17. Современные требования к описанию в публикациях процедуры и результатов статистического анализа медико-биологических данных**

В настоящей главе кратко изложены основные современные требования к описанию процедуры и результатов статистического анализа данных, а также некоторые общие требования к рукописям, представляемым в международные журналы. Часть этих положений изложена в текущей редакции “The Uniform requirements of manuscripts submitted to biomedical journals”, разработанных Ванкуверской группой редакторов медицинских журналов. Русскоязычная версия этого документа была опубликована несколько лет назад ([44], <http://www.mediasphera.ru>). Основным требованием является следующее: “Опишите статистические процедуры достаточно детально, чтобы квалифицированный читатель мог при наличии доступа к данным проверить представленные результаты”. Другие положения являются менее формальными, однако необходимыми для того, чтобы статья могла быть представлена в редакции международных журналов, в большинстве которых процедура рецензирования статей включает также анализ качества выполнения и описания статистического анализа данных. Большинство этих положений подробно освещены в книге [35].

Некоторые из нижеследующих положений упоминались и подробно освещались нами в предыдущих главах. В настоящей главе мы приводим их в сжатой форме, причем дальнейшее изложение соответствует общепринятому порядку разделов научных статей (введение, материалы и методы, результаты, обсуждение, выводы) по медико-биологическим проблемам.

## **17.1. Рекомендации по разделу “Введение”**

1. Опишите научные гипотезы, проверка которых являлась целью исследования.
2. Укажите, каким образом оригинальные данные могут быть получены для повторного анализа.

## **17.2. Рекомендации по разделу “Материалы и методы”**

Перечислим кратко основные рекомендации по описанию структуры исследования, касающиеся получения данных:

1. Укажите, что являлось объектом изучения (человек, нога, глаз и т.п.).
2. Укажите даты начала и окончания сбора данных и причины выбора таких дат.
3. Укажите название и тип организации (например, тип стационара или диспансера), в которой выполнялось исследование.
4. Опишите популяцию, из которой сделана выборка (для оценки ее репрезентативности).
5. Укажите число наблюдений на каждом этапе исследования и для каждого признака.
6. Опишите наблюдения, которые включены в исследование (в том числе отдельно выбывших из исследования — в случае динамического исследования).
7. Опишите сходство и различия между исследуемыми (сравниваемыми) группами в начале исследования.
8. Для признаков, основанных на оценках (мнениях), приведите результаты статистического анализа согласия между экспертами.
9. Дайте определения для всех исследуемых признаков, особенно клинических (укажите используемые шкалы, классификации и т.д.).

Для описания клинических испытаний дополнительно выполните следующее:

1. Укажите, соответствует ли структура исследования современным этическим нормам, в том числе этическим требованиям

- работы с животными. Отметьте факт одобрения протокола исследования этическим комитетом.
2. Подробно опишите способы формирования групп и определения необходимого числа наблюдений.
  3. Опишите процедуру формирования выборки, в том числе получение информированного согласия участника исследования (если это требовалось).
  4. Приведите критерии включения и исключения участников исследования.
  5. Опишите методику распределения наблюдений по группам (например, по группам лечения и контроля). Если использовалась рандомизация, опишите ее метод.
  6. Опишите технику маскирования (ослепления), если она использовалась.
  7. Опишите исследуемые медицинские вмешательства, в том числе плацебо, если они применялись.
  8. Опишите методы сбора данных.
  9. Опишите методы контроля качества, использованные для обеспечения полноты и точности сбора данных.
  10. Опишите все отклонения от первоначального протокола и их причины.
  11. Опишите все побочные явления.
  12. Опишите методы статистического анализа данных.

Теперь остановимся несколько подробнее на описании статистических методов анализа данных.

В разделе “Материалы и методы” следует описать задачи, которые ставились при анализе данных, например:

- оценка параметров (например, средних, средних квадратических отклонений, процентильных интервалов) изучаемых признаков в группах;
- сравнение групп;
- исследование связей двух или нескольких признаков;
- анализ выживания;
- другие задачи.

В данном разделе статьи или диссертации не следует упоминать названия конкретных статистических методов, которые были использованы, так как выбор того или иного метода зависит от результатов предыдущих этапов анализа данных. Если же для решения поставленной задачи должен быть использован единствен-



но возможный статистический метод, то целесообразно его упомянуть в данном разделе публикации. В любом случае названия статистических методов должны упоминаться по ходу изложения результатов (в разделе “Результаты”).

При анализе данных контролируемых испытаний медицинских вмешательств следует указать, проводился ли статистический анализ данных, полученных только у тех больных, которые закончили лечение, или анализировались данные, полученные у всех первоначально включенных в исследование участников. Если выбывшие из исследования составили более 20%, то следует специально указать, не утрачена ли сопоставимость групп контроля и лечения в результате выбывания больных. Рекомендации группы CONSORT (Consolidated Standards of Reporting Trials) [10, 12] по представлению результатов рандомизированного контролируемого исследования включают, в частности, следующее описание методов статистического анализа в разделе изложения протокола исследования: “опишите методы статистического анализа с детализацией основного сравнительного анализа. Укажите, проводился ли анализ исходя из фактически полученного лечения или на основе предположения, что все пациенты получили лечение, предписанное при рандомизации”. В указанных рекомендациях достаточно полно изложены требования в отношении описания статистического анализа.

Необходимо отметить, приводятся ли в тексте статьи или диссертации все сопоставления, которые выполнены в ходе анализа данных. Это необходимо знать читателю для оценки существенности проблемы множественных сравнений.

Желательно в этом разделе публикации указать, какое значение было принято в качестве порогового уровня статистической значимости. В связи с существованием проблемы множественных сравнений зачастую желательно устанавливать порог статистической значимости (вероятность совершения ошибки первого рода — вероятность гипердиагностики) на уровне хотя бы 0,01 (чем меньше уровень значимости, тем доказательнее результаты).

Необходимо указать название использованного при анализе данных пакета программ и фирму-производителя. Использование собственных разработок (аналогичных общепризнанным статистическим пакетам) не рекомендуется. Это связано с тем, что использование разных алгоритмов при программировании одних

и тех же статистических методов может приводить к несколько разным результатам, правильность алгоритмов же коммерческих пакетов программ проверена, и их высокое качество общеизвестно. Среди коммерчески доступных в России пакетов можно назвать, например, следующие:

- STATISTICA (StatSoft, USA); существует также русифицированная версия производства StatSoft-Russia, 1999;
- BIOSTATISTICA (S.A. Glantz, McGraw Hill, перевод на русский язык — “Практика”, 1998);
- SPSS (SPSS Inc., USA);
- SAS (SAS Inst., USA);
- BMDP (BMDP, USA);
- STATGRAPHICS (Manugistic Inc., USA);
- EpiInfo 2000 (Centers for Disease Control and Prevention, USA).

Если авторы пользуются оригинальными, не имеющими аналогов, математическими методами анализа данных, то желательны ссылки на предыдущие публикации в рецензируемых международных изданиях математического профиля, наличие которых может служить косвенным подтверждением надежности используемых методов.

### **17.3. Рекомендации по разделу “Результаты”**

1. Укажите, какие гипотезы проверяются.
2. Заранее укажите минимальные различия между группами по анализируемым параметрам (например, средним значениям), которые Вы будете считать клинически значимыми. Например, выявленное в исследовании статистически значимое снижение АД на 5 мм рт. ст. в результате применения нового гипотензивного средства вряд ли следует рассматривать как клинически значимое, так как эта цифра примерно соответствует погрешности метода измерения АД. Следовательно, необходимо заранее указать, какие именно различия между контрольной и экспериментальной группами будут рассматриваться как клинически значимые, т.е., например, меняющими тактику диагностики или лечения.
3. Укажите критический (пороговый) уровень значимости ( $\alpha$ -вероятность, вероятность ошибки первого рода), ниже которого Вы будете считать результаты статистически значимыми.

4. Опишите расчет мощности статистического анализа, примененного для сравнения групп (см. раздел 15.3).
5. Укажите, какие статистические тесты (критерии) использовались для каждого сравнения. Обоснуйте выбор данных статистических тестов (критериев) для проверки каждой из статистических гипотез.
6. Укажите, одно- или двусторонний тест использовался. Обоснуйте использование односторонних тестов, если они использовались.
7. Укажите, использовался ли тест для связанных или несвязанных групп.
8. В том случае, если использовались статистические процедуры внутреннего языка пакета (например, SAS, SPSS и т.д.), укажите эти процедуры и перечислите основные использованные опции этих процедур.
9. Сначала опишите результаты первичного анализа.
10. Опишите все крайние значения (аномальные значения, выбросы) отдельных признаков и способ обработки этих значений при статистическом анализе.
11. Подтвердите, что условия применимости использованного статистического теста соблюдаются.
12. Опишите абсолютные и относительные изменения или различия для всех исходов (конечных точек исследования).
13. Приведите ДИ для изменений или различий между исходами.
14. Приведите точное значение  $p$  независимо от его величины (значимо или незначимо)<sup>1</sup>.
15. Для клинически значимых различий, незначимых статистически, приведите наблюдаемую разницу, 95% ДИ, точное значение  $p$ .
16. Укажите, учитывалась ли проблема множественных сравнений, и каким образом она преодолевалась.

---

<sup>1</sup> Традиционное представление значения  $p$  только в виде  $p < 0,05$  просто показывает, что должна быть принята альтернативная гипотеза. Но значения  $p$ , равные 0,051 и 0,049, должны интерпретироваться сходно, хотя  $0,051 > 0,05$  и, следовательно, нулевая гипотеза не отклоняется (различия незначимы), а при  $p = 0,049$  должна быть принята альтернативная гипотеза (различия значимы). Приведение точного  $p$  позволяет избежать проблемы интерпретации.

Значения  $p$  должны быть приведены с точностью не более 2 значащих десятичных разрядов, а  $p = 0,001$  является наименьшим числом, которое нужно приводить. Значения  $p$ , меньшие этого, должны быть приведены в виде " $p < 0,001$ ".

17. Результаты сравнения выборок по признакам, включенным в исследование в качестве основных, должны быть приведены независимо от того, оказались ли различия между выборками статистически значимыми.
18. Опишите весь вторичный анализ, который дал интересные результаты.
19. Если возможно, приведите основные находки исследования на рисунках или в таблицах. Они не только помогают сэкономить место, но и часто представляют результаты более ясно, чем текст. Избегайте дублирования информации на рисунках и в таблицах. Вместо этого в тексте привлеките внимание к той части рисунка или таблицы, которая демонстрирует интересные факты.
20. По возможности приведите статистические находки достаточно детально, чтобы сделать возможным последующий метаанализ.
21. Приведите анализ взаимодействия изучаемых признаков, если таковое могло иметь место.
22. Учитывайте все наблюдения и объясните все пропуски в данных.

Рекомендации группы CONSORT (Consolidated Standards of Reporting Trials) [10, 12] по представлению результатов рандомизированного контролируемого исследования включают следующие требования к описанию анализа данных:

- приведите рассчитанное влияние вмешательства на основные и дополнительные критерии оценки (исходы) в виде точечных оценок и ДИ;
- приведите результаты в виде абсолютных, а не относительных показателей (например, 10/20, а не 50%);
- представьте основные данные, обработанные общепринятыми статистическими методами<sup>1</sup> достаточно детально, чтобы

---

<sup>1</sup> Используемый в данном документе термин “общепринятые статистические методы” представляется нам весьма неопределенным. Предполагается, видимо, использование традиционно и широко используемых методов статистического анализа, прежде всего одномерного — критерии Манна—Уитни, Вилкоксона, Стьюдента,  $\chi^2$  и др., а также использование для анализа коммерческих пакетов прикладных программ по статистике.

читатели могли их воспроизвести или провести альтернативный анализ;

- опишите прогностические параметры по группам вмешательства и любые попытки подгонки групп по этим параметрам;
- приведите все другие клинические исходы, ожидавшиеся или нет.

## 17.4. Рекомендации по разделу “Обсуждение”

*Результаты первичного анализа данных* должны обсуждаться в первую очередь. При этом необходимо обсуждать как положительные (статистически значимые), так и отрицательные (статистически незначимые) результаты анализа.

Необходимо обсудить *клиническую значимость* полученных результатов. При этом нельзя ограничиваться обсуждением только статистически значимых результатов. Поскольку при постановке задач могли быть выдвинуты гипотезы, не нашедшие подтверждения, их также необходимо обсудить. Даже небольшие (статистически незначимые) различия между группами могут иметь существенное клиническое значение, если они выявлены для важных признаков и на больших выборках. Одновременно многие статистически значимые различия могут быть абсолютно клинически незначимыми.

*Результаты вторичного анализа данных* должны обсуждаться с большой осторожностью, так как они являются “побочным продуктом” исследования и зачастую не являются доказательными в силу того, что их изучение не было предусмотрено структурой исследования.

*Промежуточные исходы* клинических исследований (косвенные критерии оценки) должны интерпретироваться аккуратно, так как они не являются подтверждением причинно-следственных связей клинических исходов с изучаемыми признаками.

Авторам также рекомендуется привести обсуждение собственных результатов *в контексте других публикаций* по данной проблеме.

Желательно также показать *практическую значимость* результатов (если таковая вообще предполагается). Здесь следует помнить об обобщаемости результатов исследования, кото-

рая определяется жесткостью критериев включения и исключения наблюдений в исследование.

Рекомендуется осветить необходимые (с точки зрения авторов) *направления дальнейших исследований*.

## **17.5. Описание ограничений исследования**

Включение в публикацию данного раздела является проявлением научной честности авторов. Данный раздел необходим для того, чтобы читатели смогли:

- оценить обобщаемость изложенных результатов;
- избежать подобных проблем в своих будущих исследованиях.

## **17.6. Рекомендации по разделу “Выводы”**

Выводы должны быть обоснованными, т.е. включать лишь положения, подтвержденные в ходе исследования, преимущественно в ходе первичного статистического анализа. Выводы по результатам клинических исследований должны включать преимущественно клинически значимые результаты.

В выводы следует включать как положительные, так и отрицательные результаты исследования (т.е. положения, как подтвержденные, так и опровергнутые в ходе исследования), поскольку они имеют одинаковое значение для науки. В настоящее время специальный анализ публикаций выявляет существенное смещение в сторону преимущественного опубликования положительных результатов исследований.

Результаты вторичного статистического анализа должны представляться в выводах с осторожностью.

## ЛИТЕРАТУРА

1. *Бащинский С.Е.* Evidence-Based Medicine и Международный журнал медицинской практики. Междунар журнал мед практики 1996;1:6—11.
2. *Sadowski Z.P., Alexander J.H., Skrabucha B. et al.* Multicenter randomized trial and a systematic overview of lidocaine in acute myocardial infarction. Am Heart J 1999 May; 137(5):792—8.
3. *Флетчер Р., Флетчер С., Вагнер Э.* Клиническая эпидемиология. Основы доказательной медицины. Пер. с англ. М: МедиаСфера 1998.
4. *Власов В.В.* Введение в доказательную медицину. М: МедиаСфера 2001.
5. *Адлер Ю.П., Маркова Е.В.* Планирование эксперимента при поиске оптимальных условий. М: Наука 1976.
6. *Лисенков А.Н.* Математические методы планирования многофакторных медико-биологических экспериментов. М: Медицина 1979.
7. *Максимов В.Н.* Многофакторный эксперимент в биологии. М: МГУ 1980.
8. *Планирование эксперимента в биологии и сельском хозяйстве.* Под ред. В.Н. Максимова. М: МГУ 1991.
9. *Pearl R.* Cancer and tuberculosis. Am J Hyg (now Am J Epidemiol) 1929;9:97—159.
10. *Бащинский С.Е.* Как следует представлять данные рандомизированных контролируемых исследований. Междунар журн мед практики 1997;1:7—10.
11. Новая форма представления результатов исследований, посвященных лечению. Междунар журн мед практики 1998;1:7—8.
12. *Альтман Д.Г., Шульц К.Ф., Мохер Д., Эггер М., Давидовф Ф., Элбурн Д., Гёцше П.К., Ланг Т.* Пересмотренный вариант единых стандартов представления результатов рандомизированных контролируемых испытаний (CONSORT): разъяснения и перспективы дальнейшего усовершенствования. Междунар журн мед практики 2001;5: 5—34.
13. *Боровиков В.П., Боровиков И.П.* STATISTICA. Статистический анализ и обработка данных в среде Windows. М: Финин 1997.
14. *Боровиков В.П., Ивченко Г.И.* Прогнозирование в системе STATISTICA® в среде Windows. М: Финансы и статистика 2000.
15. *Боровиков В.П.* Программа STATISTICA для студентов и инженеров (2-е издание книги «Популярное введение в программу STATISTICA»). М: Компьютер Пресс 2001.
16. *Боровиков В.П.* STATISTICA: искусство анализа данных на компьютере. СПб: Питер 2001.
17. *Поляков И.В., Соколова Н.С.* Практическое пособие по медицинской статистике. Л: Медицина 1975.
18. *Леонов В.П., Ижевский П.В.* Об использовании прикладной статистики при подготовке диссертационных работ по медицинским и биологическим специальностям Бюлл. ВАК.1997;5:56—61.
19. *Леонов В.П., Ижевский П.В.* Применение статистики в медицине и биологии: анализ публикаций 1990—1997 гг. Сибирский мед журнал 1997;3—4: 64—74.
20. *Айвзян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: Исследования зависимостей: Справ. изд. Под ред. С.А. Айвзяна. М: Финансы и статистика 1985.

21. Флейс Дж. Статистические методы для изучения таблиц долей и пропорций. Пер. с англ. Под ред. Ю.Н. Длаговещенского. М: Финансы и статистика 1989.
22. Закс Л. Статистическое оценивание. М: Статистика 1976.
23. Холлендер М., Вулф Д.А. Непараметрические методы статистики. Пер. с англ. М: Финансы и статистика 1983.
24. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. М: Мир 1982.
25. Гланц С. Медико-биологическая статистика. Пер. с англ. М: Практика 1998.
26. Altman D.G. Practical Statistics for Medical Research. London: Chapman & Hall 1991.
27. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. Пер. с англ. М: Прогресс 1976.
28. Шеффе Х. Дисперсионный анализ. Пер. с англ. М: Наука 1980.
29. Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. и др. Факторный, дискриминантный и кластерный анализ. Пер. с англ. Под ред. И.С. Енюкова. М: Финансы и статистика 1989.
30. Леонов В.П. Обработка экспериментальных данных на программируемых микрокалькуляторах. Под ред. Б.А. Гладких. Томск: Изд-во ТГУ 1990.
31. Харман Г. Современный факторный анализ. М: Статистика 1972.
32. Иберла К. Факторный анализ. М: Статистика 1980.
33. Джонстон Дж. Эконометрические методы. М: Статистика 1980
34. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Т.1—2. М: Финансы и статистика 1986—1987.
35. Методические рекомендации по разработке референтных величин лабораторных показателей. М 1983.
36. Lang T.A., Secic M. How to report statistics in medicine. Annotated guidelines for authors, editors, and reviewers. Copyright by American College of Physicians, Philadelphia, Pennsylvania, 1997, printing/binding by Port City Press, composition by Techsetters.
37. Власов В.В. Эффективность диагностических исследований. М: Медицина 1988.
38. Clarke M., Oxman A.D., editors. Cochrane Reviewers' Handbook 4.1.5 [updated April 2002]. In: The Cochrane Library, Issue 2, 2002. Oxford: Update Software. Updated quarterly.
39. Lau J., Ioannidis J.P.A., Schmid C.H. Quantitative synthesis in systematic reviews. Ann Intern Med 1997;127:820—6.
40. Mantel N., Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959; 22: 719—48.
41. Yusuf S., Peto R., Lewis J., Collins R., Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog Cardiovasc Dis 1985; 27: 335—71.
42. DerSimonian R., Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7: 177—88.
43. Cina C.S., Clase C.M., Haynes R.B. Carotid endarterectomy for symptomatic carotid stenosis (Cochrane Review). In: The Cochrane Library, Issue 2, 2002. Oxford: Update Software.
44. Единые требования к рукописям, представляемым в биомедицинские журналы. Международный комитет редакторов медицинских журналов. Международный журнал мед практики 1997;5:53—64.



## Приложение 1. Аннотированный список рекомендуемой литературы на русском языке\*

1. *Автандилов Г.Г.* Медицинская морфометрия. Руководство. М: Медицина 1990.

В первом руководстве по медицинской морфометрии изложены принципы создания представительных групп для морфометрических исследований всех уровней морфологической организации и основные методы математической обработки результатов. Описаны возможности медицинской морфометрии и ее главных разделов: антропометрии, органометрии, гистометрии, цитометрии, кариометрии, ультраструктурометрии и микроспектрофотометрии. Даны теоретические и практические сведения о стереологии. Описаны методы и приборы для объективной оценки морфологических изменений. Руководство предназначено для морфологов, патологоанатомов, судебных медиков, онкологов, патологов, биологов.

2. *Адлер Ю.П., Маркова Е.В.* Планирование эксперимента при поиске оптимальных условий. М: Наука 1976.

Книга представляет собой вводный курс по планированию эксперимента при поиске оптимальных условий. Рассматриваются полный и дробный факторный эксперимент, виды

---

\* Большинство аннотаций любезно предоставлено В.П. Леоновым. Более полный вариант данного аннотированного списка литературы можно найти на сайте <http://www.biometrica.tomsk.ru/lib/annot.html>, а перечень Интернет-ресурсов по проблемам биометрики и статистики — на <http://www.biometrica.tomsk.ru:8101/list/statleo.htm>

параметров оптимизации, выбор модели, обработка результатов эксперимента и их обсуждение, матричный подход к регрессионному анализу. Монография рассчитана на научных и инженерно-технических работников. Эта книга — одна из лучших для первого знакомства с математической теорией планирования экспериментов.

3. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ. Пер. с англ. М: Мир 1982.

Одна из лучших книг, посвященных применению статистического анализа в биомедицине. Книга может быть использована как при первом знакомстве с этим направлением, так и для углубления своих знаний в этой области. Все разделы иллюстрированы конкретными примерами, которые авторы брали из собственной практики.

4. Боровиков В.П. Программа STATISTICA для студентов и инженеров (2-е издание книги «Популярное введение в программу STATISTICA»). М: Компьютер Пресс 2001.

Книга является первым шагом к знакомству с ППП STATISTICA. На простых примерах из разных областей знаний показаны возможности системы по анализу данных. Освещаются следующие разделы статистики: описательная статистика, регрессионный, дискриминантный и кластерный анализ, анализ выживаемости.

5. Боровиков В.П. STATISTICA: Искусство анализа данных на компьютере. СПб: Питер 2001.

Отличие этой монографии от предыдущей книги этого же автора заключается в более подробном изложении ряда вопросов теории вероятностей и математической статистики. Очень подробно описан пользовательский интерфейс при работе с этим пакетом. Книга рекомендована профессионалам по анализу данных.

6. Боровиков В.П., Боровиков И.П. STATISTICA. Статистический анализ и обработка данных в среде Windows. М: Информационно-издательский дом «Филин» 1997.

Популярность пакета STATISTICA фирмы StatSoft общеизвестна. Достигается она двумя путями: наличием удобного поль-

зовательского интерфейса и наличием достаточно большого набора классических статистических процедур. Данная книга является толковым и весьма подробным описанием пользовательского интерфейса. Гораздо меньшее внимание уделено математическим аспектам реализованных процедур, в частности описанию ограничений тех или иных методов. Отсутствие таких разделов по “статистической безопасности” конкретных методов может подвести начинающих пользователей. Поэтому данную книгу можно рекомендовать лишь тем читателям, которые имеют достаточную теоретическую подготовку по математической статистике.

7. *Боровиков В.П., Ивченко Г.И.* Прогнозирование в системе STATISTICA® в среде Windows. Основы теории и интенсивная практика на компьютере. М: Финансы и статистика 2000. Рассматриваются математические основы и практические методы анализа временных рядов. В основу книги положен курс, читаемый авторами в Московском институте электроники и математики. Книга имеет гриф учебного пособия.
8. *Вейр Б.* Анализ генетических данных. Дискретные генетические признаки. Пер. с англ. М: Мир 1995. Автор этой книги первое высшее образование получил как математик, затем получил степень доктора наук по статистике и генетике. С 1976 г. работает профессором статистики и генетики на кафедре статистики в одном из университетов Новой Зеландии. Методы, описанные в этой монографии, широко применяются при реализации программы “Геном человека”.
9. *Владимирский Б.М.* Математические методы в биологии. Ростов-на-Дону: Изд-во Рост. ун-та 1983. В учебном пособии рассматриваются вопросы теории и практики использования математических методов и средств вычислительной техники для описания и моделирования процессов, протекающих в живых системах. На многочисленных примерах показано использование современных методов планирования исследований и анализа данных, существенно повышающих эффективность исследований в самых разных разделах биологии и медицины. Изложены методы оценки взаимосвязей, описательные статистики,

использование критериев согласия, элементы анализа динамических рядов, основные методы многомерного статистического анализа. В приложении даны некоторые статистические таблицы.

10. *Власов В.В.* Эффективность диагностических исследований. М: Медицина 1988.

Данная книга уникальна по многим причинам. Во-первых, она не имеет близких по своему содержанию русскоязычных аналогов. Во-вторых, изложение представляет собой гармоничное сочетание описания реальных медицинских проблем и профессионального обсуждения возникающих при этом статистических проблем, что свидетельствует о высоком уровне подготовки автора в этой области. Третье достоинство — большое количество цитируемой литературы, особенно зарубежной. И, наконец, методика изложения материала делает эту книгу одной из лучших по данной проблеме.

11. *Вучков И., Бояджиева Л., Солаков Е.* Прикладной линейный регрессионный анализ. Пер. с болг. и предисл. Ю.П. Адлера. М: Финансы и статистика 1987.

Книга систематизирует и обобщает новые направления регрессионного анализа, в том числе нарушения предпосылок (диагностика и пути преодоления), к которым относятся: нарушения предпосылки нормальности и нарушения однородности дисперсий. В качестве путей преодоления рассматриваются непараметрические методы оценивания, робастные (устойчивые) методы оценивания, анализ остатков, преобразования зависимых признаков.

12. *Гланц С.* Медико-биологическая статистика. Пер. с англ. М: Практика 1998.

Основное достоинство этой книги в том, что ее написал специалист, активно использующий биостатистику в собственных научных исследованиях. Именно поэтому язык книги прост и доходчив. В книге описаны многие из наиболее популярных статистических методов анализа экспериментальных наблюдений. К достоинствам книги можно отнести и описание условий применимости методов. Обилие примеров из реальной практики и их анализ делает книгу отличным пособием для первого знакомства с биостатистикой.

13. *Гласс Дж., Стэнли Дж.* Статистические методы в педагогике и психологии. Пер. с англ. М: Прогресс 1976.  
Одна из самых лучших книг по прикладной статистике для начинающих. Очень мало математики и много графики и толковых объяснений. В книге даны основные сведения по дескриптивным выборочным параметрам, понятие о проверке гипотез, дисперсионный анализ и основы планирования экспериментов.
14. *Гроссман С., Тернер Дж.* Математика для биологов. Пер. с англ. М: Высшая школа 1983.  
Обилие разных по содержанию разделов, не всегда логично связанных между собой, делает эту книгу введением в ряд областей математики, методы которых могут быть использованы в тех или иных задачах. Можно рекомендовать как учебник при изучении соответствующих разделов в вузе. Весьма мало материала по теории вероятностей и математической статистике.
15. *Гублер Е.В., Генкин А.А.* Применение непараметрических критериев статистики в медико-биологических исследованиях. Л: Медицина 1973.  
В книге рассмотрены различные непараметрические критерии, приведены многочисленные примеры, отображающие опыт авторов в их применении, описана методика выбора наиболее адекватного критерия в каждом случае. Обширное приложение содержит необходимые таблицы.
16. *Двойрин В.В., Клименков А.А.* Методика контролируемых клинических испытаний. М: Медицина 1985.  
Первая монография на русском языке, в которой описаны типы и организационные формы проведения клинических испытаний, проблемы формирования групп наблюдений, расчет численности групп, критерии оценки эффекта лечения, влияние дополнительной терапии, особенности изучения эффективности новых методов диагностики. Довольно подробно излагаются методы статистической оценки эффекта лечения.
17. *Дэйвисон М.* Многомерное шкалирование: Методы наглядного представления данных. Пер. с англ. М: Финансы и статистика 1988.

Книга американского ученого посвящена многомерному шкалированию, которое находит широкое применение при анализе данных в медицине, биологии, экономике, психологии, социологии и других областях. Вместе с кластерным и факторным анализом оно образует набор методов, позволяющих наглядно представить результаты и таким образом лучше изучить структуру исследуемых данных.

18. *Ермаков С.М., Жиглявский А.А.* Математическая теория оптимального эксперимента. М: Наука 1987.

Прекрасная математическая монография, написанная известными ленинградскими математиками. Можно рекомендовать специалистам, непосредственно работающим в лабораториях биостатистики и имеющих математическое образование. Не рекомендуется для первого знакомства с данным направлением статистики.

19. *Иберла К.* Факторный анализ. М: Статистика 1980.

Книга посвящена многомерному статистическому анализу. В ней в комплексе рассматриваются факторный, компонентный, дискриминантный и другие виды анализа. Представлены методы оценки весовых коэффициентов, интерпретация факторных решений. Задачи решаются применительно к современным ЭВМ. Книга предназначена для ученых и практиков, а также может служить учебным пособием для студентов и аспирантов.

20. *Компьютерная биометрика.* Под ред. В.Н. Носова. М: МГУ 1990.

В монографии рассмотрены современные методы статистической обработки данных и использование соответствующего программного обеспечения ЭВМ в биологических исследованиях. Изложены элементы теории вероятностей как математические основы статистических методов. Даны рекомендации по использованию пакетов прикладных программ (STATGRAPHICS, SYSTAT, SAS, BMDP) и алгоритмических языков ЭВМ. Книга предназначена для специалистов биологов и медиков.

21. *Леонов В.П.* Обработка экспериментальных данных на программируемых микрокалькуляторах. Под ред. Б.А. Гладких. Томск: Изд-во ТГУ 1990.

Кроме методов оценки основных статистических характеристик — среднего, дисперсии, дано изложение корреляционно-регрессионного, однофакторного и двухфакторного дисперсионного анализа, ковариационного анализа. Несколько глав посвящены методам проверки гипотез о векторах средних, дискриминантному анализу и методу главных компонент. Использование методов иллюстрируется задачами из биологии, медицины, ботаники, экономики, электроники и т. д.

22. *Лисенков А.Н.* Математические методы планирования многофакторных медико-биологических экспериментов. М: Медицина 1979.

В книге рассматриваются принципы планирования многофакторных экспериментов. Описаны элементы дисперсионного анализа, планы  $2^k$ , многоуровневые и несимметричные факторные планы, методы экстремального и отсеивающего эксперимента, процедуры множественных сравнений и непараметрические методы анализа многофакторных экспериментов. Излагаемый материал иллюстрируется примерами планирования и анализа, реальных медико-биологических экспериментов.

23. *Любичев А.А.* Дисперсионный анализ в биологии. М: МГУ 1986.

Книга посвящена специфике использования дисперсионного анализа в биологии. Насыщенность книги реальными примерами делает ее незаменимым учебником.

24. *Максимов В.Н.* Многофакторный эксперимент в биологии. М: МГУ 1980.

На примерах, взятых из практики биологических исследований, изложены основные принципы планирования многофакторных экспериментов. Подробно рассматриваются вопросы статистической обработки результатов экспериментов и их интерпретация на основе получаемых уравнений регрессии. Описаны полный и дробный факторные планы. Книга для биологов, биохимиков, физиологов, микробиологов, гидробиологов и экологов.

25. *Минцер О.П., Угаров Б.Н., Власов В.В.* Методы обработки медицинской информации. Уч. пособие. Киев: Выща школа 1991.

В книге рассмотрены различные аспекты обработки медицинской информации, изложены основы теории вероятностей, типы распределения случайных величин и их характеристики, основы теории корреляции, гармонического анализа и анализа временных рядов, методы проверки статистических гипотез и аппроксимаций данных с использованием метода наименьших квадратов.

26. *Миркин Б.Г.* Анализ качественных признаков и структур. М: Статистика 1990.

Систематически излагаются методы анализа качественных данных. Данные такого типа преобладают в клинической медицине.

27. *Мисюк Н.С., Мастыкин А.С., Кузнецов Г.П.* Корреляционно-регрессионный анализ в клинической медицине. М: Медицина 1975.

В книге, написанной специально для врачей, в доступной форме изложены методы парной и множественной корреляции, а также множественной регрессии. Отдельная глава посвящена анализу взаимосвязей качественных признаков. Приводятся примеры исследования взаимосвязей биохимических и волюметрических характеристик крови.

28. *Мюллер П., Нойман П., Шторм Р.* Таблицы по математической статистике. М: Финансы и статистика 1982.

Сборник представляет собой подбор таблиц, с помощью которых можно обрабатывать наиболее часто встречающиеся статистические задачи. Каждая таблица снабжена четкими индивидуальными определениями. Сфера пользования таблицами поясняется на простых примерах. Таблицы предназначены для работников самых разнообразных профессий и учащихся.

29. *Планирование эксперимента в биологии и сельском хозяйстве.* Под ред. В.Н. Максимова. М: МГУ 1991.

В учебном пособии популярно изложены основные методы математической теории планирования экспериментов. Изложена теория оптимальности планов, принципы регрессионного анализа. Приведены примеры практического использо-



вания многофакторных планов. Описаны пакеты прикладных программ по планированию экспериментов, факторных и полиномиальных моделей.

30. *Платонов А.Е.* Статистический анализ данных в медицине и биологии: задачи, терминология, логика, компьютерные методы. М: Изд-во РАМН 2000.

Брошюра содержит прагматический подход к анализу данных в медицинских и биологических исследованиях. Изложены принципы нескольких наиболее популярных статистических методов, условия их применимости. Может быть рекомендована читателям, впервые приступающим к анализу данных.

31. *Поллард Дж.* Справочник по вычислительным методам статистики. М: Финансы и статистика 1982.

В книге дается набор вычислительных процедур численного анализа и статистических приемов оценки гипотез. Рассматриваются основные понятия математической статистики, разного рода ошибки и способы их устранения, методы сглаживания, интерполяция и численное дифференцирование, статистические распределения, метод наименьших квадратов, регрессионный анализ. Наиболее полно представлена проверка статистических гипотез. Книга содержит множество примеров и таблиц, предназначена для научных работников, практиков, аспирантов и студентов.

32. *Рунион Р.* Справочник по непараметрической статистике. Современный подход. М: Финансы и статистика 1982.

Книга посвящена непараметрическим методам исследования (критерий  $\chi^2$ , Q-критерий Кокрена, критерий Колмогорова—Смирнова и др.). Кратко излагается теория вероятностей. Последовательно рассматриваются непараметрические критерии проверки гипотез при различных выборках. Детально показана техника вычислений. Даются задания для самостоятельной работы.

33. *Сергиенко В.И., Бондарева И.П.* Математическая статистика в клинических исследованиях. М: ГЭОТАР-МЕД 2001.

Излагаются основные методы статистического анализа данных клинических исследований, приведены математические

формулы. Показаны возможности и ограничения этих методов, основные ошибки, возникающие при их некорректном применении. Основной интерес в книге представляет глава с изложением основных подходов при изучении фармакокинетики и фармакодинамики.

34. *Снедекор Дж.У.* Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. М: Сельхозгиз 1961.

В книге описаны методы математической статистики, используемые при обработке результатов экспериментов. Рассмотрены виды распределения, различные параметрические и непараметрические методы, регрессионный, корреляционный, дисперсионный анализ, ковариация, планирование и анализ выборочных наблюдений. Книга имеет характер практического руководства, в ней нет математического обоснования и вывода рекомендуемых формул. Материал изложен в доступной форме, приведено много таблиц, необходимых для оценки статистической значимости выводов, полученных при обработке данных.

35. *Сошникова Л.А., Тамашевич В.Н., Уебе Г., Шефер М.* Многомерный статистический анализ в экономике: Уч. пособие для вузов. Под ред. В.Н. Тамашевича. М: ЮНИТИ-ДАНА 1999. Достаточно полно представлены теоретические основы и важнейшие методы многомерной статистики. В частности, рассмотрены такие методы, как многомерный регрессионный анализ, факторный анализ, многомерное шкалирование, кластерный анализ, дискриминантный анализ, метод канонических корреляций. Достаточно подробно рассмотрен и вопрос оценки аномальных наблюдений (выбросов) и получения робастных (устойчивых) статистических оценок путем удаления выбросов.

36. *Справочник по прикладной статистике.* В 2 т. Пер. с англ. Под ред. Э. Ллойда, У. Ледермана, Ю.Н. Тюрина. М: Финансы и статистика Т. 1: 1989, Т. 2: 1990.

В двух томах концентрированно изложены основные методы прикладной статистики. Интересные реальные примеры помогают читателю легко усвоить основные идеи того или иного метода.

37. *Статистические методы в клинических испытаниях*: Сб. статей под ред. А.А. Жиглявского, В.В. Некруткина. СПб: Издательство С-ПбГУ 1999.

Очень ценная книга, написанная специалистами в области прикладной статистики. Отдельные разделы книги прекрасно дополняют друг друга. Особо следует отметить раздел, посвященный методологии анализа данных. Нередко начинающие исследователи задумываются о том, какие выбрать методы анализа данных уже после того, как эти данные собраны. Авторы статей показывают порочность подобной практики. Обилие реальных примеров, проанализированных профессионалами высокого класса, делают увлекательным чтение этого сборника. Авторы еще раз убедительно доказали, что питерская школа прикладной биостатистики по-прежнему остается ведущей в России.

38. *Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере*. Под ред. В.Э. Фигурнова. М: ИНФРА-М, Финансы и статистика 1995.

Книга весьма полезна для начинающих знакомство с возможностями компьютерного анализа данных. Доступное и в то же время строгое изложение всего материала делает ее незаменимым учебником. К недостаткам издания можно отнести то, что многомерным методам в ней посвящено всего лишь 6 страниц.

39. *Ким Дж.О., Мюллер Ч.У., Клекка У.Р. и др. Факторный, дискриминантный и кластерный анализ*. Пер. с англ. Под ред. И.С. Енюкова. М: Финансы и статистика 1989.

Лучшая из всех известных автору монографий, содержащих описание и примеры использования таких сложных многомерных методов, как факторный, дискриминантный и кластерный анализ. Книга содержит три отдельных очерка по этим методам, каждый из которых можно читать и изучать приведенные в них материалы отдельно. Используя минимум математики, авторам удалось рассказать просто о сложном. Книга может быть рекомендована как для первого знакомства с этими методами, так и для тех, кто уже знаком с ними.

40. *Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа*. М: Финансы и статистика 1983.

Книга представляет собой пособие по широко применяемым методам статистической обработки данных — корреляционному и регрессионному анализу. Дает читателю, не знакомому со статистическим анализом, ключ к пониманию прикладных работ в интересующих его областях.

41. *Флейс Дж.* Статистические методы для изучения таблиц долей и пропорций. Пер. с англ. Под ред. Ю.Н. Длаговещенского. М: Финансы и статистика 1989.

Единственное издание, в котором столь глубоко и корректно рассмотрены весьма специфические задачи анализа таблиц долей и пропорций. Книга доступна для чтения и понимания даже при минимальной подготовке по математике.

42. *Флетчер Р., Флетчер С., Вагнер Э.* Клиническая эпидемиология. Основы доказательной медицины. М: МедиаСфера 1998.

Единственная на сегодняшний день книга на русском языке, в которой подробно рассматриваются основные положения клинической эпидемиологии — научной основы доказательной медицины — и ее применение в исследованиях методов лечения, профилактики, диагностики. Подробно описаны источники систематических ошибок, которые могут возникать при проведении исследований, и способы их минимизации. Рассмотрены основные подходы к статистическому анализу данных в зависимости от дизайна исследования.

43. *Фор А.* Восприятие и распознавание образов. Пер. с франц. М: Машиностроение 1989.

В книге проблема распознавания образов рассматривается комплексно: от датчиков сигнала до применения в различных областях. Математический аппарат используется в минимальной степени.

44. *Харман Г.* Современный факторный анализ. М: Статистика 1972.

В книге рассмотрены основные методы факторного анализа, проверка статистических гипотез, применение ЭВМ в факторном анализе, понятия и методы матричной алгебры. Приведены задачи и упражнения для освоения факторного анализа.

45. Хургин Я.И. Да, нет или может быть... 2-е изд. М: Наука 1983. В книге рассказывается о статистической теории управления и эксперимента — научном направлении, находящемся на стыке нескольких наук: теории управления, математической статистики и теории эксперимента. На примерах из разных областей науки и техники в доступной форме разъясняются принципиальные положения статистической теории управления и эксперимента, освещаются современные принципы построения математических моделей и обработки результатов наблюдений, статистической проверки гипотез и планирования экспериментов. Книга адресована широкому кругу читателей, интересующихся проблемами прикладной математики, в том числе специалистам, так или иначе связанным с экспериментом, но не имеющим достаточной математической подготовки, особенно в области теории вероятностей и ее применения (инженеры, геологи, химики, биологи, медики, экономисты).
46. Хургин Я.И. Как объять необъятное. М: Знание 1979. В этой научно-популярной книге живо излагаются идеи современной математической статистики, раскрывается ее основа — теория вероятностей, обсуждаются проблемы измерения, описания сложных систем и построения их математических моделей. Предназначается для широкого круга читателей. По мнению составителя этого обзора, данную книгу должен прочитать каждый, кто желает осмысленно воспринять основные идеи статистики. Автор книги — Яков Исаевич Хургин, известный популяризатор статистики. Кроме этой книги, его перу принадлежат еще несколько изданий на эту и близкие темы. Книга написана с любовью к статистике и с глубоким знанием предмета. После прочтения этой книги многие биологи и медики будут по-другому относиться к своим данным и более критично воспринимать результаты своих коллег.
47. Холлендер М., Вулф Д. Непараметрические методы статистики. Пер. с англ. Д.С. Шмерлинга. М: Финансы и статистика 1983. Книга представляет собой руководство по непараметрической статистике для пользователей. Методы непараметрической статистики характеризуются обширной областью приложения, математической простотой и доступностью при довольно высокой эффективности.

---

## Приложение 2. Список рекомендуемой литературы на английском языке

1. *Altman D.G.* Practical Statistics for Medical Research. London, Chapman & Hall 1991.
2. Data Presentation, Interpretation. Ed. H. Keller, Ch. Trendelenburg. Berlin; New York: Walter deGruyter 1989.
3. *Lang T.A., Secic M.* How to report statistics in medicine. Annotated guidelines for authors, editors, and reviewers. Copyright by American College of Physicians, Phyladelphia, Pennsylvania, printing/binding by Port City Press, composition by Techsetters 1997.
4. *Motulsky H.* Intuitive Biostatistics. NY, Oxford: Oxford University Press, 1995.

### Приложение 3. Ранги, с помощью которых вычисляется доверительный интервал для медианы распределения

Значения в таблице — ранги в упорядоченном по возрастанию списке наблюдений, которые дают примерно 90, 95% или 99% ДИ для медианы популяции на основании одной выборки.

*Пример:* 99% ДИ для популяционной медианы, вычисленной по выборке объемом 56 наблюдений, определяется наблюдениями с рангами 18 и 39.

Объем выборки	Доверительная вероятность		
	90%	95%	99%
6	1; 6	1; 6	—
7	1; 7	1; 7	—
8	2; 7	1; 8	1; 8
9	2; 8	2; 8	1; 9
10	2; 9	2; 9	1; 10
11	3; 9	2; 10	1; 11
12	3; 10	3; 10	2; 11
13	4; 10	3; 11	2; 12
14	4; 11	3; 12	2; 13
15	4; 12	4; 12	3; 13
16	5; 12	4; 13	3; 14
17	5; 13	5; 13	3; 15
18	6; 13	5; 14	4; 15
19	6; 14	5; 15	4; 16
20	6; 15	6; 15	4; 17
21	7; 15	6; 16	5; 17
22	7; 16	6; 17	5; 18
23	8; 16	7; 17	5; 19

Продолжение таблицы на стр. 287

Таблица. Продолжение

Объем выборки	Доверительная вероятность		
	90%	95%	99%
24	8; 17	7; 18	6; 19
25	8; 18	8; 18	6; 20
26	9; 18	8; 19	7; 20
27	9; 19	8; 20	7; 21
28	10; 19	9; 20	7; 22
29	10; 20	9; 21	8; 22
30	11; 20	10; 21	8; 23
31	11; 21	10; 22	8; 24
32	11; 22	10; 23	9; 24
33	12; 22	11; 23	9; 25
34	12; 23	11; 24	10; 25
35	13; 23	12; 24	10; 26
36	13; 24	12; 25	10; 27
37	14; 24	13; 25	11; 27
38	14; 25	13; 26	11; 28
39	14; 26	13; 27	12; 28
40	15; 26	14; 27	12; 29
41	15; 27	14; 28	12; 30
42	16; 27	15; 28	13; 30
43	16; 28	15; 29	13; 31
44	17; 28	16; 29	14; 31
45	17; 29	16; 30	14; 32
46	17; 30	16; 31	14; 33
47	18; 30	17; 31	15; 33
48	18; 31	17; 32	15; 34
49	19; 31	18; 32	16; 34
50	19; 32	18; 33	16; 35
51	20; 32	19; 33	16; 36
52	20; 33	19; 34	17; 36
53	21; 33	19; 35	17; 37

Продолжение таблицы на стр. 288



Таблица. Продолжение

Объем выборки	Доверительная вероятность		
	90%	95%	99%
54	21; 34	20; 35	18; 37
55	21; 35	20; 36	18; 38
56	22; 35	21; 36	18; 39
57	22; 36	21; 37	19; 39
58	23; 36	22; 37	19; 40
59	23; 37	22; 38	20; 40
60	24; 37	22; 39	20; 41
61	24; 38	23; 39	21; 41
62	25; 38	23; 40	21; 42
63	25; 39	24; 40	21; 43
64	25; 40	24; 41	22; 43
65	26; 40	25; 41	22; 44
66	26; 41	25; 42	23; 44
67	27; 41	26; 42	23; 45
68	27; 42	26; 43	23; 46
69	28; 42	26; 44	24; 46
70	28; 43	27; 44	24; 47
71	29; 43	27; 45	25; 47
72	29; 44	28; 45	25; 48
73	29; 45	28; 46	26; 48
74	30; 45	29; 46	26; 49
75	30; 46	29; 47	26; 50
76	31; 46	29; 48	27; 50
77	31; 47	30; 48	27; 51
78	32; 47	30; 49	28; 51
79	32; 48	31; 49	28; 52
80	33; 48	31; 50	29; 52
81	33; 49	32; 50	29; 53
82	34; 49	32; 51	29; 54
83	34; 50	33; 51	30; 54

Продолжение таблицы на стр. 289

Таблица. Продолжение

Объем выборки	Доверительная вероятность		
	90%	95%	99%
84	34; 51	33; 52	30; 55
85	35; 51	33; 53	31; 55
86	35; 52	34; 53	31; 56
87	36; 52	34; 54	32; 56
88	36; 53	35; 54	32; 57
89	37; 53	35; 55	32; 58
90	37; 54	36; 55	33; 58
91	38; 54	36; 56	33; 59
92	38; 55	37; 56	34; 59
93	39; 55	37; 57	34; 60
94	39; 56	38; 57	35; 60
95	39; 57	38; 58	35; 61
96	40; 57	38; 59	35; 62
97	40; 58	39; 59	36; 62
98	41; 58	39; 60	36; 63
99	41; 59	40; 60	37; 63
100	42; 59	40; 61	37; 64

## **Приложение 4. Нижние и верхние границы 95% доверительных интервалов для относительных частот, представленных в виде простых дробей**

В заголовках столбцов — числители дроби, в заголовках строк — знаменатели дроби.

*Пример:* если осложнение выявлено у 5 человек из 30, то относительная частота равна 0,167 и 95% ДИ — [0,056; 0,347].

Таблица

Знаменатель	Числитель						
	0	1	2	3	4	5	6
1	.000 .975	.025 1.00					
2	.000 .842	.013 .987	.158 1.00				
3	.000 .708	.008 .906	.094 .992	.292 1.00			
4	.000 .602	.006 .806	.068 .932	.194 .994	.398 1.00		
5	.000 .522	.005 .716	.053 .853	.147 .947	.284 .995	.478 1.00	
6	.000 .459	.004 .641	.043 .777	.118 .882	.223 .957	.359 .996	.541 1.00
7	.000 .410	.004 .579	.037 .710	.099 .816	.184 .901	.290 .963	.421 .996
8	.000 .369	.003 .527	.032 .651	.085 .755	.157 .843	.245 .915	.349 .968
9	.000 .336	.003 .482	.028 .600	.075 .701	.137 .788	.212 .863	.299 .925
10	.000 .308	.003 .445	.025 .556	.067 .652	.122 .738	.187 .813	.262 .878
11	.000 .285	.002 .413	.023 .518	.060 .610	.109 .692	.167 .766	.234 .833
12	.000 .265	.002 .385	.021 .484	.055 .572	.099 .651	.152 .723	.211 .789
13	.000 .247	.002 .360	.019 .454	.050 .538	.091 .614	.139 .684	.192 .749
14	.000 .232	.002 .339	.018 .428	.047 .508	.084 .581	.128 .649	.177 .711
15	.000 .218	.002 .319	.017 .405	.043 .481	.078 .551	.118 .616	.163 .677
16	.000 .206	.002 .302	.016 .383	.040 .456	.073 .524	.110 .587	.152 .646
17	.000 .195	.001 .287	.015 .364	.038 .434	.068 .499	.103 .560	.142 .617
18	.000 .185	.001 .273	.014 .347	.036 .414	.064 .476	.097 .535	.133 .590
19	.000 .176	.001 .260	.013 .331	.034 .396	.061 .456	.091 .512	.126 .566
20	.000 .168	.001 .249	.012 .317	.032 .379	.057 .437	.087 .491	.119 .543
21	.000 .161	.001 .238	.012 .304	.030 .363	.054 .419	.082 .472	.113 .522
22	.000 .154	.001 .228	.011 .292	.029 .349	.052 .403	.078 .454	.107 .502
23	.000 .148	.001 .219	.011 .280	.028 .336	.050 .388	.075 .437	.102 .484
24	.000 .142	.001 .211	.010 .270	.027 .324	.047 .374	.071 .422	.098 .467
25	.000 .137	.001 .204	.010 .260	.025 .312	.045 .361	.068 .407	.094 .451
26	.000 .132	.001 .196	.009 .251	.024 .302	.044 .349	.066 .394	.090 .436
27	.000 .128	.001 .190	.009 .243	.024 .292	.042 .337	.063 .381	.086 .423
28	.000 .123	.001 .183	.009 .235	.023 .282	.040 .327	.061 .369	.083 .410
29	.000 .119	.001 .178	.008 .228	.022 .274	.039 .317	.058 .358	.080 .397
30	.000 .116	.001 .172	.008 .221	.021 .265	.038 .307	.056 .347	.077 .386

Продолжение на стр. 292

## Продолжение таблицы

Знаменатель	Числитель						
	7	8	9	10	11	12	13
1							
2							
3							
4							
5							
6							
7	.590 1.00						
8	.473 .997	.631 1.00					
9	.400 .972	.518 .997	.664 1.00				
10	.348 .933	.444 .975	.555 .997	.692 1.00			
11	.308 .891	.390 .940	.482 .977	.587 .998	.715 1.00		
12	.277 .848	.349 .901	.428 .945	.516 .979	.615 .998	.735 1.00	
13	.251 .808	.316 .861	.386 .909	.462 .950	.546 .981	.640 .998	.753 1.00
14	.230 .770	.289 .823	.351 .872	.419 .916	.492 .953	.572 .982	.661 .998
15	.213 .734	.266 .787	.323 .837	.384 .882	.449 .922	.519 .957	.595 .983
16	.198 .701	.247 .753	.299 .802	.354 .848	.413 .890	.476 .927	.544 .960
17	.184 .671	.230 .722	.278 .770	.329 .816	.383 .858	.440 .897	.501 .932
18	.173 .643	.215 .692	.260 .740	.308 .785	.357 .827	.410 .867	.465 .903
19	.163 .616	.203 .665	.244 .711	.289 .756	.335 .797	.384 .837	.434 .874
20	.154 .592	.191 .639	.231 .685	.272 .728	.315 .769	.361 .809	.408 .846
21	.146 .570	.181 .616	.218 .660	.257 .702	.298 .743	.340 .782	.384 .819
22	.139 .549	.172 .593	.207 .636	.244 .678	.282 .718	.322 .756	.364 .793
23	.132 .529	.164 .573	.197 .615	.232 .655	.268 .694	.306 .732	.345 .768
24	.126 .511	.156 .553	.188 .594	.221 .634	.256 .672	.291 .709	.328 .744
25	.121 .494	.149 .535	.180 .575	.211 .613	.244 .651	.278 .687	.313 .722
26	.116 .478	.143 .518	.172 .557	.202 .594	.234 .631	.266 .666	.299 .701
27	.111 .463	.138 .502	.165 .540	.194 .576	.224 .612	.255 .647	.287 .681
28	.107 .449	.132 .487	.159 .524	.186 .559	.215 .594	.245 .628	.275 .661
29	.103 .435	.127 .472	.153 .508	.179 .543	.207 .577	.235 .611	.264 .643
30	.099 .423	.123 .459	.147 .494	.173 .528	.199 .561	.227 .594	.255 .626

Продолжение на стр. 293

## Продолжение таблицы

Знаменатель	Числитель					
	14	15	16	17	18	19
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14	.768 1.00					
15	.681 .998	.782 1.00				
16	.617 .984	.698 .998	.794 1.00			
17	.566 .962	.636 .985	.713 .999	.805 1.00		
18	.524 .936	.586 .964	.653 .986	.727 .999	.815 1.00	
19	.488 .909	.544 .939	.604 .966	.669 .987	.740 .999	.824 1.00
20	.457 .881	.509 .913	.563 .943	.621 .968	.683 .988	.751 .999
21	.430 .854	.478 .887	.528 .918	.581 .946	.637 .970	.696 .988
22	.407 .828	.451 .861	.498 .893	.546 .922	.597 .948	.651 .971
23	.385 .803	.427 .836	.471 .868	.516 .898	.563 .925	.612 .950
24	.366 .779	.406 .812	.447 .844	.489 .874	.533 .902	.578 .929
25	.349 .756	.387 .789	.425 .820	.465 .851	.506 .879	.549 .906
26	.334 .734	.369 .766	.406 .798	.443 .828	.482 .857	.522 .884
27	.319 .713	.353 .745	.388 .776	.424 .806	.460 .835	.498 .862
28	.306 .694	.339 .725	.372 .755	.406 .785	.441 .814	.476 .841
29	.294 .675	.325 .706	.357 .736	.389 .765	.423 .793	.457 .821
30	.283 .657	.313 .687	.343 .717	.374 .745	.406 .773	.439 .801

## Толковый словарь терминов

**Априори** — до проведения исследования (противоположное — *апостериори*)

**Абсолютного риска повышение** см. *Атрибутивный риск*

**Актуральный метод** — метод анализа времени до наступления изучаемого исхода с помощью таблиц дожития

**Альтернативная гипотеза** — статистическая гипотеза о существовании различий между группами, являющаяся обычно мотивацией для проведения исследования. В многофакторном анализе — предположение о том, что факторы, отличные от случайных, обуславливают различия между группами. Альтернативная гипотеза — противоположность нулевой гипотезе

**Анализ времени до наступления изучаемого исхода** — статистические процедуры для анализа времени между точкой включения участника в исследование и определенным изучаемым исходом (например, выздоровлением; для смерти — анализ выживаемости)

**Ассоциация** — более общий термин, чем корреляция. Обычно применяется при анализе взаимосвязи качественных признаков

**Атрибутивный риск** — дополнительный риск возникновения неблагоприятного исхода (например, заболевания) в связи с наличием определенной характеристики (фактора риска) у объекта исследования. Это часть риска развития болезни, которая связана с данным фактором риска, объясняется им и может быть устранена, если этот фактор риска устранить

**Байесовская статистика** — статистика, основанная на теореме Томаса Байеса, описывающей математическую взаимосвязь априорной (претестовой), условной и апостериорной (послетестовой) вероятностей. Используется главным образом в оценке диагностических тестов

**Бинарные (дихотомические) данные** — данные, выражаемые только двумя допустимыми альтернативными значениями (например, “есть” — “нет”, “здоров” — “болен”)

**Бонферрони поправка** — один из способов преодоления проблемы множественных сравнений

**Вариация** — степень рассеяния данных (значений признака) по области значений

**Вероятность** — число между 0 и 1, отражающее относительную частоту появления события. Сумма альтернативных вероятностей равна 1.

**Вмешивающийся фактор** — независимый признак, связанный с другим исследуемым независимым признаком и влияющий на исследуемый зависимый признак (исход)

**Вторичный анализ данных** — анализ данных, не предусмотренный в протоколе (плане, структуре) исследования. Результаты вторичного анализа данных являются основанием для выдвижения научных гипотез. Интерпретация результатов вторичного анализа данных должна быть весьма осторожной, поскольку часто они являются порождением систематического различия между сопоставляемыми группами

**Выборка** — часть популяции, полученная путем отбора. По результатам анализа выборки делают выводы о всей популяции, что правомерно только в случае, если отбор был случайным. Поскольку случайный отбор из популяции осуществить практически невозможно, следует стремиться к тому, чтобы выборка была по крайней мере репрезентативна по отношению к популяции

**Выпадающие значения** — значения количественного признака, располагающиеся на краях интервала допустимых значений. Их обычно мало, но они могут в значительной степени влиять на параметры распределения

**Гипотеза научная** — утверждение, которое можно подтвердить или опровергнуть на основании результатов исследования

**Гипотеза статистическая** — представление научной гипотезы в форме, приемлемой для проверки методами статистического анализа данных



**Гистограмма** — тип графика для изображения распределения данных или его параметров (среднего, дисперсии и др.)

**Двусторонний тест** — условие проверки гипотез, определяемое до сбора данных. Двусторонний тест не предполагает, что заранее известно направление смещения значения анализируемого параметра одной группы по отношению к другой. Двусторонний тест носит более универсальный характер, чем односторонний. Он также более консервативен по сравнению с односторонним тестом

**Дискретные данные** — данные, выражаемые ограниченным набором значений (обычно целыми числами). Альтернатива непрерывным данным

**Дискриминантный анализ** — многофакторный статистический метод построения моделей и выявления комбинаций наиболее информативных признаков, характеризующих номинальные категории (группы)

**Доверительный интервал (ДИ)** — интервал значений признака, рассчитанный для какого-либо параметра распределения (например, среднего значения признака) по выборке и с определенной вероятностью (например, 95% для 95% ДИ) включающий истинное значение этого параметра во всей популяции

**Доказательная медицина** — концепция использования в медицинской практике научно обоснованных фактов, полученных в ходе корректно проведенных исследований

**Достоверность** — степень, с которой измерение отражает истинное значение измеряемого признака. Достоверность исследования (внутренняя обоснованность исследования) определяется тем, в какой мере полученные результаты справедливы в отношении данной выборки

**Зависимый признак** — признак, значение которого может быть рассчитано по значениям других признаков

**Значение  $p$**  — рассчитанная в ходе статистического теста вероятность ошибочного отклонения нулевой гипотезы. Для принятия решения о том, необходимо ли отклонить нулевую гипотезу по результатам статистического теста, значение  $p$  сравнивают с принятым исследователем критическим (пороговым) уровнем значи-

мости ( $\alpha$ -ошибкой). Если значение  $p$  меньше уровня значимости, то нулевая гипотеза отклоняется, и принимается альтернативная гипотеза. В обратном случае нулевая гипотеза не отклоняется.

**Интерквартильный размах** (*интервал, отрезок*) — интервал значений признака, содержащий центральные 50% наблюдений выборки, т.е. интервал между 25-м и 75-м процентилями. Используется вместе с медианой (вместо  $M \pm s$ ) для описания данных, имеющих распределение, отличное от нормального

**Исход** — клинически значимое явление (клинический исход, событие — см.), лабораторный показатель или признак (промежуточный исход, или косвенный критерий оценки — см.), который служит объектом интереса исследователя. При проведении клинических испытаний исходы служат критериями оценки эффективности лечебного или профилактического воздействия

**Категориальные данные** см. *Качественные данные*

**Качественные данные** — номинальные и порядковые данные, которые отражают условные коды неизмеряемых категорий или условную степень выраженности признака

**Клиническая эпидемиология** — наука, позволяющая осуществлять прогнозирование того или иного исхода для каждого конкретного больного на основании изучения клинического течения болезни в аналогичных случаях с использованием строгих научных методов изучения групп больных для обеспечения точности прогнозов.

**Клинический исход** — см. *Событие*

**Коэффициент доверительный** — процент, показывающий степень уверенности автора в оценке границ доверительного интервала. Коэффициенту 0,95 соответствует 95% ДИ

**Коэффициент корреляции** — мера линейной связи количественных или порядковых признаков

**Когорта** — группа участников исследования, объединенных каким-либо общим признаком в момент ее формирования и исследуемых на протяжении длительного периода времени

**Количественные данные** — данные, измеряемые в шкале, поделенной на равные интервалы. Количественные данные могут быть непрерывными или дискретными

**Контроль исторический** — контрольная группа, сформированная и обследованная в период, предшествующий исследованию

**Контроль параллельный** — контрольная группа, формируемая одновременно с формированием основной группы

**Контролируемое исследование** — исследование, в котором данные собираются от основной (экспериментальной) и контрольной групп

**Корреляция** — статистическая связь двух признаков (количественных или порядковых), показывающая, что большему значению одного признака в определенной части случаев соответствует большее — в случае положительной (прямой) корреляции — значение другого признака или меньшее значение — в случае отрицательной (обратной) корреляции

**Косвенный критерий оценки (суррогатный исход)** — некоторый исследуемый признак (обычно лабораторный или инструментальный показатель), который, как предполагается, связан с клиническим исходом

**Медиана** — значение признака, разделяющее пополам распределение наблюдений на интервале значений признака, т.е. половина значений в выборке меньше медианы, а вторая половина — больше ее. Используется для описания центральной тенденции распределений количественных и порядковых признаков независимо от его закона

**Меры дисперсии** — статистики, описывающие вариабельность значений признака (дисперсия, стандартное отклонение, размах, интерквартильный размах)

**Мета-анализ** — статистический анализ, обобщающий результаты нескольких исследований, исследующих одну и ту же проблему (обычно эффективность методов лечения, профилактики, диагностики). Объединение исследований обеспечивает большую выборку для анализа и большую статистическую мощность объединяемых исследований. Используется для повышения доказательности или уверенности в заключении об эффективности исследуемого метода

**Мода** — описательная статистика. Соответствует значению признака, наиболее часто встречающемуся в исследуемой выборке

**Модель статистическая** — математическое уравнение, описывающее (в большей или меньшей степени) взаимосвязь между признаками (переменными)

**Многофакторный анализ** — анализ влияния двух объясняющих (независимых) признаков и более на один объясняемый (зависимый) признак

**Множественных сравнений проблема** — проблема завышения статистической значимости результатов статистических тестов, возникающая при выполнении большого числа статистических тестов на одной выборке

**Множественных сравнений процедура** — процедура попарного сопоставления нескольких групп, используемая после выявления статистически значимых различий между группами в дисперсионном анализе

**Мощность статистическая исследования** — вероятность обнаружить определенное различие, когда оно на самом деле существует; вероятность правильно отклонить нулевую гипотезу. Статистическая мощность обычно принимается равной 0,8 или 0,9. В проспективном исследовании она должна быть определена на стадии планирования и описана в разделе “Материалы и методы” при описании определения размера выборки. Статистическая мощность исследования прямо пропорциональна размеру выборки

**Мощность статистического теста** — вероятность того, что статистический тест при заданных объемах групп выявит имеющееся различие между этими группами как статистически значимое

**Независимый (объясняющий) признак** — признак, который включается в статистическую модель с целью анализа его влияния на зависимый (объясняемый) признак. С учетом его значений можно рассчитать значение зависимого признака

**Независимые (несвязанные) выборки** — выборки, в которые объекты исследования набирались независимо друг от друга. Альтернатива независимым выборкам — зависимые (связанные, парные) выборки

**Непараметрические методы статистики** — класс статистических методов, которые используются главным образом для

анализа количественных данных, не образующих нормальное распределение, а также для анализа качественных данных

**Непрерывные данные** — это количественные данные, принимающие значения на непрерывной шкале значений. Альтернативой непрерывным данным служат дискретные данные

**Номинальные данные** — это качественные данные, которые отражают условные коды неизмеряемых категорий, не подлежащих упорядочиванию. Альтернативой служат качественные порядковые данные, которые могут быть ранжированы в убывающем или возрастающем порядке

**Номограмма** — графическая схема, состоящая из нескольких шкалированных линий, расположенных таким образом, чтобы отрезок, соединяющий известные значения на двух шкалах, показывал неизвестное значение в точке пересечения с третьей шкалой. Используется, например, для расчета объема выборки при планировании исследования

**Нормальное (гауссово) распределение** — симметричное, колоколообразное распределение, у которого среднее, мода и медиана совпадают. Пологость кривой (т.е. вариация данных) может быть различной. Большинство параметрических тестов разработаны для анализа распределений, подчиняющихся закону нормального распределения

**Нулевая гипотеза** — статистическая гипотеза об отсутствии различий между группами или об отсутствии взаимосвязи признаков. Нулевая гипотеза проверяется с помощью статистических критериев

**Одномерный анализ** — анализ каждого признака независимо от других

**Односторонний тест** — статистический тест, учитывающий априорные знания о направлении (увеличении либо уменьшении) значения исследуемого параметра одной группы по отношению к этому же параметру другой группы. Требуемое для одностороннего теста минимальное различие между группами меньше, чем для двустороннего теста

**Операционные характеристики теста** — диагностические чувствительность, специфичность, эффективность, прогностическая

ценность положительного результата, прогностическая ценность отрицательного результата и др.

**Описательная статистика** — параметры выборки, описывающие набор данных, например среднее, медиана, среднее квадратическое отклонение

**Ослепление (маскирование)** — способ предотвращения осведомленности больных, медицинского персонала, статистиков о том, какие из наблюдений к какой из исследуемых групп относятся (экспериментальной либо контрольной)

**Относительный риск** — отношение риска возникновения неблагоприятного состояния в одной группе к риску этого состояния в другой группе. Используется в проспективных и наблюдательных исследованиях, когда группы формируются заранее, а возникновение исследуемого состояния еще не произошло

**Отношение шансов** — отношение шанса события в одной группе к шансу события в другой группе. Используется в ретроспективных исследованиях случай—контроль в качестве оценки относительного риска, когда группы формируются на основе исхода и целью является определение эффектов факторов риска

**Оценка** — значение параметра распределения на выборке, которое предположительно отражает соответствующее истинное свойство распределения признака в популяции. Точечная оценка — одно число (например, среднее). Интервальная оценка — интервал значений признака (например, доверительный интервал)

**Параметрические методы статистики** — класс статистических методов, используемых для анализа данных, которые образуют известное распределение (обычно нормальное). Названы так потому, что опираются на анализ параметров нормального распределения

**Параметры рассеяния** — статистические параметры распределения, отражающие вариабельность значений признака на выборке (среднеквадратическое отклонение, интерквартильный интервал и др.)

**Парный тест** — тест сопоставления связанных выборок

**Первичный анализ данных** — оценка параметров распределения выборки и проверка гипотез, сформулированных на этапе планирования исследования (до сбора данных)

**Переменная** — см. *Признак*

**Популяция** (в статистическом смысле) — группа субъектов, из которой набрана выборка и на которую следует распространять результаты исследования

**Порядковые данные** — качественные данные, значения которых могут быть ранжированы по какому-либо принципу, но интервал между значениями таких данных не может быть выражен количественно. Обычно качественно отражают условную степень выраженности какого-либо признака

**Признак (переменная)** — характеристика объекта исследования (наблюдения). Различают качественные и количественные признаки

**Проверка (статистической) гипотезы** — математический способ тестирования статистической гипотезы на конкретных данных. Решение принимается путем отклонения или неотклонения нулевой гипотезы об отсутствии различий на основе вероятностей. Неотклонение нулевой гипотезы — это признание существующих различий случайными. Принятие альтернативной гипотезы — это признание наличия воздействия изучаемого фактора (факторов) на результат

**Прогностическая ценность отрицательного результата теста** — вероятность отсутствия заболевания при отрицательном (нормальном) результате теста

**Прогностическая ценность положительного результата теста** — вероятность наличия заболевания при положительном (патологическом) результате теста

**Промежуточный анализ данных** — статистический анализ накапливаемых данных, проводящийся до окончания проспективного исследования. Необходимость промежуточного анализа может быть вызвана правилами остановки исследования. Промежуточный анализ обычно чреват проблемой множественных сравнений

**Промежуточный исход** см. *Косвенный критерий оценки*

**Проспективное исследование** — исследование, в котором группы формируются до сбора данных

**Рандомизация** — способ случайного распределения объектов исследования в основную и контрольную группы с использовани-

ем специальных средств (таблиц или счетчика случайных чисел, подбрасывания монеты и других способов случайного назначения номера группы включаемому наблюдению). С помощью рандомизации сводятся к минимуму различия между группами по известным и неизвестным признакам, потенциально влияющим на изучаемый исход

**Распределение признака на выборке** — набор частот упорядоченных значений признака на выборке. Обычно представляется в виде графика (по оси абсцисс — интервалы значений признака, по оси ординат — частоты упорядоченных значений признака)

**Распространенность** — доля наблюдений в популяции, имеющих определенную характеристику в определенный момент времени

**Ретроспективное исследование** — исследование, в котором группы формируются после сбора данных

**Риск** — вероятность того, что событие или определенный исход наступит, обычно измеряемая в процентах

**Связанные (зависимые) выборки** — выборки, в которые участники исследования набирались парами (или с использованием какого-либо иного принципа) или состоящие из одних и тех же объектов исследования, обследованных в разные моменты времени

**Скользкий экзамен** — метод проверки устойчивости, надежности, работоспособности (валидности) статистической модели путем поочередного удаления наблюдений и пересчета модели. Чем более сходны полученные модели, тем более устойчива, надежна модель

**Событие** — клинический исход, наблюдаемый в исследовании, например возникновение осложнения, рецидива, наступление выздоровления, смерти

**Специфичность диагностического теста** — вероятность отрицательного результата диагностического теста в отсутствие болезни; доля лиц с отрицательным результатом теста среди всех здоровых лиц

**Среднее значение** — описательная статистика (параметр), являющаяся мерой центральной тенденции для приближенно нормально распределенных данных. Если распределение не соответ-



ствуется закону нормального распределения, то для характеристики центральной тенденции следует использовать медиану

**Среднеквадратическое отклонение** — описательная статистика (параметр), являющаяся мерой рассеяния для приближенно нормально распределенных данных. Если распределение не соответствует закону нормального распределения, то для характеристики рассеяния следует использовать какой-либо интерпроцентильный интервал, например интерквартильный интервал

**Стратификация** — метод формирования выборки, при котором совокупность всех участников, соответствующих критериям включения в исследование, сначала разделяется на группы (страты) на основе одной или нескольких характеристик (обычно пола, возраста), потенциально влияющих на изучаемый исход, а затем из каждой из этих групп (страт) независимо проводится набор участников в экспериментальную и контрольную группы. Это позволяет исследователю соблюдать баланс важных характеристик между экспериментальной и контрольной группами

**Суррогатный исход** см. *Промежуточный исход*

**Таблица сопряженности** — таблица абсолютных частот (количества) наблюдений, столбцы которой соответствуют значениям одного признака, а строки — значениям другого признака (в случае двумерной таблицы сопряженности). Значения абсолютных частот располагаются в клетках на пересечении рядов и колонок

**Точность диагностического теста** — доля правильных результатов теста (сумма истинноположительных и истинноотрицательных результатов теста) в общем количестве результатов

**Уровень статистической значимости (критический, пороговый уровень статистической значимости)** — допускаемая исследователем величина  $\alpha$ -ошибки, т.е. максимально допускаемая исследователем вероятность ошибочного отклонения нулевой гипотезы (об отсутствии различия между группами, об отсутствии взаимосвязи признаков и т.д.) статистического теста. Обычно за величину уровня значимости принимаются значения 0,05; 0,01 или 0,001.

**$\alpha$ -Уровень** ( $\alpha$ -ошибка, ошибка первого рода, вероятность гипердиагностики) — пороговый уровень статистической значи-

мости; вероятность ошибочно отклонить нулевую гипотезу. Чем меньше  $\alpha$ -уровень, тем меньше риск совершения этой ошибки. Устанавливается исследователем произвольно (обычно принимается равным 0,05, 0,01 или 0,001)

**$\beta$ -Уровень** ( $\beta$ -ошибка, ошибка второго рода, вероятность гиподиагностики) — вероятность ошибочного принятия (не отклонения) нулевой гипотезы об отсутствии различий. Значение  $\beta$ -уровня в публикациях обычно не приводится. Значение  $\beta$ -уровня обычно устанавливается равным 0,2 или 0,1. “ $1-\beta$ ” — статистическая мощность (см.) теста в выявлении определенной разницы между группами, если она действительно существует. Если выборки малы, то статистическая мощность может быть низкой. Для больших выборок статистические тесты имеют бóльшую статистическую мощность; это означает, что истинные различия между группами выявляются с большей вероятностью

**Характеристическая кривая диагностического теста** — график, характеризующий диагностическую точность теста и строящийся в координатах (чувствительность; единица минус специфичность)

**Цензурированные данные** — один из типов данных, используемых в анализе времени до наступления исхода (таблицы дожития и др.)

**Центральной тенденции параметры** — статистические параметры распределения, отражающие наиболее типичное значение признака на выборке (среднее значение, медиана, мода)

**Чувствительность диагностического теста** — вероятность положительного результата диагностического теста при наличии болезни; доля лиц с положительным результатом теста среди всех больных лиц

**Шанс** — отношение вероятности того, что событие произойдет, к вероятности того, что оно не произойдет



# The Cochrane collaboration

## Кокрановское Сотрудничество

*Российское отделение*

### Что такое Кокрановское Сотрудничество

Кокрановское Сотрудничество (Cochrane Collaboration) – международное сообщество исследователей, поставивших своей целью отыскивать и обобщать результаты всех когда-либо проведенных рандомизированных клинических испытаний лечебных вмешательств. Со дня основания в 1993 году Кокрановское Сотрудничество превратилось во всемирную организацию, имеющую 15 центров в разных странах и объединяющую сотни ученых-медиков по 50 различным специальностям.

Участниками Сотрудничества создана самая совершенная методология аналитических исследований, подготовлено более 2500 систематических обзоров медицинских вмешательств, собрана наиболее полная база данных контролируемых клинических испытаний.

### Почему нужны систематические обзоры?

Кокрановский систематический обзор:

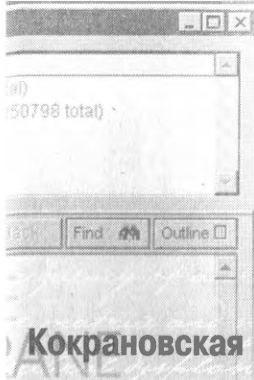
- отвечает на четко сформулированный клинический вопрос
- базируется на результатах поиска всех источников информации на разных языках
- анализирует достоверность данных исследований путем оценки надежности методов сбора и обработки клинической информации
- обобщает только доброкачественные данные
- регулярно обновляется по мере получения новых результатов испытаний

Кокрановский систематический обзор позволяет сделать вывод о том, что:

- вмешательство несомненно эффективно и его необходимо применять
- вмешательство неэффективно и его не следует применять
- вмешательство наносит вред и его следует запретить
- польза или вред не доказаны и требуются дальнейшие исследования

### Кому нужны систематические обзоры?

Кокрановские систематические обзоры незаменимы для практикующих врачей и организаторов здравоохранения, стремящихся основывать свои решения на строго доказанных фактах; исследователей, планирующих проведение клинических испытаний; составителей клинических руководств для подготовки практических рекомендаций для врачей; пациентов, заинтересованных в данных новейших исследований.



## Кокрановская библиотека

Основной продукт деятельности Кокрановского сотрудничества – электронная база данных, называемая Кокрановской библиотекой.

Строго доказанные научные факты необходимы для оказания квалифицированной медицинской помощи, однако найти и обобщить такие факты нелегко.

**БИБЛИОТЕКА КОКРАНОВСКОГО СОТРУДНИЧЕСТВА – лучший источник для получения научного обоснования эффективности лечебных вмешательств.**

Кокрановская библиотека обновляется ежеквартально и распространяется по подписке на компакт-дисках. Кокрановские систематические обзоры – самый надежный источник новейшей, тщательно отобранной и критически обработанной информации.

### **Кокрановская библиотека состоит из шести отдельных баз данных:**

- Кокрановская база данных систематических обзоров
- Кокрановский регистр контролируемых испытаний
- Реферативная база данных обзоров по эффективности медицинских вмешательств
- Кокрановский методологический регистр
- Реферативная база данных по экономической оценке медицинских вмешательств.
- Кокрановская база данных обзоров по методологии.

### **В Кокрановскую электронную библиотеку также включены:**

- учебное пособие по методологии составления систематических обзоров
- база данных рефератов аналитических исследований по оценке эффективности медицинских технологий
- глоссарий методологических терминов
- подробная информация о подразделениях Кокрановского Сотрудничества
- каталог ресурсов Интернета по доказательной медицине

Ознакомиться с Кокрановской библиотекой и научиться ей пользоваться можно бесплатно в Российском отделении Кокрановского Сотрудничества.

**Наш адрес для посещений:** 127238 Москва, Дмитровское ш., д.46 корп. 2, 7 этаж

### **Подписаться можно:**

- Через каталог РОСПЕЧАТИ (возможна полугодовая подписка).  
Индексы: для организаций – **80291**, для частных лиц – **80290**
- Прислав нам заявку по почте: 127238 Москва, а/я 54, «Медиа Сфера» для «Кокрановского Сотрудничества» (укажите количество экземпляров, Ваш адрес, телефоны, электронную почту и пр.)
- Позвонив по телефону: (095) 482-4290
- По факсу: (095) 482-4312
- По электронной почте: [cochr@cochrane.ru](mailto:cochr@cochrane.ru)
- Заглянув на наш сайт в ИНТЕРНЕТЕ \\www.cochrane.ru (автоматизированная подписка)

# Статистический анализ медицинских данных

О.Ю. Реброва

