МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

О.Г. Берестнева, О.В. Марухина, Г.Е. Шевелёв

ПРИКЛАДНАЯ МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Рекомендовано в качестве учебного пособия
Редакционно-издательским советом
Томского политехнического университета

Издательство
Томского политехнического университета
2012

УДК 519.22(075.8)
ББК 22.172я73
Б48

Берестнева О.Г.,

Б48 Прикладная математическая статистика: учебное пособие / О.Г. Берестнева, О.В. Марухина, Г.Е. Шевелёв. – Томск: Изд-во Томского политехнического университета, 2012. – 188 с.

В пособии в общедоступной форме изложены основные математико-статистические методы, применяемые в прикладных гуманитарных и технических исследованиях. Изложение материала сопровождается примерами статистической обработки результатов экспериментальных исследований.

Предназначено для бакалавров направления 230700 «Прикладная информатика», также может быть полезно для аспирантов, научных сотрудников и преподавателей вузов.

УДК 519.22(075.8)
ББК 22.172я73

Рецензенты

Доктор технических наук, профессор СибГМУ
В.А. Фокин

Доктор биологических наук, профессор ТГПУ
А.М. Уразаев

ПРЕДИСЛОВИЕ

Методы прикладной математической статистики активно применяются в технических исследованиях, экономике, теории и практике управления (менеджмента), социологии, медицине, геологии, истории и т. д. С результатами наблюдений, измерений, испытаний, опытов, с их анализом имеют дело специалисты во всех отраслях практической деятельности, почти во всех областях теоретических исследований.

В настоящее время статистическая обработка данных проводится, как правило, с помощью соответствующих программных продуктов. Разрыв между математической и прикладной статистикой проявляется, в частности, в том, что большинство методов, включенных в статистические пакеты программ (например, в заслуженные *STATGRAPHICS* и *SPSS* или в более новую систему *STATISTICA*), даже не упоминается в учебниках по математической статистике. В результате специалист по математической статистике оказывается зачастую беспомощным при обработке реальных данных, а пакеты программ применяют (что еще хуже – и разрабатывают) лица, не имеющие необходимой теоретической подготовки.

В главе 1 рассмотрены приемы работы с универсальными статистическими пакетами, такими как *STATGRAPHICS*, *SPSS*, *STATISTICA*. Эти пакеты выделяются развитым аппаратом статистического анализа, удобными графическими средствами, высокой точностью вычислений, содержат удобный редактор отчетов. Наличие недостатков и, следовательно, привлекательность того или иного пакета пользователь может определить при непосредственной работе. Достаточно большой объем информации представлен в графическом виде, дается интерпретация основных понятий на уровне графических образов с последующим сравнительным анализом.

Глава 2 содержит проверку статистических гипотез о законах распределения на доступных примерах с привлечением формул разных авторов. Для ряда критериев даны примеры реакции статистических пакетов.

Глава 3 содержит материал по исследованию взаимосвязи между изучаемыми признаками разных групп. Обсуждается выбор меры связи между переменными, измеренными в разных шкалах.

В главе 4 рассматривается оценка влияния известных факторов на исследуемую переменную (дисперсионный анализ), а в главе 5 – выделение новых латентных переменных (факторный анализ). При ознакомлении с материалом главы 5 находится много объяснений методов, полезных для пользователя ПК. Для кластерного анализа приведен анализ выбора алгоритма кластерного анализа, применимого в каждом случае, что позволяет выбрать приемлемый вариант с наилучшими результатами.

Прикладная математическая статистика нацелена на решение реальных задач. Поэтому в ней возникают новые постановки математических задач анализа статистических данных, развиваются и обосновываются новые методы. Обоснование часто проводится математическими методами, то есть путем доказательства теорем. Большую роль играет методологическая составляющая – как именно ставить задачи, какие предположения принять с целью дальнейшего математического изучения. Велика роль современных информационных технологий, в частности, компьютерного эксперимента.

ГЛАВА 1
КОМПЬЮТЕРНЫЕ ПРОГРАММЫ АНАЛИЗА ДАННЫХ. ОСНОВНЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ И ИХ ИСПОЛЬЗОВАНИЕ В ПРИКЛАДНЫХ ИССЛЕДОВАНИЯХ

1.1. Компьютерные программы анализа данных

Рынок компьютерных программ анализа данных очень разнообразен. На нем представлены продукты более тысячи наименований. Такое разнообразие отражает многоплановость задач анализа экспериментальных данных в различных областях человеческой деятельности. Выбор исследователем компьютерного статистического пакета анализа данных зависит от характера решаемых задач, квалификации, имеющегося оборудования и т. д. В табл. 1.1 представлена классификация статистических пакетов.

Таблица 1.1

Классификация статистических пакетов

Тип	Зарубежные	Отечественные
Профессиональные	SAS	нет
Универсальные	STATGRAPHICS, SPSS, STATISTICA, S-PLUS	STADIA, Olymp
Специализированные	Большое разнообразие	Класс-Мастер, Эвристика, DataScope, САНИ

В данном пособии будут рассмотрены приемы работы с универсальными статистическими пакетами, такими как *STATGRAPHICS*, *SPSS*, *STATISTICA*.

Следует отметить тот факт, что пакеты *SPSS* и *STATISTICA* предназначены для пользователей, владеющих статистическими методами анализа данных на профессиональном уровне. Поэтому начинать осваивать работу с универсальными статистическими пакетами рекомендуется с *STATGRAPHICS*. Этот пакет обладает статистическим консультантом (StatAdvisor), который поможет интерпретировать полученные результаты, определит значимые эффекты или укажет на возможные недостатки в проведенном анализе. Изучив принципы работы с данным пакетом, не составит особого труда перенести полученные навыки на другие пакеты.

Из множества привлекательных свойств рассматриваемых статистических пакетов выделим и охарактеризуем следующие:

1. Наличие достаточно широкого спектра статистических алгоритмов (порядка 100).
2. Сотни типов двумерных и трехмерных графиков.
3. Обмен данными с другими программными продуктами.
4. Большой набор возможностей манипулирования данными (сортировка, трансформация, кодировка, изменение шкалы измерения).
5. Комбинирование текста и графики для составления статистических отчетов.
6. Коррекция и преобразование элементов графических отображений (изменение цвета, заливки, шрифта, надписей, меток, масштабов и т. д.).
7. Взаимодействие пользователя с данными посредством графики (идентификация объекта, разгонка точек на диаграммах рассеивания, окраска «интересных» объектов).

Естественно, что приведенные достоинства далеко не полностью отображают все возможности анализа данных, которыми располагают компьютерные статистические пакеты, но уже достаточно информации, для того чтобы стало ясно, что работа с ними эффективна и удобна.

Наличие недостатков и, следовательно, привлекательность того или иного пакета пользователь может определить при непосредственной работе. Поэтому выбор в пользу предпочтения какого-либо программного продукта предоставляется самому исследователю, а нам лишь остается познакомить его с основными принципами работы со статистическими программами.

1.1.1. STATGRAPHICS

После инициализации системы *STATGRAPHICS* открывается рабочее окно, представленное на рис. 1.1.

Внизу экрана расположен набор пиктограмм, которые позволяют активизировать следующие окна:

– пиктограмма открывает окно работы с электронной таблицей. Работа с данной таблицей ничем не отличается от работы с другими известными электронными таблицами для *Windows* типа *Exel*. Для переименования переменных и задания их типа необходимо выделить столбец и вызвать контекстное меню (правая кнопка мыши), а затем выбрать пункт меню *Modify Column*. Преобразование переменных (например, арифметическое, логическое), а также генерацию новых признаков можно осуществить при помощи пункта – *Generate Data*.

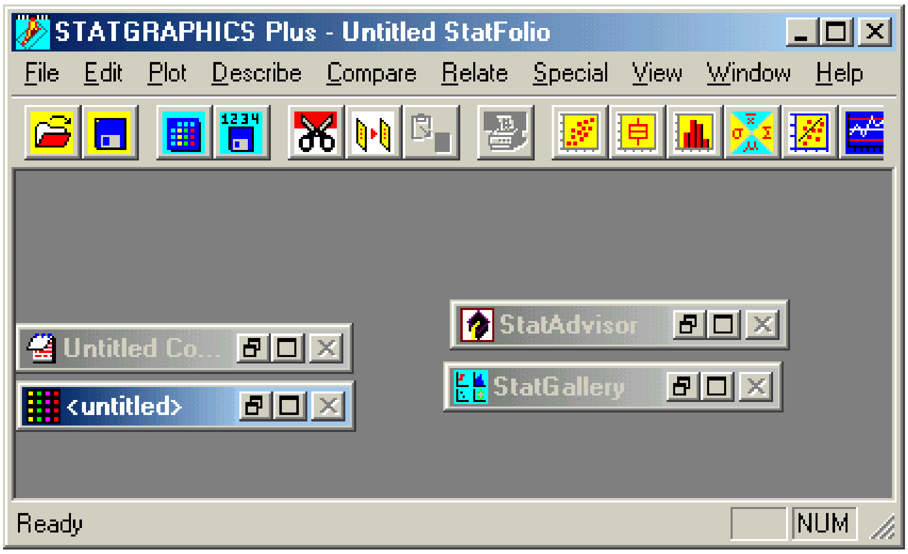


Рис. 1.1. Рабочее окно системы STATGRAPHICS

– пиктограмма открывает окно интеллектуального помощника, который поможет интерпретировать полученные результаты, определить значимые эффекты или укажет на возможные недостатки в проведенном анализе.

– пиктограмма активизирует окно составления отчета. В галерее отчетов возможно располагать в одном окне или на одном листе до девяти различных фрагментов текста и графических иллюстраций (рис. 1.2–1.3). При этом трансформация и перемещение всех составляющих отчета производится быстро и просто.

– пиктограмма раскрывает окно для ввода комментариев к проводимому статистическому анализу. Сохранив протокол анализа (*File – Save StatFolio As…*), исследователь сможет повторить проведенный анализ на новом массиве данных. Все заданные таблицы и графические отображения будут выданы автоматически.

Универсальный статистический пакет *STATGRAPHICS* включает более 100 статистических и системных процедур, применяющихся в медицине, социологии, психологии, педагогике, в бизнесе, на производстве и других областях. Каждой группе процедур соответствует отдельный пункт меню, которое расположено в верхней части рабочего окна. Опишем несколько основных пунктов меню.

Пункт *Plot.* Раскрывающийся список *Plot* представленный на рис. 1.4 содержит в себе графические процедуры визуализации экспериментальных данных.

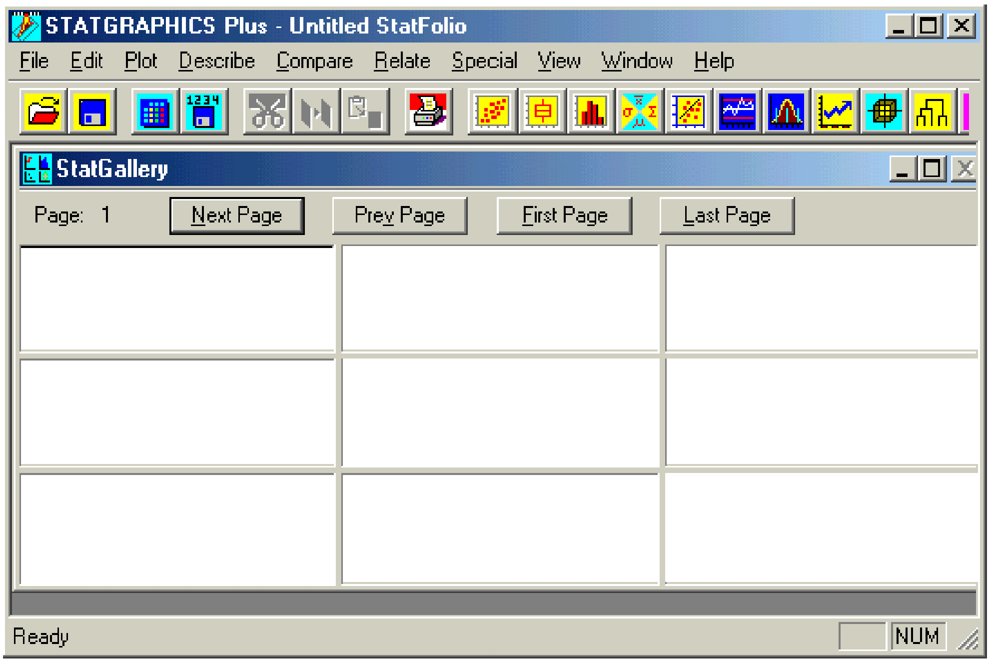


Рис. 1.2. Окно составления отчетов

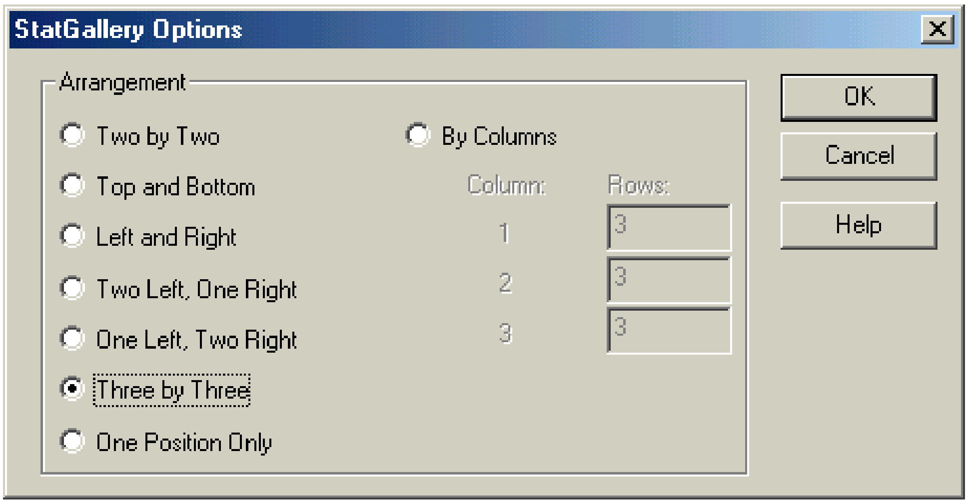


Рис. 1.3. Панель выбора количества фрагментов в окне отчета

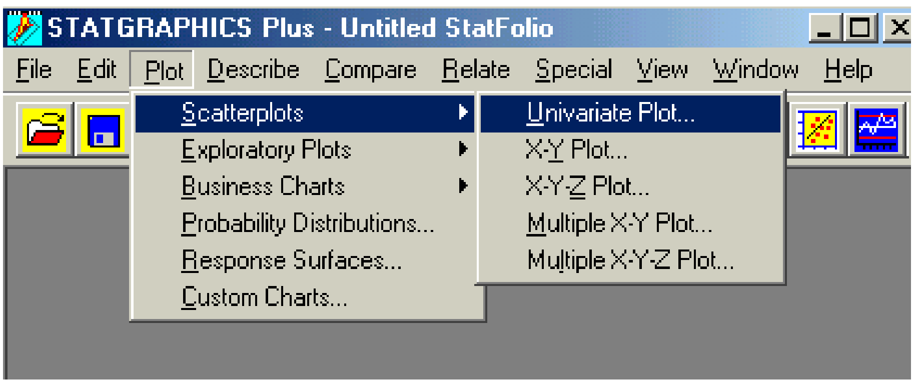


Рис. 1.4. Пункт Plot – графическое отображение данных

Пункт *Describe* содержит статистические методы анализа данных по одной и множеству переменных, процедуры подбора распределений, средства табуляции, кросс-табуляции данных (рис. 1.5).

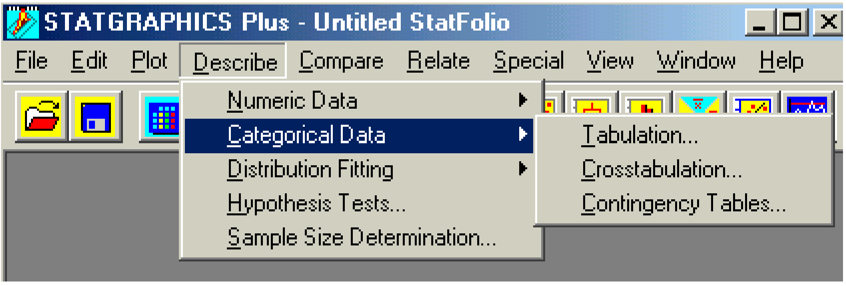


Рис. 1.5. Пункт Describe – статистические методы анализа

Пункт *Compare* включает методы сравнения двух и более выборок данных, процедуры одно- и многофакторного дисперсионного анализа и др. (рис. 1.6).

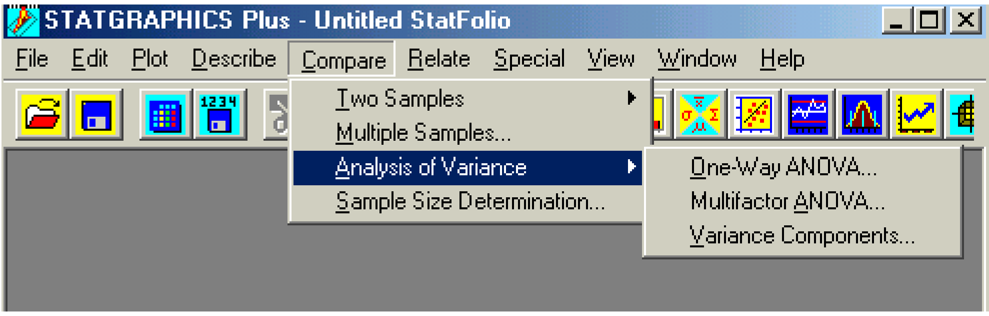


Рис. 1.6. Пункт Compare

Пункт *Relate* содержит процедуры простого, полимиального и множественного регрессионного анализа (рис. 1.7).

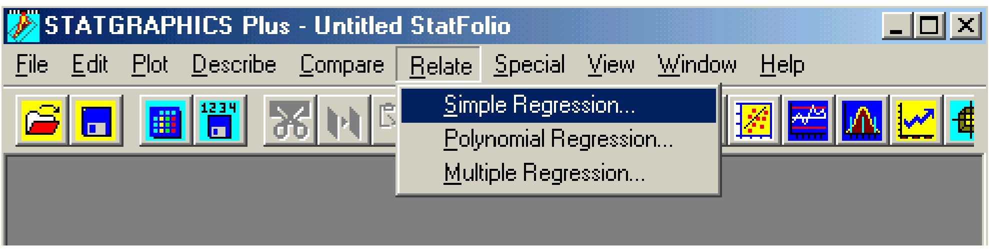


Рис. 1.7. Пункт Relate

Пункт *Special*. В данном пункте меню предлагаются дополнительные модули анализа данных: контроль качества, планирование эксперимента, анализ временных рядов, многомерные методы анализа и расширенный регрессионный анализ (рис. 1.8).

В дополнительном модуле *Многомерные методы* собраны процедуры, обеспечивающие проведение анализа по методу главных компонент, факторного и кластерного анализа, дискриминантного и канонического корреляционного анализа (рис. 1.8).

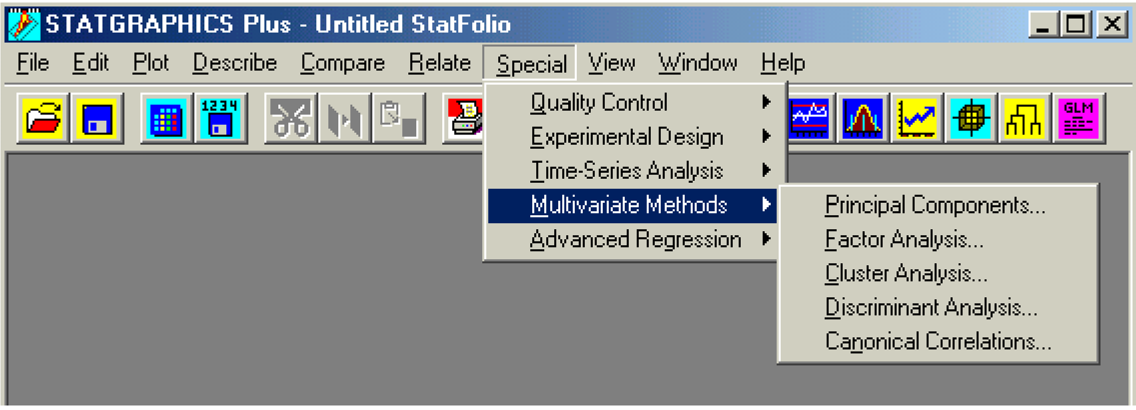
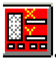
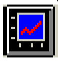
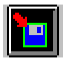


Рис. 1.8. Пункт Special – набор дополнительных модулей анализа данных

Несколько слов необходимо добавить о возможности вызова диалоговых окон задания параметров реализации выбранного статистического метода и построения графических иллюстраций.

После того, как будет выбран рабочий файл с исходными данными и активизирован один из методов анализа данных, на экране откроется окно с результатами сводного анализа. В верхней части окна пользователю становятся доступными следующие пиктограммы (рис. 1.9): окно ввода исходных данных , табличные опции , графические опции  и пиктограмма  сохранения полученных результатов. В зависимости от выбранного метода анализа на панели также будут активизироваться различные пиктограммы для проведения интерактивного анализа экспериментальных данных.

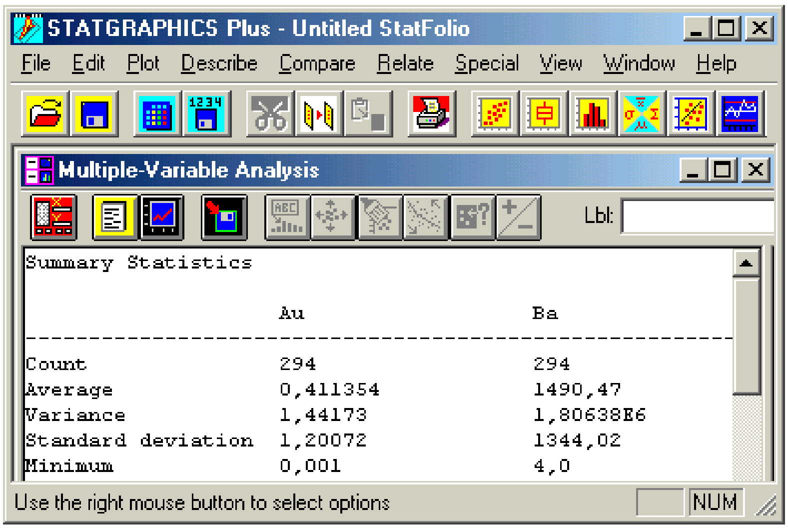


Рис. 1.9. Окно вывода основных характеристик переменных

1.1.2. SPSS

После загрузки программного продукта *SPSS* на экране появляется рабочее окно (рис. 1.10) и диалоговое окно ввода имени рабочего файла с расширением *.*sav*. Вы также можете расширить область рабочих файлов, открыв раскрывающийся список с типами файлов, которые поддерживает данный статистический пакет.

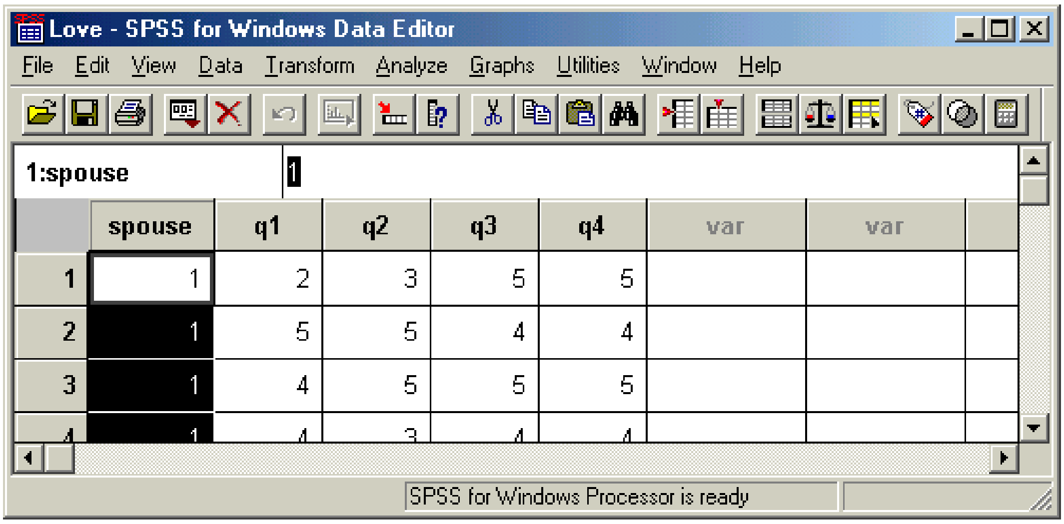


Рис. 1.10. Рабочее окно пакета SPSS

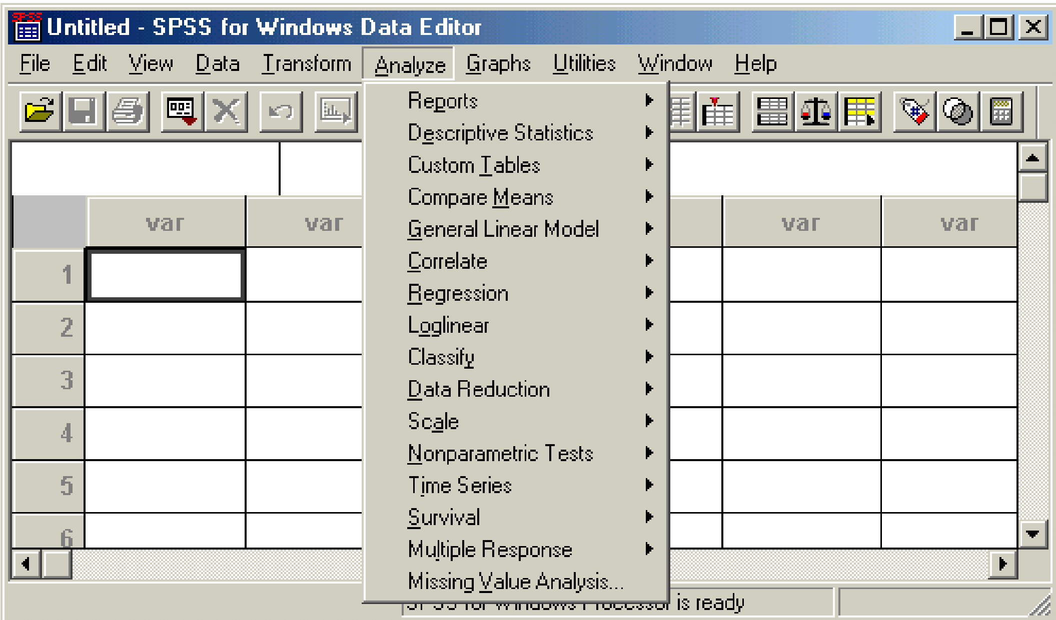


Рис. 1.11. Меню Analyze

В базовой системе функционируют следующие процедуры:

Пункт *Analyze* содержит статистические методы анализа (рис. 1.11.): описательные статистики, критерии сравнения средних, проверка взаимосвязей, построение регрессионных моделей, шкалирование, непараметрические тесты и другие.

Пункт меню *Graphs* содержит графические процедуры (рис. 1.12): гистограммы, графики, полигоны, ящики с усами, графики построения на вероятностной бумаге (*P-P…)* и многие другие.

Пункт *Transform* (рис. 1.13) включает методы трансформации переменных: кодирование, автоматическое кодирование, ранжирование, категоризацию переменных и т. д.

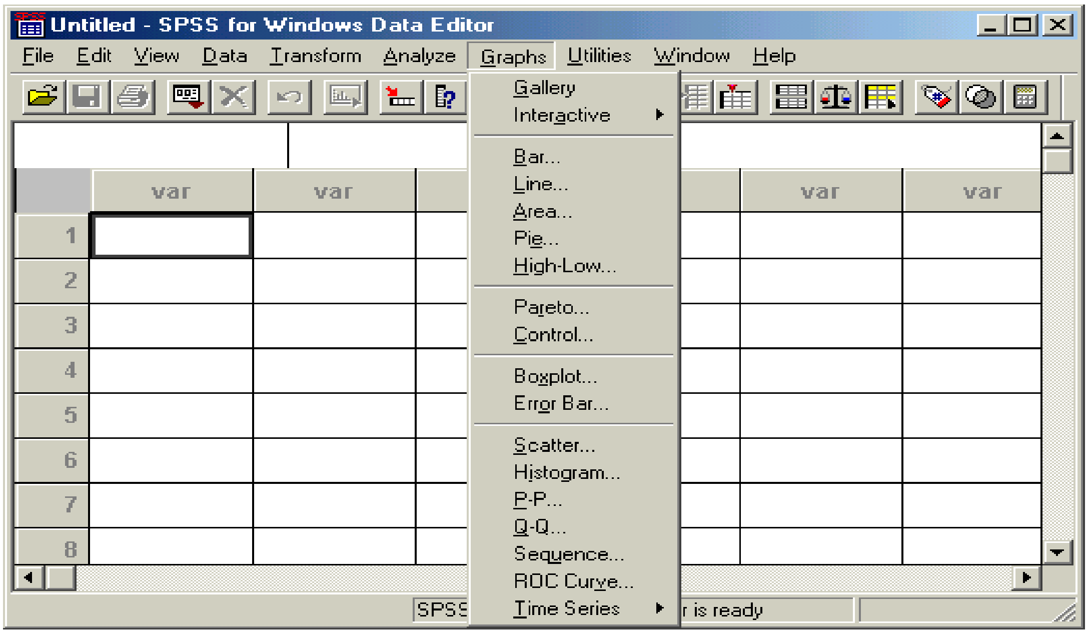


Рис. 1.12. Пункт Graphs

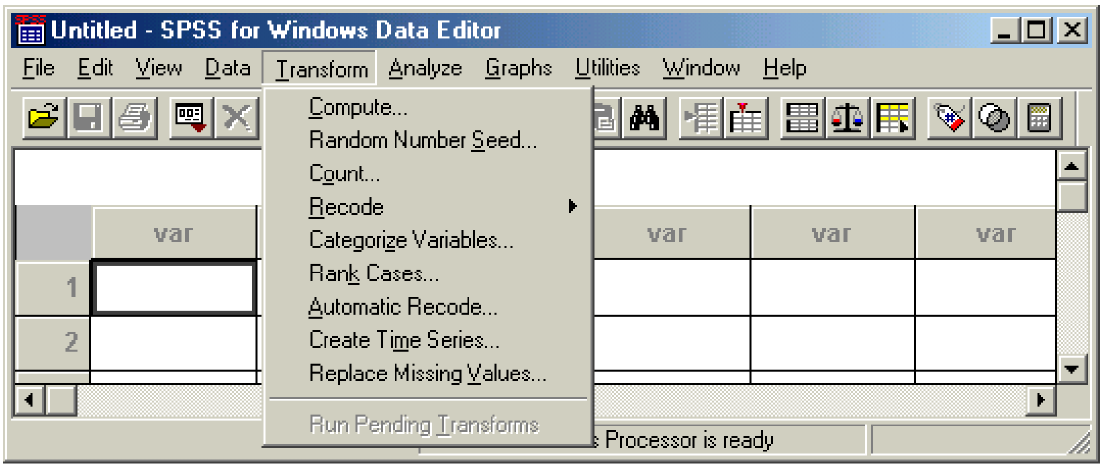


Рис. 1.13. Пункт Transform

Пункт *Data* (рис. 1.14) содержит процедуры сортировки, разбиения исходного файла по заданным параметрам, определения и вставки переменных и др.

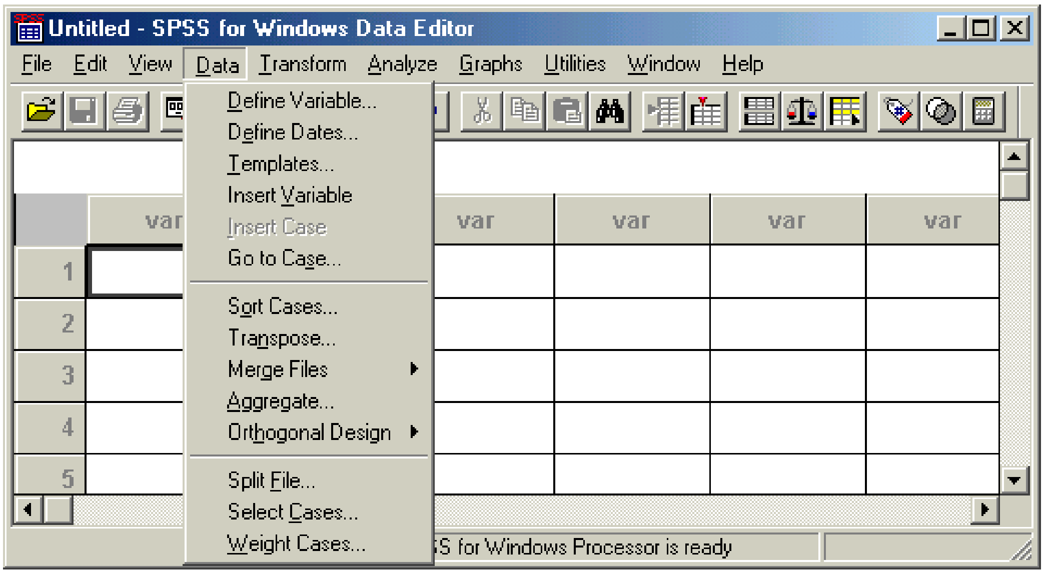


Рис. 1.14. Пункт Data

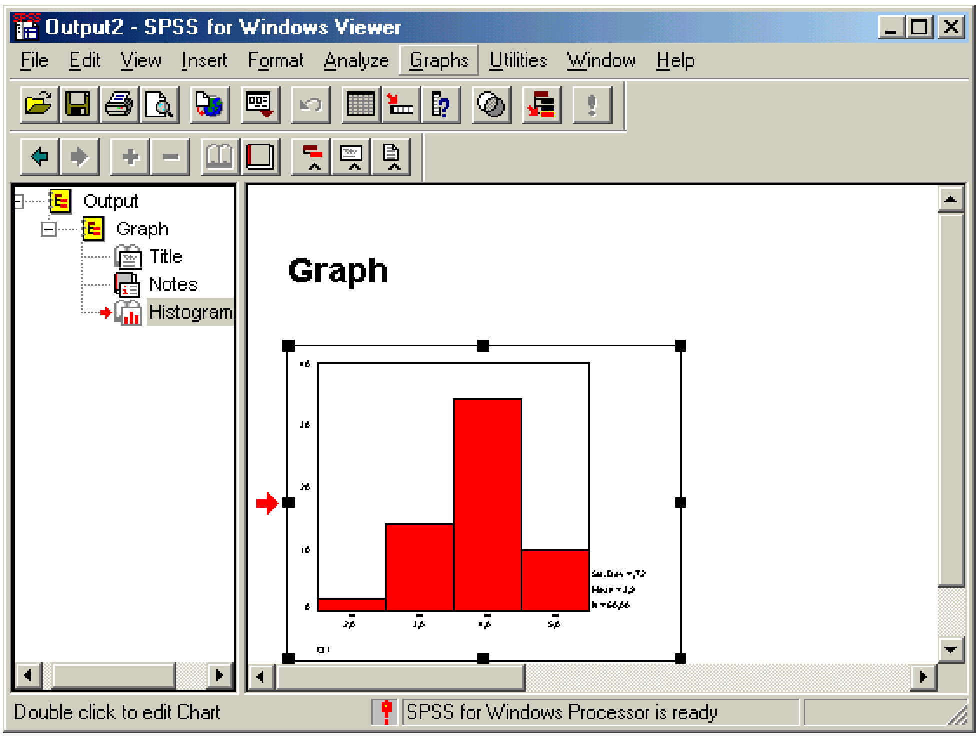


Рис. 1.15. Окно навигатора редактора отчета

Программа *SPSS* в настоящее время считается одним из лидеров среди универсальных статистических пакетов, но тем не менее следует еще раз напомнить, что она предназначена для статистиков-профессионалов и без детальной справочной документации для пользователей, не владеющих английским языком, будет очень трудно освоить самостоятельно работу с данным пакетом. Далее, по ходу изложения теоретического материала, в пособии будут приводиться примеры работы с *SPSS*.

Результаты проведенного анализа отображаются в навигаторе вывода *SPSS*. Большинство процедур представляют результаты в виде мобильных (имеется возможность непосредственного редактирования) таблиц (рис. 1.15).

1.1.3. STATISTICA

Загрузив программную оболочку пакета *STATISTICA* на экране, появляется рабочее окно (рис. 1.16):

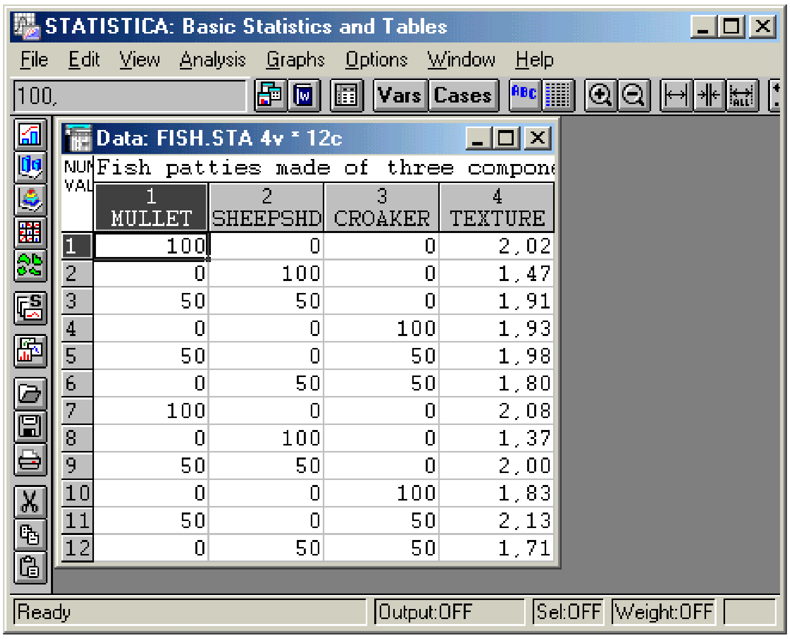


Рис. 1.16. Рабочее окно пакета STATISTICA

Главное отличие рассматриваемого пакета от двух других в том, что статистические процедуры системы сгруппированы в нескольких специализированных статистических модулях (рис. 1.17):

- основные статистики и таблицы;
- множественная регрессия;
- дисперсионный анализ (*ANOVA/MANOVA*);
- дискриминантный анализ;
- непараметрическая статистика и подгонка распределений;
- факторный анализ;
- многомерное шкалирование;
- анализ надежности;
- кластерный анализ;
- лог-линейный анализ;
- каноническая корреляция;
- анализ временных рядов и прогнозирование;
- нелинейное оценивание и т. д.

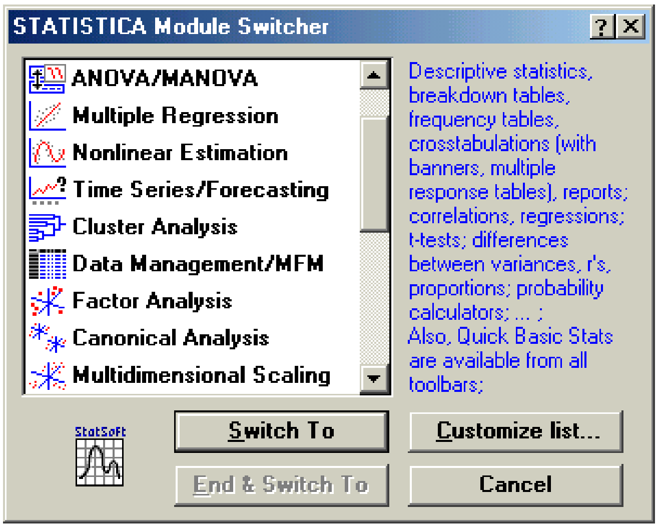


Рис. 1.17. Переключатель статистических модулей

Ниже дается краткое описание статистических процедур, доступных в отдельных статистических модулях.

Модуль *Основные статистики и таблицы* – предварительная обработка данных, разведочный анализ, определение и визуальное оценивание зависимостей между экспериментальными данными, разбиение на группы и т. д. Обычно с этого модуля начинается работа в системе (рис. 1.18).

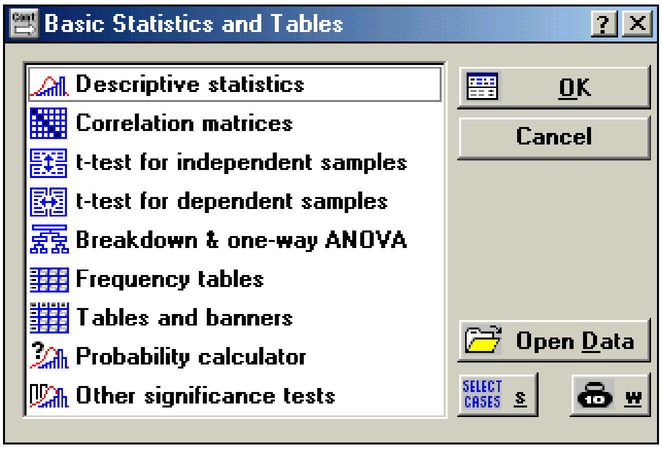


Рис. 1.18. Окно выбора в модуле Основные статистики таблицы

Модуль *Множественная регрессия* – построение зависимости между многомерными переменными, подбор линейной модели и оценка ее адекватности. Модуль *Дисперсионный анализ* – оценка степени воздействия фактора на измеряемые данные.

В модуле *Дискриминантный анализ* решается задача, как по результатам измерений отнести объект к одному из нескольких классов, например в медицине при обследовании больных, или в педагогике при оценке успешности метода обучения.

Модуль *Непараметрическая статистика* и подгонка распределений – проверка различных гипотез о характере распределения данных (рис. 1.19). Модуль содержит непараметрические критерия согласия, в частности, критерий Колмогорова–Смирнова, ранговые критерии Манна–Уитни, Вальда–Вольфовица, Уилкоксона и многие другие.

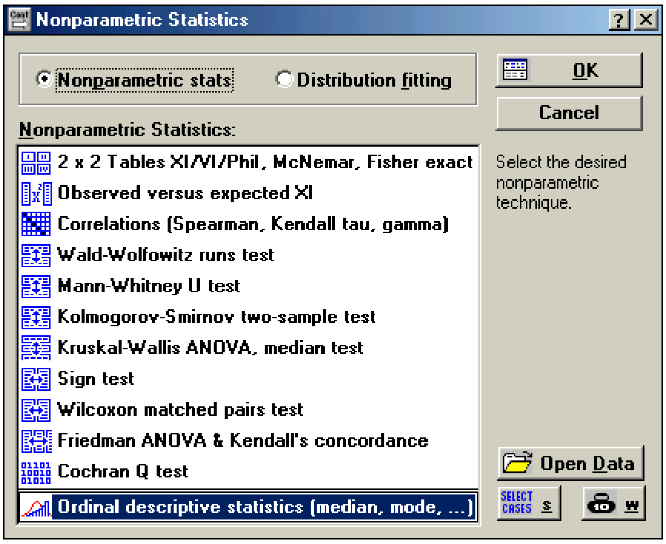


Рис. 1.19. Модуль непараметрической статистики и подгонки распределения

Модуль *Факторный анализ* – сжатие данных или выделение общих факторов, влияющих на наблюдаемые характеристики сложного объекта и объясняющие связи между ними, например основные социально-психологические факторы или факторы, влияющие на результаты голосования.

Модуль *Многомерное шкалирование* – представление о близости объектов какой-либо пространственной моделью, в которой объекты интерпретируются, например, как города на обычной карте, а различия между ними есть просто расстояния.

Модуль *Кластерный анализ* – сложная иерархическая классификация данных, кластеризация данных (рис. 1.20). В дополнение к стандартному выводу кластерного анализа предлагается исчерпывающее множество описательных статистик.

Модуль *Нелинейное оценивание* – определение нелинейных зависимостей в данных, подгон к ним функциональных кривых (рис. 1.21).

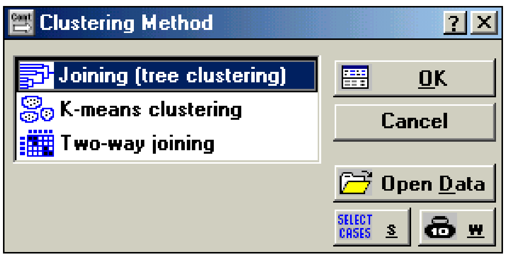


Рис. 1.20. Окно выбора в модуле кластерного анализа

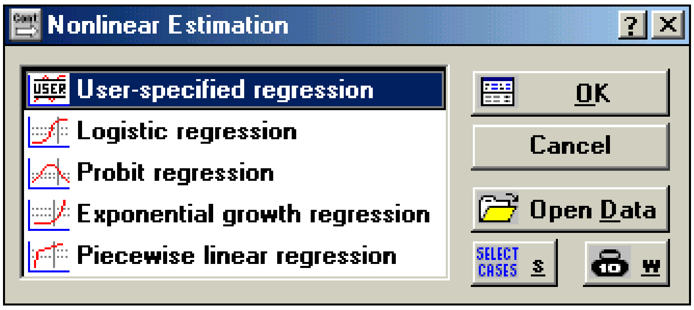


Рис. 1.21. Окно выбора в модуле нелинейного оценивания

Для визуализации значений в пакете *STATISTICA* доступны различные типы графиков, включая научные, деловые, специализированные статистические графики как на плоскости, так и в пространстве.

Статистические графики могут быть вызваны следующими способами: через меню *Graphs*, при помощи пиктограммы на панели инструментов или при помощи контекстного меню, вызываемого правой кнопкой мыши. Галерея графиков представлена на рис. 1.22.

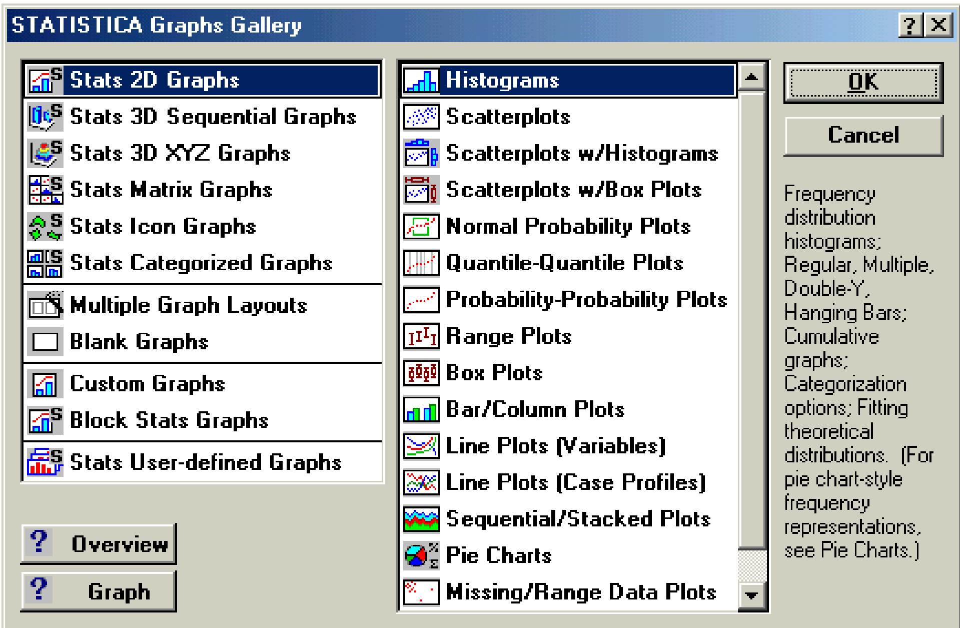


Рис. 1.22. Графическая галерея системы STATISTICA

Результаты обработки исходных экспериментальных данных при помощи той или иной статистической процедуры выводятся в виде специальных электронных таблиц, которые называются таблицами вывода результатов (*Scrollsheet*) (рис. 1.23).

Для того чтобы все результаты анализа, которые появляются на экране, автоматически заносились в файл регистрации необходимо создать автоотчет. Для этого необходимо из меню *File*, в диалоговом окне *Page/Output Setup* выбрать опцию *Automatically Print All Scrollsheet* или с помощью двойного клика мыши на поле *Output* в строке состояния внизу окна *STATISTICA* (рис. 1.24).

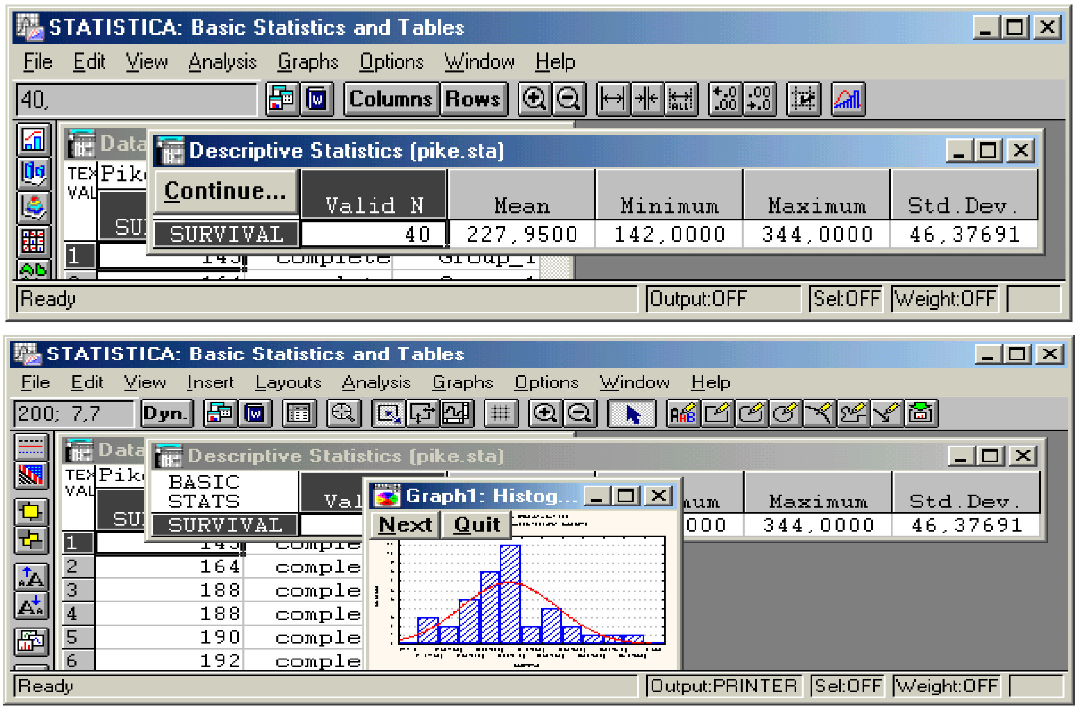


Рис. 1.23. Таблицы вывода результатов статистического анализа: с описательными статистиками (вверху), с графиком (внизу)

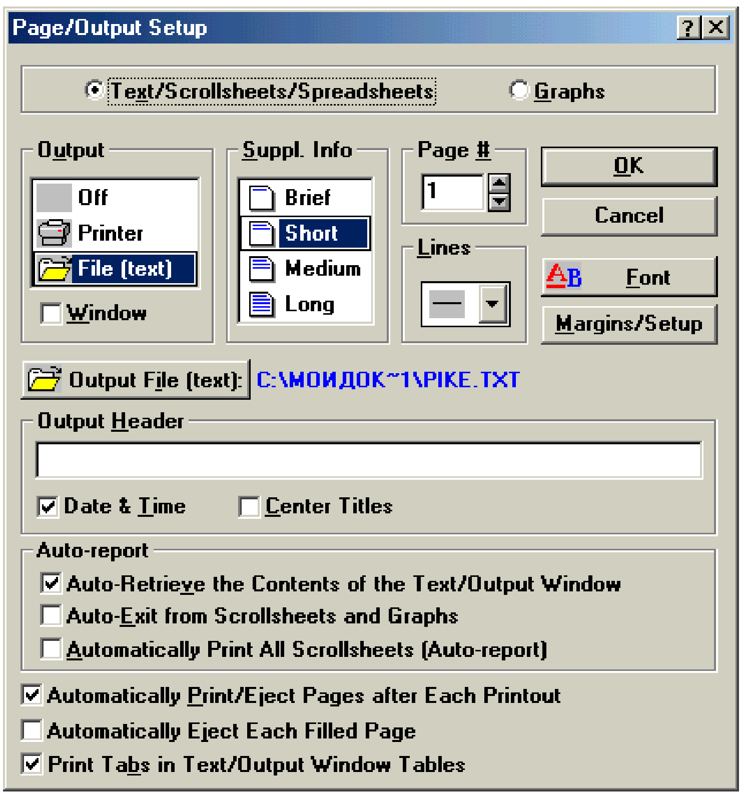


Рис. 1.24. Диалоговое окно задания параметров канала вывода

По мере изложения теоретического материала в учебном пособии будут представлены практические примеры работы со статистическими пакетами анализа данных.

Для этих целей будем использовать Интернет-ресурс http://www.psychol-ok.ru/.

С главной страницы необходимо перейти по ссылке «*Библиотека*», найти раздел «*Психодиагностические методики и математические методы обработки данных*» и выбрать интересующий пункт (рис. 1.25).

Психодиагностические методики и математические методы обработки данных

- Автоматический расчет T - критерия Вилкоксона
- Автоматический расчет t-критерия Стьюдента
- Автоматический расчет U-критерия Манна-Уитни
- Автоматический расчет теста СМИЛ (MMPI) (он-лайн тест)
- Автоматический расчет теста Сонди (он-лайн тест)
- Автоматический расчет углового преобразования Фишера
- Анализ семейного воспитания (он-лайн тест)
- Клинический опросник для выявления и оценки невротических состояний (он-лайн тест)
- Коэффициент ранговой корреляции Спирмена
- Тест-опросник удовлетворенности браком (он-лайн тест)
- Характер взаимодействия супругов в конфликтных ситуациях (он-лайн тест)
- χ^2 - критерий Пирсона

Рис. 1.25. Критерии и тесты для автоматического расчета

После выбора критерия предоставляется краткое его описание. В конце страницы находятся поля для ввода данных, необходимых при расчете.

1.2. Основные понятия математической статистики

В прикладных экспериментальных исследованиях мы чаще всего имеем дело не с одним, а с группой объектов (испытуемых, обследуемых), т. е. изучаются массовые явления, в сфере которых проявляют свое действие статистические законы. Например, если психолог провел психодиагностику одного человека – это единичное явление. Если же психолог провел психологическое тестирование по одной и той же методике нескольких человек или проводит длительное наблюдение за одним и тем же человеком – это массовое явление независимо от того, каким был объект наблюдения – единичным или групповым.

Генеральной совокупностью называют множество относительно однородных, но индивидуально различимых объектов (наблюдений, измерений, описаний), объединенных для совместного (группового) изучения, а отобранная тем или иным способом часть генеральной совокупности называют *выборкой*. Для получения достоверных результатов анализа данных необходимо, чтобы выборка была *представительной* (репрезентативной).

Количество элементов в выборке называется *объемом выборки* и обозначают буквой n.

В качестве элементов выборки используются *переменные* или *признаки*, которые являются характеристиками объектов исследования, например вес, возраст, время реакции, беглость чтения, количество детей, число студентов. Некоторые из этих переменных *непрерывны* (т. е. измерения их могут дать любое значение внутри некоторой области), таковы вес, возраст, уровень тревожности, самочувствие и т. п., а некоторые – *дискретны* (т. е. их измерения могут давать только отдельные целые значения), например количество детей.

Измерение любой непрерывной переменной должно сопровождаться определением *точности* процесса измерения. Например, скорости хронометрируются с точностью до десятой доли секунды; рост может быть измерен с точностью до сантиметра; возраст – с точностью до дня.

1.3. Табулирование данных

Результат любого эксперимента фиксируют в той или иной форме (например, табл. 1.2), а затем используют для той цели, ради которой и проводился эксперимент.

Переменные принято обозначать последними заглавными буквами (x, y, z), а их числовые значения – соответствующими строчными буквами с индексом ($x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_n$ или $z_1, z_2, \ldots, z_n$).

Таблица 1.2

Таблица экспериментальных данных

Номер объекта (испытуемого)	Номер переменной (признака)					
	x_1	x_2	…	x_j	…	x_p
1	x_{11}	x_{12}	…	x_{1j}	…	x_{1p}
2	x_{21}	x_{22}	…	x_{2j}	…	x_{2p}
.	.	.	…	.	…	.
.	.	.	…	.	…	.
i	x_{i1}	x_{i2}	…	x_{ij}	…	x_{ip}
.	.	.	…	.	…	.
.	.	.	…	.	…	.
n	x_{n1}	x_{n2}	…	x_{nj}	…	x_{np}

В приведенной таблице используются следующие обозначения:

n – общее количество объектов (испытуемых);

p – общее количество признаков;

x_j – j-й признак;

x_{ij} – значение j-го признака, измеренного у i-го объекта.

Так, например, если мы измерили значения коэффициента интеллекта в группе из 20 испытуемых, то $n = 20$. Переменную «коэффициент интеллекта» мы можем обозначить как x_1. Тогда x_{11} будет обозначать значение коэффициента интеллекта у первого испытуемого; x_{21} – у второго; x_{ij} – у i-го испытуемого и x_{n1} – значение коэффициента интеллекта у n-го испытуемого.

1.3.1. Ранговый порядок

Первый этап представления данных – это обычно упорядочивание результатов, например, от максимального до минимального значения. Такое представление называется *несгруппированным рядом*. В небольшой группе этого часто вполне достаточно.

Для того чтобы построить несгруппированный ряд воспользуемся пакетом STATGRAPHICS. Выберите в меню пункт *Describe – Categorical Data – Tabulation* (рис. 1.25).

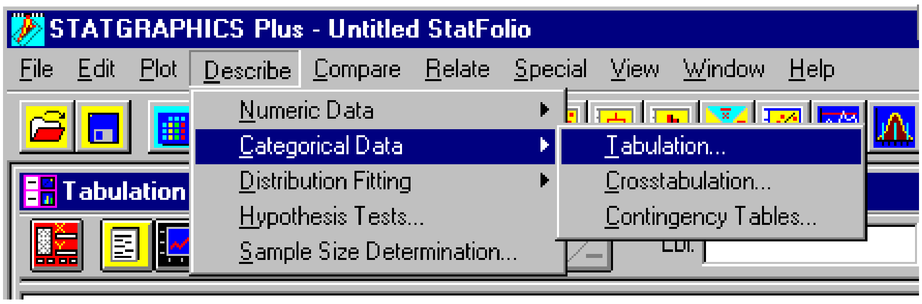
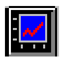


Рис. 1.25. Выбор процедуры Табуляции в пакете STATGRAPHIS

После выбора исследуемого признака и нажатия кнопки *OK*, на экране отображается окно со следующими данными: указывается общее количество элементов выборки, а также количество неповторяющихся значений. Чтобы продолжить анализ исследуемой выборки, можно воспользоваться пиктограммами задания табличных  и графических  опций. На рис. 1.26 представлена упорядоченная частотная таблица, начиная с наименьшего значения. В представленной таблице отображаются значения частоты повторения (*Frequency)*, относительной частоты (*Relative Frequency*), накопленной частоты (*Comulative Frequency*) и накопленной относительной частоты (*Comulative Relative Frequency*) для каждого неповторяющегося значения.

На следующем этапе представления данных, возможно, понадобиться выполнить ранжирование упорядоченных данных, например в критерии Манна–Уитни (см. гл. 3). Ниже приведены *правила ранжирования.*

1. Наименьшему значению начисляется ранг 1.
2. Наибольшему значению начисляется ранг, соответствующий общему количеству ранжируемых значений (объему выборки n). Для примера представленного на рис. 1.26 наибольший ранг равен 20.

```
Frequency Table for Time
```

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	58	1	0,0500	1	0,0500
2	59	2	0,1000	3	0,1500
3	69	1	0,0500	4	0,2000
4	70	1	0,0500	5	0,2500
5	71	1	0,0500	6	0,3000
6	75	3	0,1500	9	0,4500
7	83	1	0,0500	10	0,5000
8	84	2	0,1000	12	0,6000
9	93	1	0,0500	13	0,6500
10	95	2	0,1000	15	0,7500
11	97	2	0,1000	17	0,8500
12	100	1	0,0500	18	0,9000
13	109	1	0,0500	19	0,9500
14	112	1	0,0500	20	1,0000

Рис. 1.26. Результат упорядочивания значений исследуемой выборки

3. В случае если несколько значений равны, им начисляется ранг, представляющий собой среднее значение из тех рангов, которые они получили бы, если не были равны. Например (см. рис. 1.26), так как существует 2 объекта, ранг которых выше 1, то следующие два ранга, 2 и 3, усреднены, что дает значению 59 ранг равный 2,5. Имеется три результата равных 75, тогда для каждой оценки 75 ранг равен 8 (среднее значение следующих трех рангов 7, 8, 9).
4. Общая сумма рангов должна совпадать с расчетной, которая определяется по формуле

$$\sum\left(R_i\right) = \frac{n(n+1)}{2}, \tag{1.1}$$

где n – общее количество ранжируемых наблюдений (значений).

1.3.2. Распределение частот

Полученный список на рис. 1.26 можно сократить, классифицируя оценки по распределению частот (f), иногда называемому просто распределением.

Для большого числа оценок – скажем, 100 или более – на следующем этапе может иметь смысл обобщения данных. Как правило, существует настолько широкий диапазон оценок тестирования, что целесообразнее сгруппировать их по величинам, например в группы, объединяющие все результаты от 58 до 70 включительно, от 71 до 83 включительно и т. д., каждая такая группа называется разрядом оценок.

В случае полного размещения по группам обычно говорят о распределении сгруппированных частот.

Осуществим процедуру группирования, используя возможности пакета *STATGRAPHICS.* Для этого необходимо выбрать в меню пункт *Describe – Numeric Data – One-Variable Analysis…* Далее активизруйте диалоговое окно для выбора табличных опций с помощью пиктограммы  и выберите процедуру *Frequency Tabulation* (рис. 1.27).

На экране отобразится окно с результатом распределения частот. Для того чтобы изменить параметры распределения (задать ширину разряда, нижнюю и верхнюю границу) в контекстном меню выберите пункт *Pane Option* (рис. 1.28).

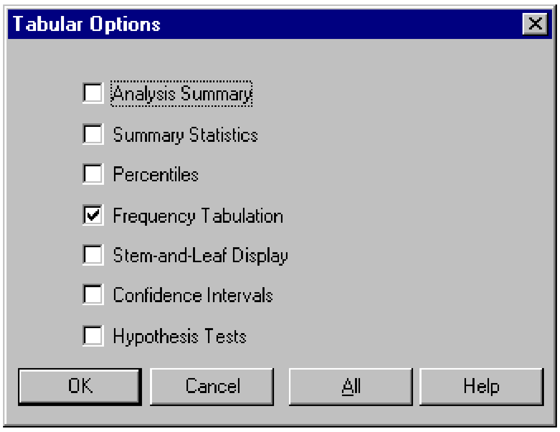


Рис. 1.27. Диалоговое окно выбора табличных опций

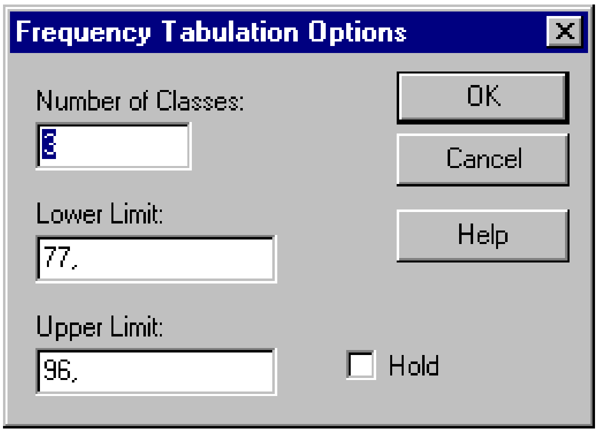


Рис. 1.28. Параметры процедуры группирования

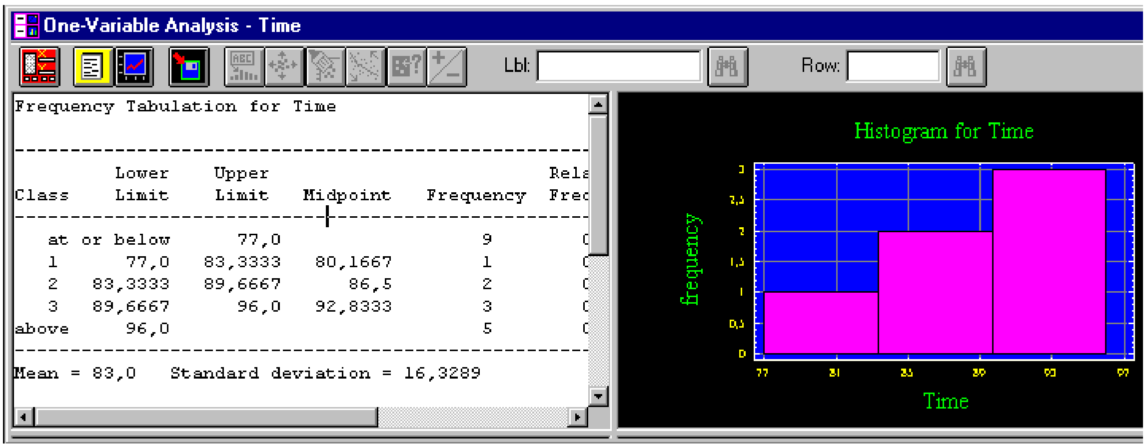
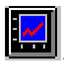


Рис. 1.29. Результаты табулирования данных

Результат группирования и гистограмма распределения частот представлены на рис. 1.29. Вывод на экран гистограммы распределения можно задать в диалоговом окне графических опций .

1.4. Типы измерительных шкал

При решении практических задач анализа диагностической информации исследователю приходится иметь дело с переменными, измеренными в разных шкалах, например, количественных, порядковых, номинальных.

Проиллюстрируем сложную и многоплановую структуру диагностической информации на примере медицинской психодиагностики. Специалисту в этой области приходится оперировать следующими видами информации:

- *психодиагностической* (данные психодиагностического эксперимента, измеренные в количественных, порядковых, номинальных шкалах);
- *медицинской* (результаты параклинических методов исследования, имеющие, как правило, количественный и интервальный характер);
- *консультационально-биологической* (пол, возраст и др.);
- *социально-демографической* (образование, профессия, семейное положение и др.).

Рассмотрим более подробно основные свойства разных типов шкал, эмпирические операции, допустимые на уровне этих шкал, а также статистические приемы обработки и анализа исходных, или, как их чаще называют, первичных результатов исследования.

Шкалы наименований, или номинативные (номинальные) шкалы. Шкала наименований представляет собой взаимно-однозначное отображение некоторой эмпирической системы в числовой системе. Таким образом, шкала наименований отображает взаимооднозначное соответствие между классами эквивалентности, т. е. классами эмпирических объектов – обозначений. Само название «шкала наименований» указывает на то, что в этом случае шкальные значения играют роль лишь названий классов эквивалентности [22].

Шкалы наименований подчиняются законам равенства. То есть объект А может быть равен объекту В по некоторому признаку Х; но по отношению к третьему объекту С по признаку Х он может быть неравным. Любая другая связь между шкальными значениями, за исключением равенства, не имеет отношения к данному случаю, так как для данного типа шкал не существует никакого дополнительного определения.

Шкала наименований представляет собой наиболее общую форму шкал. Все типы шкал в каждом отдельном случае являются некоторыми видами шкал наименований, но обладающими при этом теми или иными дополнительными свойствами. При построении шкал наименований должны быть выполнены следующие требования [22]:

1. Каждый член некоторого множества объектов должен быть отнесен лишь к одному классу объектов (или к собирательному классу «прочие объекты»).

2. Ни один из объектов не может быть отнесен одновременно к двум или большему числу классов.

К примеру, если принять, что глаза у людей могут быть только светлыми или темными, то все люди по этому признаку разделяются на две группы. При этом люди с множеством оттенков глаз: голубых, серо-зеленых и серых попадут в класс «люди со светлыми глазами», а те, у которых глаза карие и темно-коричневые, – в класс «люди с темными глазами». Из приведенного примера видно, что отношения эквивалентности по заданному признаку между классифицируемыми объектами, как правило, грубее реальных отношений, существующих между объектами.

С формальной точки зрения установление классов эквивалентности как будто не вызывает никаких затруднений. В действительности, как это было показано предыдущим примером, понятие «равенство» можно трактовать более узко или более широко в зависимости от «тонкости» или «грубости» используемой классификации по заданному признаку. Рассмотрим еще один пример. Так, если нужно упорядочить события по признаку «мороз/оттепель», то температуры, обозначаемые как +1° и –1°, будут входить в два разных неэквивалентных класса, в то время как температуры +1° и +10° попадут в один класс и по признаку «мороз/оттепель» будут рассматриваться как эквивалентные события.

Приведенные примеры показывают, что при построении шкал наименований главными являются качественные различия, а количественные не принимаются во внимание. Поэтому числа, используемые в качестве обозначений классов эквивалентности в этих шкалах, не отражают количественных различий выраженности изучаемого признака.

В примере с температурой мы имели дело с дихотомической (делением на два класса), или альтернативной классификацией. Эти классификации можно образовать по логическому принципу «А/не-А», т. е. согласно принципу наличия или отсутствия определенного признака. Примерами такого рода классификации могут быть: «нормальный/анормальный», «женатый/холостой», «решает задачу/не решает задачу» и т. п. В случае так называемой истинной дихотомии классы могут быть четко разделены по определенному признаку, например «мужской/женский пол» [27].

Однако бывают классификации с менее жесткими переходами признака, т. е. с довольно произвольными границами между классами эквивалентности, например «способен к концентрации внимания/не способен к концентрации внимания». Именно с такого рода классификациями чаще всего и имеет дело психолог. Это так называемые квазидихотомические классификации. Построение и использование шкал с квазидихотомическими границами классов вызывает ряд затруднений. Первая трудность, которая при этом возникает, состоит в установлении границы классов. В частности, каков же будет в нашем примере критерий «способности» к концентрации внимания, как определить точку в континууме «концентрация внимания», дифференцирующую людей на «способных» и «неспособных» к концентрации внимания?

В [22] приводится пример из области психологии мышления. На первый взгляд, альтернатива «решил задачу/не решил задачу» вполне может быть расценена как истинно дихотомическая классификация. И действительно, в принципе для отнесения любого конкретного решения к классу «решил задачу» достаточно соотнести получаемый в нем результат с результатом, полученным достаточно большой группой людей, аналогичным образом решивших данную задачу. Все остальные решения можно тогда отнести к классу «не решил задачу». Однако возникает вопрос: действительно ли данный человек решил эту задачу? Вполне возможно, во-первых, что решение было случайным, т. е. случайно данный результат совпал с результатом решения других людей, и, во-вторых, что этот класс задач заранее был известен данному человеку. Но, как правило, такого рода сопровождающие факторы, например, в психодиагностических тестах, совершенно не учитываются [22].

В шкале наименований с числами, которые мы приписываем объектам или классам объектов, нельзя производить никаких арифметических действий. Числа, обозначающие классы, нельзя суммировать, вычитать, умножать и делить. Дело в том, что структура шкалы остается инвариантной по отношению к перемене обозначений (наименований) и к изменению последовательности, т. е. разного рода перестановкам. Следовательно, операция присвоения чисел классам объектов является совершенно произвольной, и ей не соответствуют операции, производимые с реальными объектами. Поэтому классы объектов можно обозначить любыми символами – произвольными числами, буквами или другими знаками при одном условии: каждый символ будет использован исключительно для обозначения одного класса объектов и одновременно ни один класс объектов не будет обозначаться двумя или большим числом символов.

Шкалы порядка (ранговые или ординальные) шкалы. В порядковых измерениях символы, в частности числа, присваивают классам объектов так, чтобы первые отображали не только равенство или неравенство, эквивалентность или неэквивалентность, но и упорядоченность объектов в отношении измеряемого свойства. В шкалах порядка классы объектов, как и в случае шкал наименований, являются дискретными. И хотя цифры, обозначающие классы можно сравнивать, всегда надо помнить, что в шкалах порядка они используются только в целях кодирования.

Например, если какой-то один класс объектов обозначен большим числом, чем другой, то мы понимаем, что по измеряемой характеристике первый превосходит второй, но при этом нам неизвестно, насколько велико это различие. Дело в том, что в самих измерительных операциях, связанных с установлением порядка, не содержится никаких данных о величине различий. В [22] в качестве примера рассматриваются оценки знаний материала студентами во время экзаменов. Различия между оценками 5 – «отлично» и 4 – «хорошо» указывают лишь на то, что уровень знаний «отличника» выше уровня знаний «хорошиста». Однако на основе такого рода оценок нельзя сказать, насколько или во сколько раз эти уровни знаний отличаются друг от друга.

Таким образом, шкала порядка отображает монотонное возрастание или убывание измеряемого признака с помощью монотонно возрастающих или монотонно уменьшающихся чисел. Оценить направление изменения признака можно только в том случае, если шкала порядка содержит не меньше трех классов, которые образуют последовательность. Из-за того, что в шкале порядка устанавливается последовательность классов, любые преобразования, связанные с перестановками элементов этой шкалы, недопустимы.

Упорядочение объектов может быть униполярным или биполярным. При униполярном установлении порядка объекты или классы объектов соотносят, используя в качестве индикатора степень выраженности одного-единственного свойства. Например, шкала порядка для оценки умственной отсталости может содержать следующие классы: «нет отклонения от нормы/отклонение слабое/отклонение среднее/ отклонение сильное» [22].

При биполярном упорядочении исходят, как правило, из полярных проявлений какого-то свойства, которые фиксируются в виде двух «точек отсчета» на шкале. Примером биполярной шкалы в психологическом исследовании является методика семантического дифференциала. В этом случае для построения шкалы первоначально производят отбор некоторого множества понятий, которые могут характеризовать, по мнению исследователя, изучаемые психические свойства испытуемого. Затем каждому понятию находят антоним (например: «общительный – замкнутый», «сильный – слабый», «уравновешенный – неуравновешенный»). Очевидно, что между каждыми двумя такими понятиями располагается несколько промежуточных оценочных категорий. Словесное определение промежуточных категорий очень часто вызывает у исследователей значительные трудности, поскольку в языке, как правило, мы легче находим понятия для обозначения экстремальных степеней выраженности какого-то свойства и труднее – для промежуточных.

Примерами использования в психологии порядковых шкал могут служить первичные результаты тестовых испытаний группы лиц, первичные результаты при использовании некоторых личностных опросников, работы со шкалами самооценки и т. п. Можно сказать, что результаты большинства психологических исследований представляют собой ординальные величины, т. е. выражающиеся порядковыми числами. Об этом необходимо помнить, поскольку характер первичных результатов накладывает ряд ограничений на возможность использования тех или других статистических приемов их обработки и анализа. Поскольку в порядковых шкалах не определена единая точка отсчета величин, то и для элементов шкалы порядка, как и для элементов шкал наименований, непригодны способы расчета, требующие арифметических действий – в частности, сложения и вычитания.

Шкалы интервалов. Когда шкала обладает всеми свойствами порядковой шкалы и дополнительно к этому определены еще расстояния между ее единицами, то такую шкалу называют шкалой интервалов. Иначе говоря, классы объектов шкал интервалов всегда дискретны и упорядочены по степени возрастания (или убывания) измеряемого свойства. Кроме того, в этих шкалах одинаковым разностям степени выраженности измеряемого свойства соответствуют равные разности между приписываемыми им числами. Шкалы интервалов имеют равные единицы измерения, однако способ их определения является произвольным, следовательно, и сами единицы произвольны. При этом неизвестна абсолютная величина отдельных значений по шкале, поскольку шкала интервалов не имеет естественной нулевой точки отсчета. Последняя может быть произвольно смещена.

Шкалам интервалов присущи все те отношения, которые характерны для номинативных и порядковых шкал. Кроме того, для них возможно использование арифметических действий. Основными операциями с элементами интервальных шкал являются операции установления равенства, разности, сопоставление больше/меньше в отношении измеряемых свойств, а также утверждение равенства интервалов и равенства разностей между значениями одной шкалы.

Хотя психологические измерения дают нам преимущественно ординальные величины, их обработка часто осуществляется с помощью приемов, допустимых на уровне интервальных шкал. То есть большинство исследователей исходят из равенства интервалов между полученными при измерении величинами. Такой подход основывается чаще всего на следующих предпосылках [22]:

- измеряемая переменная (то или иное свойство объектов) в генеральной совокупности имеет нормальное распределение;

- различные показатели одной и той же переменной обнаруживают линейную корреляцию.

Действительно, на основании этого можно допустить, что интервалы в шкале равны, так как чем более линейна зависимость, тем более равными должны быть интервалы в шкале.

Итак, при конструировании шкалы интервалов используют три произвольные операции:

1) установление величин единиц измерения;
2) определение нулевой точки;
3) определение направления, в котором ведут отсчет по отношению к нулевой точке.

Примером интервальных шкал, используемых в психологии, являются стандартизованные тестовые шкалы Векслера, шкалы Терстена [22].

Шкалы отношений. Конструирование шкал отношений предполагает, наряду с наличием свойств предыдущих шкал, существование постоянной естественной нулевой точки отсчета, в которой измеряемый признак полностью отсутствует. Следовательно, шкалы отношений характеризуются тем, что в них, во-первых, классы объектов разделены и упорядочены согласно измеряемому свойству; во-вторых, равным разностям между классами объектов соответствуют равные разности между приписываемыми им числами; в-третьих, числа, приравниваемые классам объектов, пропорциональны степени выраженности измеряемого свойства. Последнее не было свойственно рассмотренным выше шкалам.

Основными операциями, допустимыми на уровне шкал отношений, являются все те операции, которым подчиняются шкалы всех перечисленных выше типов, и дополнительно – операции установления равенства отношений между отдельными значениями шкалы. Это возможно благодаря существованию на шкале естественного, абсолютного нуля. Поэтому лишь для данной шкалы числа, являющиеся точками (значениями) на шкале, соответствуют реальному количеству измеряемого свойства, что позволяет производить с ними любые арифметические действия – оперирование суммами, произведениями и частными.

Считалось, что шкалы отношений не встречаются в психологических измерениях. Однако Стивенс, исходя из постулата о допустимости непосредственного измерения психических процессов, показал возможность построения шкал отношений в психофизике. Для этой цели он разработал ряд измерительных процедур, предусматривающих прямое шкалирование. Среди них наиболее известными стали методики фракционирования и мультипликации предъявляемых стимулов. К этой же группе методик можно отнести и методики оценки величин стимулов и непосредственной оценки их отношений. Общим для всех перечисленных методик прямого шкалирования является то, что в качестве измерительного инструмента выступает сам испытуемый, который оценивает количественные отношения между раздражителями [22].

Табл. 1.3 подводит итог и дополняет сказанное относительно шкал измерения. Более подробно данная тема рассмотрена в [6], [11], [12], [16], [18].

Таблица 1.3

Основные характеристики и примеры измерительных шкал [6]

Шкала	Характеристики	Примеры
Номинальная	Объекты классифицированы, а классы обозначены номерами. То, что номер одного класса больше или меньше другого, еще ничего не говорит о семействах объектов, за исключением того, что они различаются	Раса, цвет глаз, номера на футболках, пол, клинические диагнозы, автомобильные номера, номера страховых полисов
Порядковая	Соответствующие значения чисел, присваиваемых предметам, отражают количество свойства, принадлежащего предметам. Равные разности чисел не означают равных разностей в количествах свойств	Твердость минералов, награды за заслуги, ранжирование по индивидуальным чертам личности, военные ранги
Интервальная	Существует единица измерения, при помощи которой предметы можно не только упорядочить, но и приписать им числа так, чтобы равные разности чисел, присвоенных предметам, отражали равные различия в количествах измеряемого свойства. Нулевая точка интервальной шкалы произвольна и не указывает на отсутствие свойства	Календарное время, шкалы температур по Фаренгейту и Цельсию
Отношений	Числа, присвоенные предметам, обладают всеми свойствами объектов интервальной шкалы, но, помимо этого, на шкале существует абсолютный ноль. Значение ноль свидетельствует об отсутствии оцениваемого свойства. Отношения чисел, присвоенных в измерении, отражают количественные отношения измеряемого свойства	Рост, вес, время, температура по Кельвину (абсолютный ноль)

1.5. Основные характеристики математической статистики

1.5.1. Меры центральной тенденции

Различные совокупности данных предполагают разные определения «центрального положения». Эти описательные индексы можно использовать для ответа на вопрос вроде «Каков рост типичного выпускника средней школы?». Существуют три такие меры (мода, медиана и среднее), которые рассмотрены ниже. Более подробная информация о мерах центральной тенденции представлена в работе [6].

Наиболее просто получаемой мерой центральной тенденции является *мода.* Мода (чаще всего обозначается как *M*) – это такое значение в множестве наблюдений, которое встречается наиболее часто.

В совокупности значений (2, 6, 6, 8, 9, 9, 9, 10) модой является 9, потому что оно встречается чаще любого другого значения. Обратите внимание, что мода представляет собой наиболее частое значение (в данном примере 9), а не частоту появления этого значения (в примере равную 3).

Медиана (обозначается *Md*) – это значение, которое делит упорядоченное множество данных пополам, так что одна половина значений оказывается больше медианы, а другая – меньше.

Если данные содержат нечетное число различных значений, например 11, 13, 18, 19, 20, то медиана есть центральное значение для случая, когда они упорядочены, т. е. $Md = 18$.

Если данные содержат четное число различных значений, например 4, 9, 13, 14, то медиана есть точка, лежащая посредине между двумя центральными значениями, когда они упорядочены: $Md = (9 + 13)/2 = 11$.

Теперь определим третью меру – выборочное среднее (называемое иногда «средним» или «арифметическим средним»).

Среднее совокупности *n* значений обозначается через $\bar{x}$ и определяется как

$$\bar{x} = \frac{(x_1 + x_2 + ... + x_n)}{n} \qquad (1.2)$$

или

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i . \qquad (1.3)$$

Каждая из приведенных выше мер центральной тенденции обладает характеристиками, которые делают ее ценной в определенных условиях.

Мода наиболее просто вычисляется – ее можно определить на глаз. Кроме того, для очень больших групп данных это достаточно стабильная мера центра распределения. Во многих распределениях значительного числа измерений, используемых в педагогике и психологии, мода близка к двум другим мерам – медиане и среднему.

На величину среднего влияют значения всех результатов, особенно те результаты, которые можно назвать «выбросами», т. е. данные, находящиеся далеко от центра группы оценок. Медиана и мода не требуют для определения всех значений. Посмотрим, что произойдет, например, со средним, медианой и модой, когда удвоится максимальное значение в следующем множестве результатов наблюдений:

	Среднее	Медиана	Мода
1 множество: 1, 3, 3, 5, 6, 7, 8	4,7	5	3
2 множество: 1, 3, 3, 5, 6, 7, 16	5,9	5	3

Другой пример. Мы можем знать средние, медианы и моды для трех разных классов учащихся школы, и необходимо найти те же характеристики для объединения всех трех классов.

В отношении среднего все просто, но для определения медианы и моды приходится вернуться к исходным данным и выполнить новые вычисления.

Пусть средние и числа учащихся для трех классов A, B и C будут:

$$\bar{x}_A = 11{,}9; \qquad n_A = 24;$$
$$\bar{x}_B = 14{,}2; \qquad n_B = 30;$$
$$\bar{x}_C = 10{,}8; \qquad n_C = 28.$$

Таблица 1.4

Выбор меры центральной тенденции в зависимости от типа измерительной шкалы

Тип шкалы	Меры центральной тенденции
Номинальная	Мода
Ранговая	Мода, медиана
Интервальная	Мода, медиана, среднее
Отношений	Мода, медиана, среднее

Общий итог n по всем трем классам $n = n_A + n_B + n_C = 82$. Среднее объединенной группы есть просто сумма всех 82 значений, деленная на 82. Сумма 24 значений класса A равна $n_A \cdot \bar{x}_A = 24 \cdot 11{,}9 = 285{,}6$, ибо среднее – это сумма, деленная на число значений, т. е. $\Sigma x = n\bar{x}$. Аналогично для групп B и C суммы составляют соответственно $30 \cdot 14{,}2 = 426{,}0$ и $28 \cdot 10{,}8 = 302{,}4$. $285{,}6 + 426{,}0 + 302{,}4 = 1014{,}0$. Если мы объединим все 82 значения и просуммируем их, то также получим 1014,0. Таким образом, общее среднее всех трех классов равно 1014,0/82 = 12,4. Символически среднее объединенной группы есть

$$\overline{x} = \frac{n_A \overline{x}_A + n_B \overline{x}_B + n_C \overline{x}_C}{n_A + n_B + n_C}. \tag{1.4}$$

Для определения медианы или моды объединенных выборок необходимо иметь исходные данные.

В табл. 1.4 приводятся данные о возможности использования тех или иных мер центральной тенденции в зависимости от типа измерительных шкал.

1.5.2. Меры изменчивости

К мерам изменчивости переменной относятся следующие характеристики: размах, дисперсия, среднее квадратическое (стандартное) отклонение и среднее линейное отклонение, коэффициент вариации.

Размах измеряет на числовой шкале расстояние, в пределах которого изменяются оценки. Поскольку существуют несколько иные определения размаха, то надо разграничить два его типа: включающий и исключающий.

Исключающий размах – это разность максимального и минимального значений в выборке [6].

Например, исключающий размах значений 0, 2, 3, 5, 8 равен 8 – 0 = 8. Значения: –0,2; 0,4; 0,8; 1,6 имеют исключающий размах, равный 1,6 – (–0,2) = 1,8.

Включающий размах – это разность между *естественной верхней границей* интервала, содержащего максимальное значение, и *естественной нижней границей* интервала, включающего минимальное значение [6]. Или к величине исключающего размаха добавляют значение единицы точности измерения.

Например, рост пяти мальчиков измеряется с точностью до 1 сантиметра. Получены следующие значения:

1	2	3	4	5
150 см	155 см	157 см	165 см	168 см

Таким образом, включающий размах равен 168 – 149 = 18+1=19.

Размах представляет собой меру рассеяния, разброса, неоднородности или изменчивости. Эта величина возрастает с ростом рассеяния и уменьшением однородности. Размах является довольно грубой, но достаточно распространенной мерой изменчивости.

Более точной мерой изменчивости является *дисперсия* (обозначается D, σ_x^2 или s_x^2), которая определяется по формуле

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}. \tag{1.5}$$

Ценность дисперсии заключается в том, что, являясь мерой варьирования числовых значений признака вокруг его среднего значения, она измеряет внутреннюю изменчивость значений признака, зависящую от разностей между наблюдениями. Преимущество дисперсии перед другими показателями вариации состоит также и в том, что она разлагается на составные компоненты, позволяя тем самым оценивать влияние различных факторов на величину учитываемого признака.

Мерой изменчивости, тесно связанной с дисперсией, является стандартное отклонение. *Среднее квадратическое* или *стандартное отклонение*, обозначаемое s_x (или σ_x), определяется как положительное значение квадратного корня из дисперсии:

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}. \tag{1.6}$$

Стандартное отклонение часто является полезной мерой вариации, так как для многих распределений мы приблизительно знаем, какой процент данных лежит внутри одного, двух, трех и более стандартных отклонений среднего. Например, мы можем знать, что 70 % значений лежит между $\bar{x} - s_x$ и $\bar{x} + s_x$.

Еще одна мера изменчивости, *среднее отклонение*, используется реже. *Отклонение* каждого значения от среднего обозначается как $x_i - \bar{x}$. Совокупность всех n отклонений характеризует изменчивость в исходных данных. Однако сумма положительных и отрицательных отклонений вовсе не является мерой общей изменчивости в группе данных, ибо она всегда точно равна нулю. Если рассматривать отклонения как расстояние каждого объекта до $\bar{x}$ без учета знака, то сумма этих расстояний будет характеризовать изменчивость данных, а именно

$$\bar{d} = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}. \tag{1.7}$$

Среднее отклонение не часто используется как мера изменчивости в связи с тем, что среднее отклонение не имеет теоретического обоснования в отличие, например, от дисперсии.

Любое множество n данных со средним $\bar{x}$ и стандартным отклонением s_x можно преобразовать в другое множество со средним 0 и стандартным отклонением 1 таким образом, что преобразованные значения будут непосредственно выражаться в отклонениях исходных значений от среднего, измеренных в единицах стандартного отклонения.

Новые значения называют значениями z:

$$z_i = \frac{x_i - \bar{x}}{s_x}. \tag{1.8}$$

Значение z – не только удобное средство информации о положении некоторого значения, связанного со средним и измеренного в единицах стандартного отклонения, но и шаг вперед к преобразованию множества X в произвольную шкалу с удобными характеристиками среднего и стандартного отклонения. Сами оценки z могут не подходить для некоторых целей. Отрицательные оценки, например, могут оказаться неудобными, а множество z будет, конечно, содержать дроби. Преобразование самих z позволяет устранить эти несущественные трудности.

Известно, что значения cz, полученные умножением каждого z на константу c, будут иметь стандартное отклонение $|c|$, а для $cz + d$ среднее равно d, *а именно*

$$cz + d = \mathrm{c} \times 0 + \mathrm{d} = \mathrm{d}.$$

В практике довольно часто приходится сравнивать изменчивость признаков, выраженных разными единицами. В таких случаях используют не абсолютные, а относительные показатели вариации. Дисперсия и среднее отклонение как величины, выражаемые теми же единицами, что и характеризуемый ими признак, для оценки изменчивости разноименных величин непригодны. Одним из относительных показателей вариации является *коэффициент вариации.* Этот показатель представляет собой среднее квадратическое отклонение, выраженное в процентах от величины среднего значения:

$$Cv = \frac{s_x}{\bar{x}} \cdot 100 \ \ \%. \tag{1.9}$$

Различные признаки характеризуются различными коэффициентами вариации. Но в отношении одного и того же признака значение этого показателя Cv остается более или менее устойчивым и при симметричных распределениях обычно не превышает 50 %. Согласно [17] при сильно асимметричных рядах распределения коэффициент вариации может достигать 100 % и даже выше. Варьирование считается слабым, если не превосходит 10 %, средним, когда Cv составляет 11…25 %, и значительным при $Cv > 25$ %.

1.5.3. Асимметрия

Одно из наиболее важных свойств распределения частот – *степень асимметрии.* Практически точно симметричные полигоны частот и гистограммы почти никогда не встречаются. Степень асимметрии распределения частот для выборки называется просто его *асимметрией.* Легко выявить и распознать асимметрию, если рассматривать полигон частот или гистограмму, но это не всегда возможно или удобно. Поэтому используют различные обобщенные статистические характеристики, оценивающие вид и степень асимметрии группы наблюдений.

Мера асимметрии (As) для группы данных выражается формулой

$$As = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3 / n}{s_x^3}. \qquad (1.10)$$

Непосредственное расстояние, на котором лежит оценка от среднего группы, часто обозначают в единицах стандартного отклонения этой группы с помощью z_x, т. е.

$$z_x = \frac{x_i - \overline{x}}{s_x}.$$

Тогда мера асимметрии в (1.10) принимает вид [6]:

$$As = \frac{\sum z_{x_i}^3}{n} = \overline{z}^3. \qquad (1.11)$$

Пусть измеряется асимметрия двух распределений (рис. 1.30).

Среднее данных распределения A на рис. 1.30 равно приблизительно 13. Когда для случая A строятся значения z, получается значительно больше положительных значений (поскольку максимальная оценка 19 и расположена на 6 единиц выше среднего), которые по абсолютной величине существенно превзойдут любые отрицательные (минимальное значение 10 лежит на 3 единицы ниже среднего). Для асимметрии алгебраический знак сохраняется, так как находится куб числа $(-2)^3 = -8$. Поэтому для распределения A вклад отрицательных значений z, возведенных в куб, $\sum z_x^3 / n$ будет меньше вклада «кубов» больших положительных значений. Следовательно, величина $\sum z_x^3 / n$ в отношении (1.11) будет положительной и большой. Мы говорим, что распределение *A* имеет *положительную асимметрию*, так как его мера асимметрии положительна.

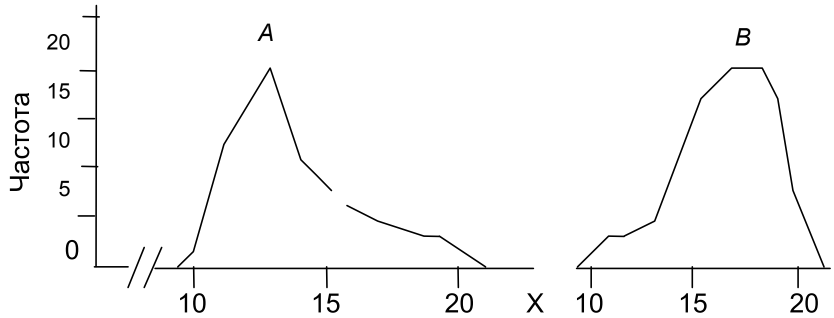


Рис. 1.30. Два асимметричных распределения частот [6]

Распределение *B* на рис. 1.30 имеет *отрицательную асимметрию*. Величина $\sum z_x^3 / n$ для распределения *B* отрицательна. Попытайтесь сами убедиться в этом. Положительная асимметрия распределения *A* более отчетливо выражена, чем отрицательная для распределения *B*.

В симметричном распределении мера асимметрии в отношении (1.11) равна нулю. Это естественно, поскольку точная симметрия означает, что каждое отрицательное значение z уравновешивается положительным значением равной величины.

В [6] приводится упрощенная формула для расчета асимметрии, основанная на том факте, что на величину среднего крайние значения оказывают большее влияние, чем на медиану (*Md*). В положительно асимметричных распределениях среднее больше медианы, которая, в свою очередь, больше моды. В отрицательно асимметричных распределениях, наоборот, – среднее меньше медианы, которая, в свою очередь, меньше моды. Это значит, что положение среднего по отношению к медиане информирует нас в какой-то степени об асимметрии распределения. Это верно для умеренно больших выборок, например объемом 50 или более. Простейшая мера асимметрии, основанная на этих фактах, определяется следующим образом [6]:

$$As = \frac{3(\bar{x} - Md)}{s_x}. \qquad (1.12)$$

Асимметрия из отношения (1.12), может принимать значения в диапазоне от –3 до +3. Когда распределение симметрично, отношение

(1.12) даст ноль. Такую меру асимметрии можно использовать для сравнения различных распределений, так как деление на s_x сделало эту меру независимой от изменчивости распределения.

1.5.4. Эксцесс

Иногда важно получить представление о том, является ли полигон частот или гистограмма островершинными или плоскими. *Эксцесс* – греческое слово, обозначающее свойство «остроконечности» кривой. Карл Пирсон формализовал понятие «эксцесс» в статистике и предложил метод его оценки.

На рис. 1.31 изображены три кривые, отличающиеся по «остроконечности», или эксцессу.

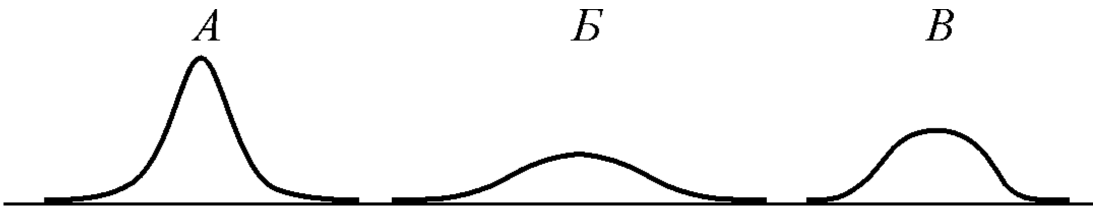


Рис. 1.31. «Островершинная», «плоская» и «промежуточная» («средневершинная») кривые (А, Б, В соответственно)

Первая кривая (*А*) является совсем острой: подобная кривая называется *островершинной.* Вторая (*Б*) – сравнительно плоская: такие кривые называются *плосковершинными.* «Островершинность», или степень эксцесса, третьей кривой (*В*) представляет собой норму, по отношению к которой измеряется эксцесс других кривых. Третья кривая на рис. 1.31 – нормальная кривая, которая будет обсуждаться в соответствующей главе; принято говорить, что она является *средневершинной.*

Теперь мы рассмотрим способ измерения эксцесса кривой. Однако сначала необходимо подчеркнуть, что понятие «эксцесс» применимо лишь к одномодальным распределениям и относится к крутизне кривой в окрестности единственной моды (если распределение имеет две моды, то принято говорить об эксцессе кривой в окрестности каждой моды).

Обычная мера эксцесса (*Ex*) определяется следующей формулой:

$$Ex = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^4 / n}{s_x^4} - 3. \tag{1.13}$$

Соотношения между величиной статистики асимметрии и «островершинностью» распределения, для которого она вычислялась, показаны в табл. 1.5.

На рис. 1.32–1.34 представлены диалоговые окна для выбора вычисляемых характеристик в трех статистических пакетах:

- среднего (Average), медианы (Median), моды (Mode);
- дисперсии (Variance);
- стандартного отклонения (Standard Deviation);
- минимума (Minimum) и максимума (Maximuma);
- размаха (Range);
- коэффициента ассимметрии (Skewness);
- значение эксцесса (Kurtosis) и др.

Таблица 1.5

Соотношение величины статистики эксцесса
с «островершинностью» распределения частот

Характер распределения	Описание «островершинности»	Величина эксцесса
Нормальное, например кривая *В* на рис. 1.31	Средневершинное	0
Островершинное, например кривая *А* на рис. 1.31	Островершинное	Больше 0 (может быть очень большой)
Плоское, например кривая *Б* на рис. 1.31	Плосковершинное	Меньше 0

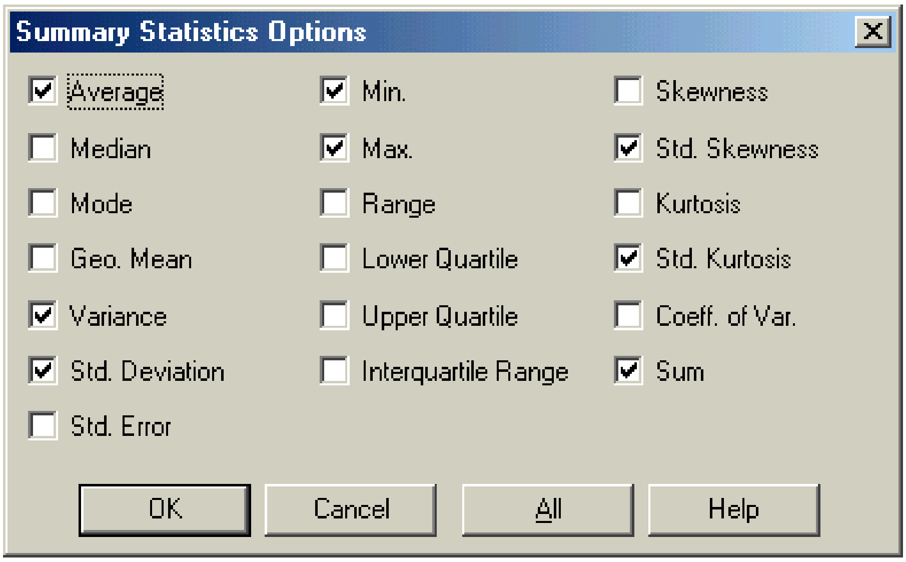


Рис. 1.32. Вычисляемые статистические характеристики пакета STATGRAPHICS

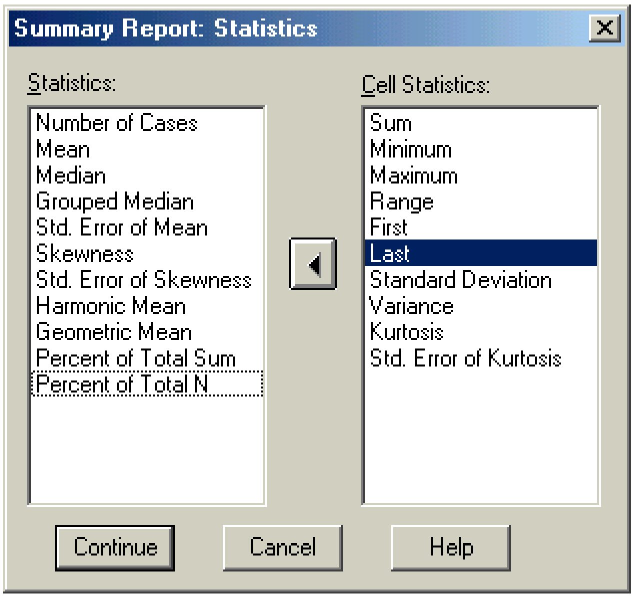


Рис. 1.33. Вычисляемые статистические характеристики пакета SPSS

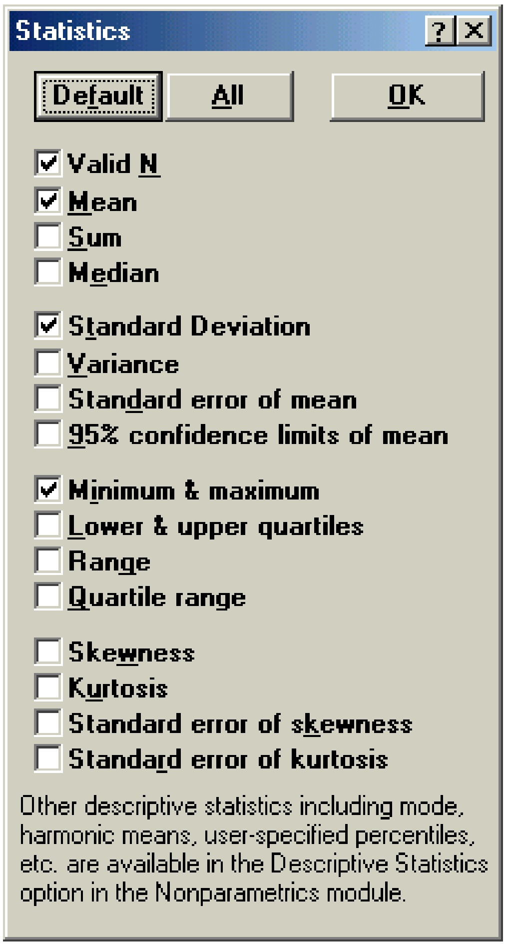


Рис. 1.34. Вычисляемые статистические характеристики пакета STATISTICA

Глава 2
ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Во многих случаях исследователю требуется на основе тех или иных данных решить, справедливо ли некоторое суждение. Например, верно ли, что два набора данных исходят из одного и того же источника? Что *A* – лучший ученик, чем *B*? Что от дома до работы быстрее доехать на метро, а не на автобусе, и т. д. Если мы считаем, что исходные данные для таких суждений в той или иной мере носят случайный характер, то и ответы можно дать лишь с определенной степенью уверенности, и имеется некоторая вероятность ошибиться. Например, предложив двум персонам *A* и *B* выстрелить по три раза в мишень и осмотрев результаты стрельбы, мы лишь предположительно можем сказать, кто из них лучший стрелок: ведь возможно, что победителю просто повезло, и он по чистой случайности стрелял намного точнее, чем обычно, либо наоборот, проигравшему не повезло, так как он стрелял намного хуже, чем обычно. Поэтому при ответе на подобные вопросы хотелось бы не только уметь принимать наиболее обоснованные решения, но и оценивать вероятность ошибочности принятого решения [24].

Рассмотрение таких задач в строгой математической постановке приводит к понятию *статистической гипотезы.* В этой главе рассматриваются вопросы о том, что такое статистические гипотезы и какие существуют способы их проверки.

2.1. Основные понятия

2.1.1. Статистические модели

Прежде чем приступить к описанию статистических гипотез, вспомним понятие случайного выбора. Как справедливо отмечено в [24], весь статистический анализ основан на *идее случайного выбора.* Мы принимаем тезис, что имеющиеся данные появились как результат случайного выбора из некоторой генеральной совокупности, нередко – воображаемой. Обычно мы полагаем, что этот случайный выбор произведен природой. Впрочем, во многих задачах эта генеральная совокупность вполне реальна, и выбор из нее произведен активным наблюдателем.

Поскольку мы приняли вероятностную точку зрения на происхождение наших данных (т. е. считаем, что они получены путем случайного выбора), то все дальнейшие суждения, основанные на этих данных, будут иметь вероятностный характер. Всякое утверждение будет верным лишь с некоторой вероятностью и с некоторой тоже положительной вероятностью оно может оказаться неверным.

Какую вероятность следует считать малой? На этот вопрос нельзя дать количественного ответа, пригодного во всех случаях. Ответ зависит от того, какой опасностью грозит нам ошибка. Довольно часто при проверке статистических гипотез, например, полагают малыми вероятности, начиная с 0,01…0,05 [24].

2.1.2. Статистические гипотезы

В обычном языке слово «гипотеза» означает предположение. В том же смысле оно употребляется и в научном языке, используясь в основном для предположений, вызывающих сомнения. В математической статистике термин «гипотеза» означает предположение, которое не только вызывает сомнения, но и которое мы собираемся в данный момент проверить.

При построении статистической модели приходится делать много различных допущений и предположений, и далеко не все из них мы собираемся или можем проверить. Эти предположения относятся как к выборочному пространству, так и к распределению вероятностей на нем.

Формулирование гипотез систематизирует предположения исследователя и представляет их в четком и лаконичном виде. Благодаря гипотезам исследователь не теряет путеводной нити в процессе расчетов и ему легко понять после их окончания, что, собственно, он обнаружил.

Статистические гипотезы подразделяются на *нулевые* и *альтернативные*, *направленные* и *ненаправленные*.

Нулевая гипотеза – это гипотеза об отсутствии различий.

Она обозначается как H_0 и называется нулевой потому, что содержит число 0: $x_1 - x_2 = 0$, где x_1, x_2 – сопоставляемые значения признаков.

Нулевая гипотеза – это то, что мы хотим опровергнуть, если перед нами стоит задача доказать значимость различий [24].

Альтернативная гипотеза – это гипотеза о значимости различий.

Она обозначается как H_1. Альтернативная гипотеза – это то, что мы хотим доказать, поэтому иногда ее называют *экспериментальной* гипотезой [24].

Бывают задачи, когда мы хотим доказать как раз незначимость различий, т. е. подтвердить нулевую гипотезу. Например, если нам нужно убедиться, что разные испытуемые получают хотя и различные, но уравновешенные по трудности задания, или что экспериментальная и контрольная выборки не различаются между собой по каким-то значимым характеристикам. Однако чаще нам все-таки требуется доказать *значимость различий*, ибо они более информативны для нас в поиске нового. Нулевая и альтернативная гипотезы могут быть направленными и ненаправленными.

Направленные гипотезы

H_0: x_1 не превышает x_2,

H_1: x_1 превышает x_2.

Ненаправленные гипотезы

H_0: x_1 не отличается от x_2,

H_1: x_1 отличается от x_2.

Если исследователь обнаружил, что в одной из групп индивидуальные значения испытуемых по какому-либо признаку, например по социальной смелости, выше, а в другой ниже, то для проверки значимости этих различий нам необходимо сформулировать направленные гипотезы.

Если мы хотим доказать, что в группе А под влиянием каких-то экспериментальных воздействий произошли более выраженные изменения, чем в группе Б, то нам тоже необходимо сформулировать направленные гипотезы.

Если же мы хотим доказать, что различаются формы распределения признака в группе А и Б, то формулируются ненаправленные гипотезы.

Проверка гипотез осуществляется с помощью критериев статистической оценки различий. Критерии делятся на параметрические и непараметрические.

Параметрические критерии – это критерии, включающие в формулу расчета параметры распределения, т. е. средние и дисперсии (t-критерий Стьюдента, критерий F и др.).

Непараметрические критерии – это критерии, не включающие в формулу расчета параметров распределения и основанные на оперировании частотами или рангами (Q-критерий Розенбаума, T-критерий Уилкоксона и др.).

При нормальном распределении признака параметрические критерии обладают большей мощностью, чем непараметрические критерии. Они способны отвергать нулевую гипотезу, если она неверна. Поэтому во всех случаях, когда сравниваемые выборки взяты из нормально распределяющихся совокупностей, следует отдавать предпочтение параметрическим критериям.

В случае очень больших отличий распределений признака от нормального вида следует применять непараметрические критерии, которые в этой ситуации оказываются часто более мощными. В ситуациях, когда варьирующие признаки выражаются не числами, а условными знаками, применение непараметрических критериев оказывается единственно возможным.

2.1.3. Статистические критерии

При проведении научных исследований, если гипотезу можно проверить непосредственно, – тогда не возникает никаких методических проблем. Но если прямого способа проверки у нас нет, приходится прибегать к проверкам косвенным. Это значит, что приходится довольствоваться проверкой некоторых следствий, которые логически вытекают из содержания гипотезы. Если некоторое явление логически неизбежно следует из гипотезы, но в природе не наблюдается, то это значит, что гипотеза неверна [24]. С другой стороны, если происходит то, что при гипотезе происходить не должно, это тоже означает ложность гипотезы. Заметим, что подтверждение следствия еще не означает справедливости гипотезы, поскольку правильное заключение может вытекать и из неверной предпосылки. Поэтому, строго говоря, косвенным образом *доказать* гипотезу нельзя, хотя *опровергнуть* – можно.

Проверку статистических гипотез осуществляют при помощи так называемых статистических критериев.

Статистический критерий – это решающее правило, обеспечивающее надежное поведение, т. е. принятие истинной и отклонение ложной гипотезы с высокой вероятностью [31].

Ниже приведена схема проверки гипотез с помощью статистических критериев.

1. По расчетной формуле вычисляется эмпирическое (или фактическое) значение критерия. Обозначим это значение в общем случае как $F_{\text{эмп}}$ (или $F_{\text{ф}}$).
2. По таблицам критических значений для выбранного критерия находится так называемая критическая точка (или критическое значение). Обозначим это значение $F_{\text{кр}}$ (иногда в тексте F_{st}). Для того чтобы правильно воспользоваться статистическими таблицами, необходимо знать число степеней свободы и определить уровень значимости.
3. По соотношению эмпирического и критического значений критерия мы можем судить о том, подтверждается ли или опровергается нулевая гипотеза. Например, если $F_{\text{эмп}} > F_{\text{кр}}$, H_0 отвергается.

В большинстве случаев для того, чтобы мы признали различия значимыми, необходимо, чтобы эмпирическое значение критерия превышало критическое, хотя есть критерии (например, критерий Манна–Уитни или критерий знаков), в которых мы должны придерживаться противоположного правила.

Число степеней свободы

В некоторых случаях расчетная формула $F_{\text{эмп}}$ включает в себя количество наблюдений в исследуемой выборке, обозначаемое как n.

В этом случае эмпирическое значение критерия одновременно является тестом для проверки статистических гипотез. По специальной таблице мы определяем, какому уровню статистической значимости различий соответствует данная эмпирическая величина. Примером такого критерия является критерий (φ*, вычисляемый на основе углового преобразования Фишера).

В большинстве случаев, однако, одно и то же эмпирическое значение критерия может оказаться значимым или незначимым в зависимости от количества наблюдений в исследуемой выборке (n) или от так называемого количества степеней свободы, которое обозначается как v или df.

Число степеней свободы v равно числу классов вариационного ряда минус число условий, при которых он был сформирован [15]. К числу таких условий относятся объем выборки (n), средние и дисперсии.

Если мы расклассифицировали наблюдения по классам какой-либо номинативной шкалы и подсчитали количество наблюдений в каждой ячейке классификации, то мы получаем так называемый частотный вариационный ряд. Единственное условие, которое соблюдается при его формировании – объем выборки n. В [24] приводится следующий пример. Допустим, у нас 3 класса: «Умеет работать на компьютере – умеет выполнять лишь определенные операции – не умеет работать на компьютере». Выборка состоит из 50 человек. Если в первый класс отнесены 20 испытуемых, во второй – тоже 20, то в третьем классе должны оказаться все остальные 10 испытуемых. Мы ограничены одним условием – объемом выборки. Поэтому даже если мы потеряли данные о том, сколько человек не умеют работать на компьютере, мы можем определить это, зная, что в первом и втором классах – по 20 испытуемых. Мы не свободны в определении количества испытуемых в третьем разряде, «свобода» простирается только на первые две ячейки классификации:

$$v = c - 1 = 3 - 1 = 2. \qquad (2.1)$$

Аналогичным образом, если бы у нас была классификация из 10 разрядов, то мы были бы свободны только в 9 из них, если бы у нас было 100 классов – то в 99 из них и т. д.

Способы более сложного подсчета числа степеней свободы при двухмерных классификациях приведены в разделах, посвященных критерию χ^2 и дисперсионному анализу.

Зная n и/или число степеней свободы, мы по специальным таблицам можем определить критические значения критерия и сопоставить с ними полученное эмпирическое значение. Обычно это записывается так: «при n = 22 критические значения критерия составляют…» или «при v = 2 критические значения критерия составляют…» и т. п.

Уровень значимости

Уровень значимости – это вероятность отклонения нулевой гипотезы, в то время как она верна.

Обычно при проверке статистических гипотез принимают три уровня значимости: 5%-й (вероятность ошибочной оценки $p = 0,05$), 1%-й ($p = 0,01$) и 0,1%-й ($p = 0,001$). В психологических исследованиях часто считают достаточным 5%-й уровень значимости. При этом нулевую гипотезу не отвергают, если в результате исследования окажется, что вероятность ошибочности оценки относительно правильности принятой гипотезы превышает 5 %, т. е. $p > 0,05$. Если же $p < 0,05$, то принятую гипотезу следует отвергнуть на взятом уровне значимости. Ошибка при этом возможна не более чем в 5 % случаев, т. е. она маловероятна.

При более ответственных исследованиях уровень значимости может быть уменьшен до 1 % или даже до 0,1 %.

Ошибка, состоящая в том, что мы отклонили нулевую гипотезу, в то время как она верна, называется *ошибкой I рода.*

Вероятность такой ошибки обычно обозначается как α. Если вероятность ошибки – это α, то вероятность правильного решения: $1 - \alpha$.

Сидоренко в работе [24] для облегчения процесса принятия решения предлагает вычерчивать «ось значимости».

Критические значения критерия при уровнях значимости $p < 0,05$ и $p < 0,01$ обозначены, соответственно, как $F_{0,05}$ и $F_{0,01}$, эмпирическое значение критерия как $F_{эмп}$. На рис. 2.1 оно заключено в эллипс.

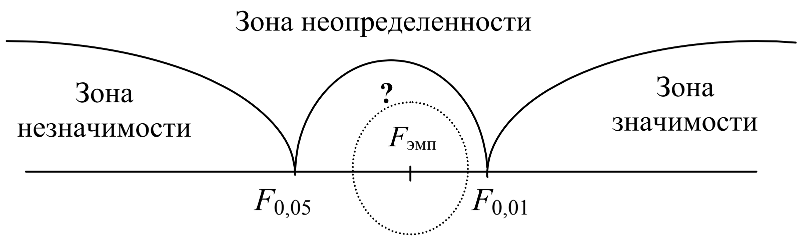


Рис. 2.1. Пример «оси значимости»
для некоторого статистического критерия [24]

Вправо от критического значения $F_{0,01}$ простирается «зона значимости» – сюда попадают эмпирические значения, превышающие $F_{0,01}$ и, следовательно, безусловно значимые.

Влево от критического значения $F_{0,05}$ простирается «зона незначимости» – сюда попадают эмпирические значения F, которые ниже $F_{0,05}$, и, следовательно, безусловно незначимы.

Если эмпирическое значение критерия ($F_{эмп}$) попадает в область между $F_{0,05}$ и $F_{0,01,}$ т. е. в «зону неопределенности»: мы уже можем отклонить гипотезу о недостоверности различий (H_0), но еще не можем принять гипотезу об их достоверности (H_1).

Согласно [24] исследователь может считать достоверными уже те различия, которые не попадают в зону незначимости, заявив, что они достоверны при $p \leq 0{,}05$, или указав точный уровень значимости полученного эмпирического значения критерия, например $p = 0{,}02$. Это можно сделать с помощью таблиц критических значений, приведенных в приложении (например, см. [6, 24]).

Следует отметить, что при компьютерном анализе данных таблицы критических значений не нужны. Это связано с тем, что в данных пакетах для каждого статистического критерия выходная информация, наряду с эмпирическими значениями критерия, включает в себя точное значение уровня значимости, при котором мы имеем право отвергнуть нулевую гипотезу. Если это значение (p) меньше 0,05, то нулевая гипотеза отвергнута.

Уровень статистической значимости или критические значения критериев определяются по-разному при проверке направленных и ненаправленных гипотез.

При направленной гипотезе используется односторонний критерий, при ненаправленной гипотезе – двусторонний. Двусторонний критерий более строг, поскольку он проверяет различия в обе стороны, и поэтому то эмпирическое значение критерия, которое ранее соответствовало уровню значимости $p \leq 0{,}05$, теперь соответствует лишь уровню $p \leq 0{,}10$. Подробно о двусторонних критериях см. [24].

Мощность критериев

Мощность критерия – это его способность выявлять различия, если они есть. Иными словами, это его способность отклонить нулевую гипотезу об отсутствии различий, если она неверна.

Ошибка, состоящая в том, что мы приняли нулевую гипотезу, в то время как она неверна, называется *ошибкой II рода.*

Вероятность такой ошибки обозначается как β. Величину $1 - \beta$ обычно называют *мощностью критерия.* Ясно, что мощность критерия может принимать любые значения от 0 до 1. Чем ближе мощность критерия к единице, тем более эффективен (более «мощен») критерий. Многие известные статистические критерии получены путем нахождения наиболее мощного критерия при заданных предположениях о гипотезе и альтернативе.

Из-за различного подхода к гипотезе и альтернативе наше отношение к ошибкам первого и второго рода также неодинаково. При построении статистических критериев мы фиксируем максимальную допустимую вероятность ошибки первого рода (т. е. уровень значимости критерия) и стремимся выбрать критическое множество таким образом, чтобы минимизировать вероятность ошибки второго рода (или хотя бы сделать так, чтобы эта вероятность была как можно меньше по мере удаления истинного распределения от гипотетического или гипотетических) [24].

2.1.4. Классификация основных задач прикладного эксперимента и методы их решения

Множество задач психологического исследования предполагает те или иные сопоставления. Сопоставляются группы испытуемых по какому-либо признаку, чтобы выявить различия между ними по этому признаку.

Можно сопоставлять два признака, измеренные на одной и той же выборке испытуемых, для того, чтобы установить степень согласованности их измерений, их сопряженность, корреляцию между ними.

Также мы можем сопоставлять индивидуальные значения, полученные при разных комбинациях каких-либо существенных условий, с тем чтобы выявить характер взаимодействия этих условий в их влиянии на индивидуальные значения признака.

Краткая классификация задач и статистических методов их решения приведена в табл. 2.1.

Таблица 2.1

Классификация задач и методов их решения

<table>
<tr><th>Задачи</th><th>Условия</th><th>Методы</th><th>Примечание</th></tr>
<tr><td rowspan="4">1. Выявление различий в уровне исследуемого признака</td><td rowspan="3">2 выборки испытуемых</td><td>Q-критерий Розенбаума;</td><td>Раздел 3.4</td></tr>
<tr><td>U-критерий Манна–Уитни;</td><td>Раздел 3.4</td></tr>
<tr><td>φ*-критерий (угловое преобразование Фишера)</td><td>Раздел 3.5</td></tr>
<tr><td>3 и более выборок испытуемых</td><td>H-критерий Крускала–Уолиса</td><td>Раздел 3.4</td></tr>
<tr><td rowspan="2">2. Выявление различий в средних значениях исследуемого признака</td><td>а) 2 выборки испытуемых</td><td rowspan="2">t-критерий Стьюдента</td><td rowspan="2">Раздел 2.3</td></tr>
<tr><td>б) исследуемый признак является нормально распределенным</td></tr>
</table>

Окончание табл. 2.1

Задачи	Условия	Методы	Примечание
3. Оценка сдвига значений исследуемого признака	2 выборки испытуемых	Т-критерий Уилкоксона;	Раздел 2.4
		G-критерий знаков;	Раздел 2.4
		φ*-критерий (угловое преобразование Фишера)	Раздел 2.5
4. Выявление степени согласованности изменений	двух нормально распределенных признаков	Коэффициент корреляции Пирсона	Раздел 3.1
	двух качественных или количествен-ных признаков	Непараметрические коэффициенты корреляции	Раздел 3.2
5. Анализ изме-нений признака под влиянием контролируемых условий	Под влиянием одного фактора	Однофакторный дисперсионный анализ	Глава 4
	Под влиянием двух факторов одновре-менно	Двухфакторный дисперсионный анализ	См. [6], [34]
6. Выявление скрытых законо-мерностей и вы-деление новых «обобщенных» переменных	Признаки измерены в количественной или ранговой шкале	Факторный анализ	Глава 5
7. Выделение однородных групп по сово-купности изме-ряемых признаков	Признаки измерены в количественной шкале	Кластерный анализ	Глава 5

2.2. Параметрический критерий проверки статистических гипотез *t*-критерий Стьюдента

2.2.1. *t*-распределение

Одним из наиболее распространенных параметрических критериев является критерий Стьюдента, в основе которого лежит *t*-распределение. Английский математик В. Госсет (печатавшийся под псевдонимом Стьюдент) в 1908 г. нашел *закон распределения значений*:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}. \qquad (2.2)$$

Оказалось, что отношение разности между выборочной и генеральной средними к ошибке выборочной средней непрерывно распределяется согласно следующей формуле:

$$f(t) = C\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n-1}{2}} \text{ для } -\infty < t < +\infty, \qquad (2.3)$$

где C – константа, зависящая только от степеней свободы $k = n - 1$.

Открытый Стьюдентом и теоретически обоснованный Р. Фишером *закон t-распределения* служит основой так называемой теории малой выборки, которая характеризует распределение выборочных средних в нормально распределяющейся совокупности в зависимости от объема выборки. Данное распределение зависит только от числа степеней свободы $k = n - 1$, причем с увеличением объема выборки n t-распределение быстро приближается к нормальному с параметрами $\mu = 0$ и $\sigma = 1$ и уже при $n > 30$ не отличается от него. Это видно из табл. 2.4, в которой наряду с табулированными значениями функции нормального распределения приведены табулированные значения t-распределения для разных значений t.

Таблица 2.2 [7]

Распределение	Нормированное отклонение t						
	0,5	1,0	1,5	2,0	2,5	3,0	3,5
Нормальное	0,383	0,683	0,866	0,955	0,988	0,997	0,9995
Стьюдента при $n = 3$	0,333	0,577	0,728	0,816	0,870	0,905	0,927
$n = 20$	0,377	0,670	0,850	0,940	0,978	0,993	0,998
$n = 30$	0,383	0,683	0,866	0,955	0,988	0,997	0,9995

Более наглядное представление о характере t-распределения дает рис. 2.2, на котором на фоне нормальной кривой изображена (более пологая) кривая t-распределения при $n = 2$. Это распределение симметрично и отражает специфику распределения средней арифметической в случае малой выборки в зависимости от ее объема *(n)*.

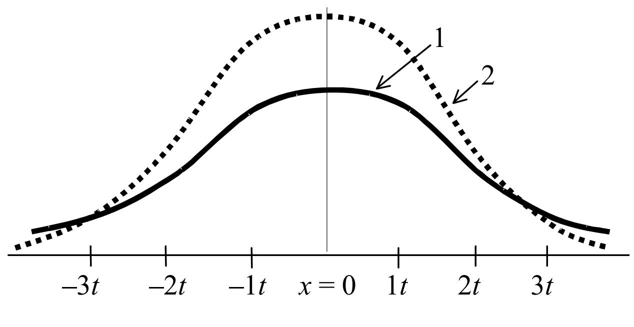


Рис. 2.2. Кривая t-распределения (1) при n = 3 на фоне нормальной кривой (2)

Для выборок, объем которых превышает 30 единиц, величина t распределяется нормально и не зависит от числа наблюдений. Если же $n < 30$, характер t-распределения находится в зависимости от числа наблюдений n.

Для практического использования t-распределения составлена специальная таблица (см. табл. IX приложения), в которой содержатся критические точки ($t_{кр}$) для разных уровней значимости α и чисел степеней свободы k.

2.2.2. Оценка разности средних для независимых выборок

Нулевая гипотеза сводится к предположению, что разность средних в выборке 1 и выборке 2 равна нулю. Эмпирическое значение t-критерия выражается в виде отношения разности выборочных средних к своей ошибке:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_d} = \frac{d}{s_d}. \tag{2.3}$$

H_0-гипотезу отвергают, если фактически установленная величина t-критерия (обозначаемая символом $t_{эмп}$) превзойдет или окажется равной критическому (стандартному) значению $t_{кр}$ этой величины для принятого уровня значимости α и числа степеней свободы $k = n_1 + n_2 - 2$, т. е. при условии

$$t_{эмп} \geq t_{кр}.$$

Ошибку разности средних s_d определяют по следующим формулам [15]:

а) для равночисленных выборок, т. е. при $n_1 = n_2$:

$$\begin{aligned} s_d &= \sqrt{s_{x_1}^2 + s_{x_2}^2} = \sqrt{\frac{\sum(x_i - \bar{x}_1)^2}{n(n-1)} + \frac{\sum(x_i - \bar{x}_2)^2}{n(n-1)}} = \\ &= \sqrt{\frac{\sum(x_i - \bar{x}_1)^2 + \sum(x_i - \bar{x}_2)^2}{n(n-1)}}; \end{aligned} \tag{2.4}$$

б) для неравночисленных выборок, т. е. при $n_1 \neq n_2$:

$$\begin{aligned} s_d &= \sqrt{\frac{(n_1 - 1)s_1^2 + (n_1 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{n_1 + n_2}{n_1 n_2}\right)} = \\ &= \sqrt{\frac{\sum(x_i - \bar{x}_1)^2 + \sum(x_i - \bar{x}_2)^2}{n_1 + n_2 - 2}\left(\frac{n_1 + n_2}{n_1 n_2}\right)}. \end{aligned} \tag{2.5}$$

Следует заметить, что вышеизложенное применение t-критерия предполагает, что дисперсии сравниваемых групп одинаковы: $\sigma_1^2 = \sigma_2^2$. Если это не так, то величину критерия находят по другим формулам [15].

Неопровержение H_0-гипотезы нельзя рассматривать как доказательство равенства между неизвестными параметрами совокупностей, из которых извлечены сравниваемые выборки. В таких случаях вопрос о преимуществе одной статистической совокупности перед другой остается открытым. Ведь не исключено, что при повторных испытаниях H_0-гипотеза может оказаться несостоятельной. Более того, и в тех случаях, когда H_0-гипотеза опровергается, не следует спешить с окончательным выводом.

Правильное применение t-критерия предполагает нормальное распределение совокупностей, из которых извлечены сравниваемые выборки, и равенство генеральных дисперсий. Если эти условия не выполняются, то t-критерий применять не следует. В таких случаях более эффективными будут непараметрические критерии (см. раздел 2.4).

Пример 2.1

В двух группах испытуемых – опытной ($n_1 = 9$) и контрольной ($n_2 = 11$) – изучали умение ориентироваться в новой обстановке. После месячных курсов показатели коммуникативности, выраженные в баллах, варьировали следующим образом:

Группа	Значения показателя коммуникативности											$\bar{x}$
опытная	80	76	75	64	70	68	72	79	83	–	–	**74,1**
контрольная	70	78	60	80	62	68	73	60	71	66	69	**68,8**

Необходимо выяснить отличается ли среднее опытной группы от среднего контрольной? Сформулируем нулевую гипотезу H_0: $\bar{x}_1 = \bar{x}_2$. Напомним, что нулевую гипотезу выдвигают для того, чтобы попытаться ее опровергнуть.

Хотя t-критерий и позволяет сравнить средние значения, графический вывод также может оказаться полезным, например, для того, чтобы определить, различаются ли разбросы в двух группах (т. е. для проверки предположения о равенстве дисперсий) и есть ли выбросы (крайние значения), влияющие на среднее и увеличивающее стандартное отклонение (рис. 2.3).

Вертикальная линия внутри каждого ящичка соответствует медиане значений показателя. Хорошо было бы, если для исследуемых данных медиана располагалась примерно в центре каждого ящичка и чтобы разбросы ящичков (50 % каждого распределения) были похожи.

В нашем примере оба показателя содержат выбросы, которые приводят к тому, что средние оказываются ближе к друг другу, чем медианы, что, в свою очередь, может помешать выявлению существенных различий. Не исключена и противоположная ошибка – наличие выбросов может привести к тому, что средние будут различаться сильнее, чем медианы, и результаты будут казаться более значимыми.

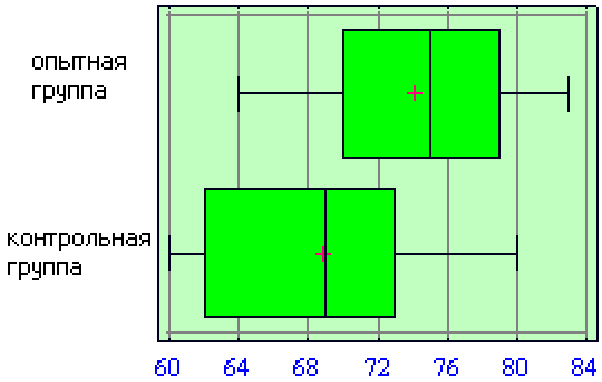


Рис. 2.3. Ящичковые диаграммы разброса распределений в опытной и контрольной группах

Активизировав прикладной статистический пакет SPSS, выберите из меню *Analyze – Compare Means – Independent – Samples T Test* и задайте необходимые параметры. Результаты сравнения представлены в табл. 2.3.

Таблица 2.3

Сравнение средних для независимых выборок

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95 % Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	0,005	0,942	1,829	18	0,084	5,29	2,89	–0,79	11,37
Equal variances not assumed	–	–	1,844	17,687	0,082	5,29	2,87	–0,75	11,33

Когда *SPSS* выводит на экран таблицу критерия для независимых выборок, она по умолчанию расщепляется на две части. Сначала проверяется гипотеза о равенстве дисперсий (*Levene Test for Equality of Variances*) посредством критерия Ливиня, а далее сравниваются средние. В процедуре сравнения средних вычисляет два критерия. Первый предполагает, что дисперсии совокупностей, из которых набираются члены групп, равны, т. е. их распределения имеют одинаковую форму (*Equal variances assumed*). Второй критерий не требует равенства дисперсий (*Equal variances not assumed*).

Критерий Ливиня определяет, различается ли разброс в разных группах. Нулевая гипотеза, в данном случае, это гипотеза о том, что дисперсии двух совокупностей равны. В таблице уровень значимости (*Sig.)* равен 0,942, поэтому будем работать с полученными значениями расположенными в верхней строке (*Equal variances assumed*).

Из табл. 2.3 находим, что *t*-статистика равна 1,829 с 18 степенями свободы (df) и ей соответствует вероятность (*Significance*) 0,084. Итак, мы можем заключить, что средние значения показателя коммуникативности в опытной и контрольной группе достоверно не различаются, т. е. испытуемые обеих групп в новой обстановке ориентируются без значимых отличий.

В табл. 2.3 также даны: значения разности средних (*Mean Difference*), стандартная ошибка разности (*Std. Error Difference*), которая используются при построении доверительного интервала для разности средних, и 95 % доверительный интервал среднего.

Рассмотрим реализацию расчета *t*-критерия Стьюдента на интернет-ресурсе http://www.psychol-ok.ru/. После ввода выборок необходимо указать, что они не связанные. Результат выполнения представлен ниже:

Таблица 2.4

Расчет отклонений для заданных выборок

№	Выборки		Отклонения от среднего		Квадраты отклонений	
	В.1	В.2	В.1	В.2	В.1	В.2
1	80	70	5,89	1,18	34,6921	1,3924
2	76	78	1,89	9,18	3,521	84,2724
3	75	60	0,89	–8,82	0,7921	77,7924
4	64	80	–10,11	11,18	102,2121	124,9924
5	70	62	–4,11	–6,82	16,8921	46,5124
6	68	68	–6,11	–0,81999999999999	37,3321	0,6724
7	72	73	–2,11	4,18	4,4521	17,4724
8	79	6	4,89	–8,82	23,9121	77,7924

Окончание табл. 2.4

№	Выборки		Отклонения от среднего		Квадраты отклонений	
	В.1	В.2	В.1	В.2	В.1	В.2
9	83	71	8,89	2,18	79,0321	4,7524
10		66		–2,82		7,9524
11		69		0,18000000000001		0,0324
Суммы:	667	757	0,01	–0,02	302,8889	443,6364
Среднее:	74,11	68,82				

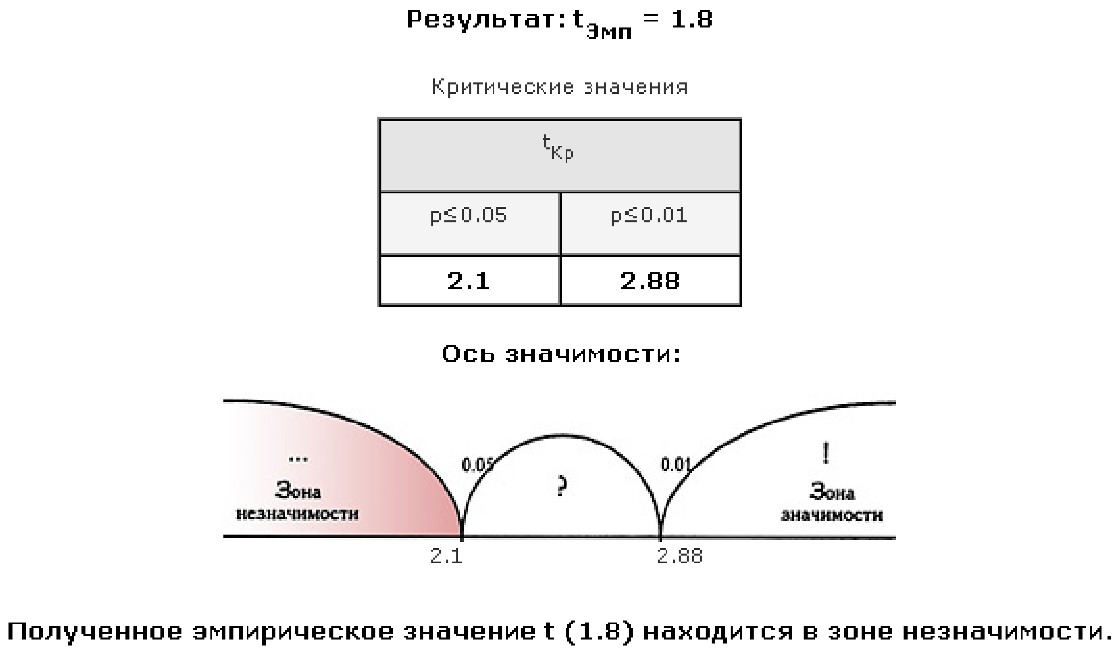


Рис. 2.4. Результат вычисления критерия Стьюдента

2.2.3. Оценка разности средних для зависимых выборок

Этот критерий применяют, когда некоторые показатели одного и того же объекта измеряются дважды: до и после некоторого вида обработки (например, лечения или обучения). Этот критерий подходит также для пар, в которых объекты подбираются по переменной, имеющей отношение к исследуемому измерению. Например, проверяя новый метод обучения иностранному языку, вы должны каждому студенту, обучающемуся новым способом, поставить в соответствие студента, занимающегося по традиционной программе.

В таких случаях *средняя разность* определяется по формуле [15]

$$\bar{d} = \frac{\sum d_i}{n} = \left(\bar{x}_1 - \bar{x}_2\right), \tag{2.6}$$

где $d_i = x_{1i} - x_{2i}$.

Ошибку средней разности $\overline{d}$, обозначаемую символом s_d, определяют по формулам:

$$s_d = \sqrt{\frac{\sum\left(d_i - \overline{d}\right)^2}{n(n-1)}} = \sqrt{\frac{\sum d_i^2 - \frac{\left(\sum d_i\right)^2}{n}}{n(n-1)}} \tag{2.7}$$

или

$$s_d = \sqrt{\frac{1}{n-1}\left(\frac{\sum d_i^2}{n} - \overline{d}^2\right)}. \tag{2.8}$$

H_0-гипотеза сводится к предположению, что $\overline{d} = 0$. t-критерий выразится в виде отношения средней разности к своей ошибке, т. е. $t = \overline{d} / s_{\overline{d}}$. Если $t_{эмп} \geq t_{кр}$ (для принятого уровня значимости и числа степеней свободы $k = n - 1$), то нулевая гипотеза должна быть отвергнута.

Пример 2.2

На протяжении ряда лет в школе изучали влияние учебной нагрузки на уровень тревожности учеников младших классов. В табл. 2.5 приведены выборки с попарно связанными вариантами. Поэтому обрабатывать полученные данные нужно с учетом тех условий, в которых проводили эксперимент. Из табл. 2.5 видно, что уровень тревожности в конце учебного года несколько выше, чем в начале года.

Таблица 2.5

Процент школьников, у которых наблюдался высокий уровень тревожности

	1993	1994	1995	1996	1997	1998
В начале учебного года	31,1	24,0	24,6	28,6	29,1	30,1
В конце учебного года	31,6	24,2	24,8	29,1	29,9	31,0

Воспользуемся SPSS для сравнения средних значений показателей тревожности в этих двух выборках. Для этого необходимо выбрать из меню *Analyze – Compare Means – Paired – Samples T Test.* Поместите в список исследуемые переменные. На экране отобразится окно со следующими таблицами (табл. 2.6, табл. 2.7).

Таблица 2.6

Статистика связанных выборок (Paired Samples Statistics)

	Mean	N	Std. Deviation	Std. Error Mean
VAR1	27,9167	6	2,9363	1,1987
VAR2	28,4333	6	3,1728	1,2953

Для данных выборок получаем, что у школьников в конце года уровень тревожности несколько выше (28,4), чем в начале года (27,9).

Таблица 2.7

t-критерий связанных выборок (Paired Samples Test)

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95 % Confidence Interval of the Difference				
				Lower	Upper			
VAR1 – VAR2	0,5167	0,2927	0,1195	0,2095	0,8238	4,324	5	0,008

Средняя разность $\bar{d} = 0{,}52$. 95%-й доверительный интервал разности от 0,21 до 0,82. Значение *p* для *t* – статистики, равной 4,32, очень мало (меньше 0,0005), откуда следует, что разница в 0,52 статистически значима. Следовательно, можно сделать вывод, что учебная нагрузка оказывает влияние на уровень тревожности учеников младших классов.

2.2.4. Оценка разности между долями

Выборочная доля зависит от числа объектов в выборке, имеющих учитываемый признак. Общее число таких объектов в генеральной совокупности определяет *генеральную долю P*. Оценкой разности между генеральными долями $P_1 - P_2 = D$ служит разность между выборочными долями $p_1 - p_2 = d$. Отношение этой разности к своей ошибке дает случайную величину $t = d / s_{d_p}$, которая подчиняется *t*-распределению Стьюдента. H_0-гипотезу, или предположение о том, что $P_1 = P_2$, отвергают, если $t_{эмп} \geq t_{кр}$ для степеней свободы, равных $k = n1 + n2 - 2$ и принятого уровня значимости α.

Ошибка разности между долями, взятыми из приблизительно равновеликих выборок (когда численности групп различаются не более чем на 25 %), вычисляют по формуле [15]:

$$s_{d_p} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}, \tag{2.9}$$

где $q = 1 - p$.

Если доли выражены в процентах от общего числа наблюдений, ошибку разности между ними определяют по формуле [15]:

$$s_{d_p} = \sqrt{\frac{p_1(100-p_1)}{n_1} + \frac{p_2(100-p_2)}{n_2}}. \tag{2.10}$$

Сопоставляемые группы n_1 и n_2 могут быть выражены абсолютными числами m_1 и m_2. Ошибка наблюдаемой между ними разности определяется по следующей формуле [15]:

$$s_{d_p} = \sqrt{\frac{m_1(n_1 - m_1)}{n_1} + \frac{m_2(n_2 - m_2)}{n_2}}, \tag{2.11}$$

но так как $m_1/n_1 = p_1$; $m_2/n_2 = p_2$; $(n_1 - m_1)/n_1 = q_1$; $(n_2 - m_2)/n_2 = q_2$, то формулу (2.11) можно представить и в таком виде [15]:

$$s_{d_p} = \sqrt{n_1 p_1 (1 - p_1) + n_2 p_2 (1 - p_2)} = \sqrt{n_1 p_1 q_1 + n_2 p_2 q_2}. \tag{2.12}$$

Когда сравнивают доли из неравновеликих выборок и при $75\ \% < p < 25\ \%$, ошибку разности между ними определяют по формуле [15]:

$$s_{d_p} = \sqrt{p(1-p)\frac{n_1 + n_2}{n_1 n_2}} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}; \tag{2.13}$$

p определяют как средневзвешенную из p_1 и p_2 долей, или же из абсолютных численностей групп:

$$p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}. \tag{2.14}$$

В этих формулах n_1 и n_2 – численности групп, на которых определяют доли $m_1/n_1 = p_1$, $m_2/n_2 = p_2$. Если доли выражают в процентах от n, то вместо $q = 1 - p$ нужно брать $q = 100 - p$. Если же неравновеликие группы выражены абсолютными числами m_1 и m_2, ошибку разности между ними определяют по формуле

$$s_{d_p} = \sqrt{\frac{m_1 + m_2}{n_1 + n_2}\left(1 - \frac{m_1 + m_2}{n_1 + n_2}\right)\frac{n_1 + n_2}{n_1 n_2}}. \tag{2.15}$$

Пример 2.3

Рассматривались две группы студентов: с высокой и низкой успеваемостью. Было установлено, что в первой группе – 14 холостых и 9 женатых студентов, а во второй группе – 12 холостых и 8 женатых студентов. Разница между количеством холостых студентов в 1-й и 2-й группе составила 14 – 12 = 2.

Для того чтобы оценить различие полученных долей воспользуемся пакетом *STATGRAPHICS*. Выберите в меню *Compare – Two Samples – Hypothesis Tests…* (рис. 2.5).

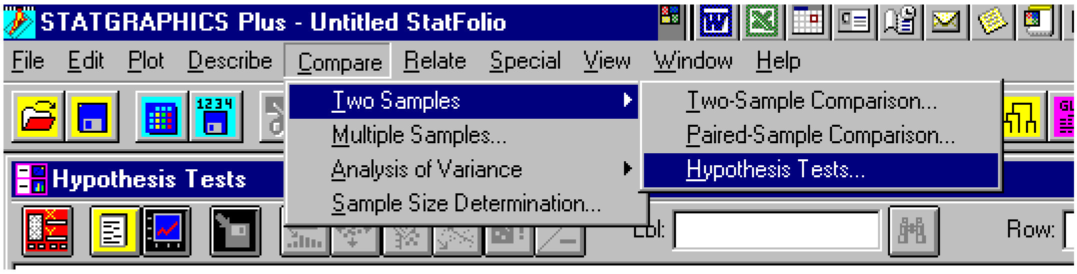


Рис. 2.5. Выбор процедуры сравнения между долями

Появится окно *Hypothesis Tests* для задания анализируемых переменных. В окне диалога активизируем в разделе *Compare* пункт *Poisson Rates*, и далее необходимо задать значения доли для первой выборки в поле *Sample 1 Rate*, а поле *Sample 2 Rate* – значение доли второй выборки (рис. 2.6), завершив ввод информации нажатием кнопки *OK*. Если доли выражены в процентах от общего числа наблюдений, то проценты заносятся также в поле *Sample Rate*. Размер выборок вводится в поле *Sample Size*.

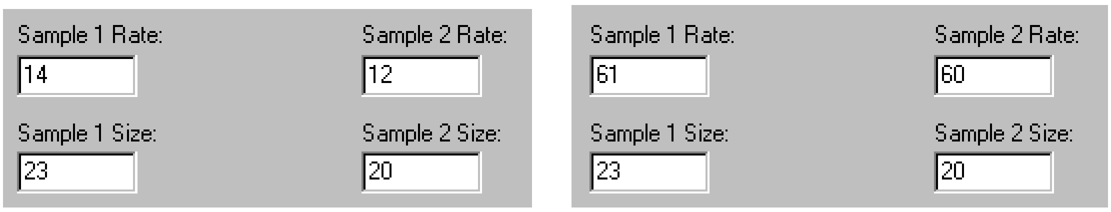


Рис. 2.6. Ввод данных:ввод количественной доли (слева);
доля в процентах(справа)

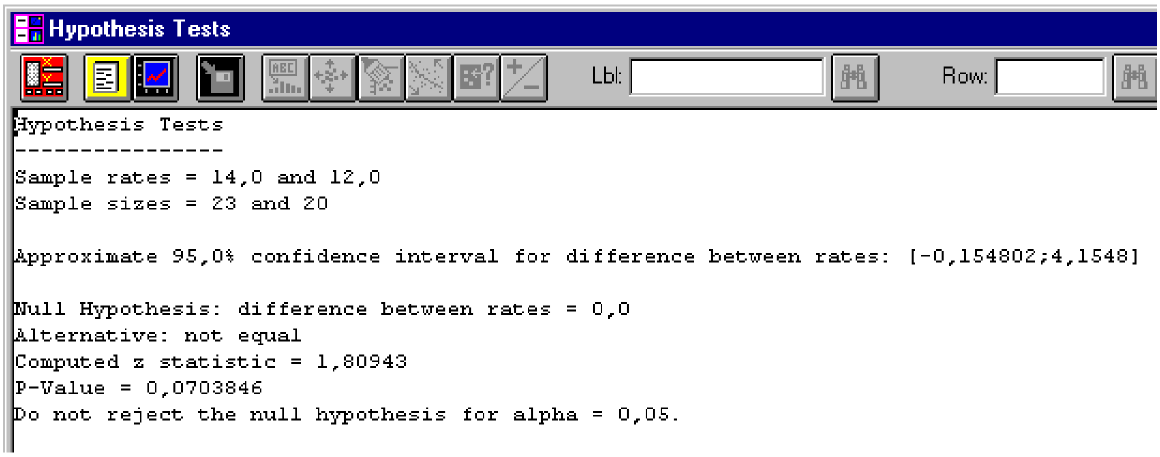


Рис. 2.7. Сводка проведенного сравнения разницы между долями

После того как мы нажали кнопку *OK* на экран выдается сводка проведенного сравнения (рис. 2.7). Из представленной сводки мы получаем сведения: *z* – статистику (1.80943) и соответствующий ей уровень значимости равный 0,07, а также сведения о том, что нельзя отклонить нулевую гипотезу о равенстве исследуемых долей: разница между количеством холостых студентов в разных группах оказалась статистически недостоверной.

2.2.5. Оценка разности между коэффициентами вариации

Разность между коэффициентами вариации сравниваемых групп, извлеченных из нормально распределяющихся совокупностей, можно оценить с помощью t-критерия Стьюдента. Приближенной оценкой разности $Cv_1 - Cv_2 = d_{Cv}$ служит ее отношение к своей ошибке, которая равна корню квадратному из суммы ошибок коэффициентов вариации сравниваемых групп, т. е.

$$s_{d_{Cv}} = \sqrt{s_{Cv_1}^2 + s_{Cv_2}^2}\,, \tag{2.16}$$

$$s_{Cv} = \sqrt{\frac{(cv)^2}{n-1}\left[0{,}5 + \left(\frac{cv}{100}\right)^2\right]}\,, \tag{2.17}$$

Нулевую гипотезу отвергают, если $t_{эмп} > t_{кр}$ для принятого уровня значимости и числа степеней свободы $k = n_1 + n_2 - 2$.

Пример 2.4

Экспериментальная ($n_1 = 8$) и контрольная ($n_2 = 9$) группы учеников 2-го класса характеризуются следующими средними показателями скорости чтения (количество слов в минуту): $\bar{x}_1 = 63{,}8$ и $\bar{x}_2 = 52{,}6$. Соответственно, $s_1 = \sqrt{25{,}963} = 5{,}095$ и $s_2 = \sqrt{35{,}97} = 5{,}997$. Отсюда

$$Cv_1 = 100(5{,}095/63{,}8) = 8{,}0\ \%;$$

$$Cv_2 = 100(5{,}997/52{,}6) = 11{,}4\ \%.$$

Сформулируем нулевую гипотезу H_0: разность между коэффициентами вариации показателя скорости чтения в экспериментальной и контрольной группе учеников равна нулю.

Разница $d_{Cv} = 11{,}4 - 8{,}0 = 3{,}4\ \%$.

Определяем ошибки этих показателей по формуле (2.17):

$$s_{Cv_1} = \sqrt{\frac{(8{,}0)^2}{8-1}\left[0{,}5 + \left(\frac{8{,}0}{100}\right)^2\right]} = \sqrt{9{,}143 \cdot 0{,}506} = \sqrt{4{,}63} = 2{,}15;$$

$$s_{Cv_2} = \sqrt{\frac{(11{,}4)^2}{9-1}\left[0{,}5 + \left(\frac{11{,}4}{100}\right)^2\right]} = \sqrt{16{,}245 \cdot 0{,}513} = \sqrt{5{,}04} = 2{,}24\,.$$

Ошибка разности $s_{d_{Cv}} = \sqrt{(2{,}15)^2 + (2{,}24)^2} = \sqrt{9{,}64} \approx 3{,}11$.

$$t_{эмп} = 3{,}40/3{,}11 = 1{,}09.$$

Эта величина не превосходит критическую точку $t_{кр}$ = 2,13 для $k = 8 + 9 - 2 = 15$ и $\alpha = 5$ %, что не дает основания для отвержения нулевой гипотезы, т. е. не выявлены существенные различия между коэффициентами вариации.

2.3. Проверка статистической гипотезы о нормальном распределении

2.3.1. Применение коэффициентов асимметрии и эксцесса для проверки нормальности распределения

Эмпирический вариационный ряд и его график – вариационная кривая – не позволяют с полной уверенностью судить о законе распределения совокупности, из которой взята выборка. На величине любого варьирующего признака сказывается влияние многочисленных, в том числе и случайных, факторов, искажающих четкую картину варьирования. Между тем знание закона распределения позволяет избежать возможных ошибок в оценке генеральных параметров по выборочным характеристикам.

Гипотезу о законе распределения можно проверить разными способами, в частности с помощью *коэффициентов асимметрии As* и *эксцесса Ex*. При нормальном распределении эти показатели равны нулю. В действительности такое равенство почти не наблюдается. Выборочные показатели *As* и *Ex*, определяемые по формулам (1.10) и (1.13), являются случайными величинами, которые сопровождаются ошибками репрезентативности.

$$s_{As} = \sqrt{\frac{6}{n+3}}; \tag{2.18}$$

$$s_{Ex} = \sqrt{\frac{24}{n+5}} = 2\sqrt{\frac{6}{n+5}}. \tag{2.19}$$

В качестве критерия нормальности распределения служат t_{As} и t_{Ex}, являющиеся отношениями выборочных коэффициентов *As* и *Ex* к их ошибкам репрезентативности, которые определяют обычно по следующим приближенным формулам:

$$t_{As} = \frac{|As|}{s_{As}} \geq 3;\ t_{Ex} = \frac{|Ex|}{s_{Ex}} \geq 3. \tag{2.20}$$

Показатели асимметрии и эксцесса свидетельствуют о достоверном отличии эмпирических распределений от нормального в том случае, если они превышают по абсолютной величине свою ошибку репрезентативности в 3 и более раз.

В связи с тем, что выборочные распределения коэффициентов асимметрии и эксцесса в случае нормальности распределения признака при не слишком больших объемах выборок (особенно это характерно для *Ex*) могут быть довольно далеки от нормального вида, использование квадратических ошибок (ошибок репрезентативности) для *As* и *Ex* при n, меньшем нескольких сотен наблюдений, оказывается рискованным. Так как согласно [15], значения t_{As} и t_{Ex} есть не что иное, как эмпирические значения t-критерия Стьюдента для проверки гипотез о равенстве нулю асимметрии и эксцесса, то более предпочтительным при малом объеме исследуемых выборок следует считать проверку нормальности распределения по значениям этих коэффициентов с применением статистических таблиц (см. [6, 24]). В них указаны критические точки для разных уровней значимости α и объемов выборки n. Если коэффициенты *As* и *Ex* превосходят критические точки, содержащиеся в этих таблицах, гипотеза о нормальности распределения должна быть отвергнута или если

$$t_{As} < t_{\text{кр}} \text{ и } t_{Ex} < t_{\text{кр}},$$

то мы принимаем гипотезу H_0, т. е. распределение является нормальным.

Можно использовать и другую схему проверки по формулам, представленным в [23]:

1. Определяются эмпирические показатели асимметрии и эксцесса по формулам (1.10) и (1.13).
2. Рассчитываются критические значения показателей асимметрии и эксцесса по формулам Е.И. Пустыльника [23]:

$$As_{\text{кр}} = 3\sqrt{\frac{6\cdot(n-1)}{(n+1)(n+3)}}; \tag{2.21}$$

$$Ex_{\text{кр}} = 5\sqrt{\frac{24\cdot n\cdot(n-2)(n-3)}{(n+1)^2\cdot(n+3)(n+5)}}. \tag{2.22}$$

3. Сопоставляются эмпирические значения с критическими. Если эмпирические значения показателей ниже критических, делается вывод о том, что распределение признака не отличается от нормального:

$$As_{\text{эмп}} < As_{\text{кр}};\ Ex_{\text{эмп}} < Ex_{\text{кр}}. \tag{2.23}$$

Например, при изучении формы распределения были получены значения $As = -0{,}556$ и $Ex = 0{,}872$. Для $\alpha = 1$ % и $n = 200$ табличные значения $As_{\text{кр}} = 0{,}403$ и $Ex_{\text{кр}} = 0{,}632$. Так как эмпирически определенные величины

As и *Ex* превышают табличные критические значения, можно сделать вывод о наличии у этого распределения значимых асимметрии и эксцесса.

Пример 2.5

В табл. 2.8 приведены значения показателя длительности попыток решения анаграмм для 16 испытуемых. Необходимо проверить гипотезу о нормальности распределения данного показателя.

Таблица 2.8

Показатель длительности попыток решения анаграмм [17]

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_i	11	13	12	9	10	11	8	10	15	14	8	7	10	10	5	8

По статистической таблице критических значений t-критерия Стьюдента [6, 24] $t_{\text{кр}}$ для $\nu = 16 - 2 = 14$, уровня значимости $p = 0{,}05$: $t_{\text{кр}} = 2{,}14$. Вычислим среднее арифметическое $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{161}{16} = 10{,}06$ и стандартное отклонение:

$$\sigma = \sqrt{\frac{\sum\left(x_i - \bar{x}\right)}{n-1}} = \sqrt{\frac{102{,}944}{16-1}} = \sqrt{6{,}893} = 2{,}62.$$

Показатели асимметрии и эксцесса с их ошибками репрезентативности:

$$As = \frac{\sum(x_i - \bar{x})^3}{n \cdot \sigma^3} = \frac{+30{,}468}{16 \cdot 2{,}62^3} = +0{,}106;$$

$$s_{As} = \sqrt{\frac{6}{n+3}} = \sqrt{\frac{6}{19}} = 0{,}56;$$

$$Ex = \frac{\sum(x_i - \bar{x})^4}{n \cdot \sigma^4} - 3 = \frac{1725{,}467}{16 \cdot 2{,}62^4} - 3 = -0{,}711;$$

$$s_{Ex} = 2 \cdot \sqrt{\frac{6}{n+5}} = 2 \cdot \sqrt{\frac{6}{21}} = 1{,}07.$$

$$t_{As} = \frac{|0{,}106|}{0{,}56} = 0{,}189; \; t_{Ex} = \frac{|-0{,}711|}{1{,}07} = 0{,}475.$$

Оба показателя не превышают в три раза свою ошибку репрезентативности, из чего мы можем заключить, что распределение данного признака не отличается от нормального.

Произведем проверку другим способом (по формулам Е.И. Пустыльника). Рассчитаем критические значения для показателей *As* и *Ex*:

$$As_{\text{кр}} = 3\sqrt{\frac{6\cdot(16-1)}{(16+1)(16+3)}} = 3\sqrt{\frac{90}{323}} = 1{,}58;$$

$$Ex_{\text{кр}} = 5\sqrt{\frac{24\cdot 16\cdot(16-2)(16-3)}{(16+1)^2\cdot(16+3)(16+5)}} = 5\sqrt{\frac{69888}{115311}} = 3{,}89;$$

$As_{\text{эмп}} = 0{,}106$; $Ex_{\text{эмп}} = -0{,}711$;

$As_{\text{эмп}} < As_{\text{кр}}$; $Ex_{\text{эмп}} < Ex_{\text{кр}}$.

Поскольку эмпирические значения показателей ниже критических, делается вывод о том, что распределение признака не отличается от нормального.

Проверка гипотезы о нормальном законе распределения исследуемой выборки с помощью t-критерия Стьюдента в статистическом пакете *STATGRAPHICS* рассмотрена ниже.

2.3.2. Критерий хи-квадрат (χ^2-распределение)

Проверку гипотез о нормальном законе распределения также производят с помощью других критериев. Один из них, нашедший широкое применение в педагогических, психологических и медицинских исследованиях, – *критерий согласия*, или *соответствия* χ^2 (хи-квадрат). Величина критерия χ^2 определяется по формуле

$$\chi^2 = \sum_{i=1}^{k}\frac{(f-f')^2}{f'} = \sum_{i=1}^{k}\left(\frac{d^2}{f'}\right), \tag{2.24}$$

где f – эмпирические частоты, f' – ожидаемые частоты, d – разность между эмпирическими и вычисленными частотами.

Величина критерия χ^2 всегда положительна. При полном совпадении эмпирических частот с вычисленными или ожидаемыми частотами $\sum(f_i - f_i')^2 = 0$ и $\chi^2 = 0$.

Распределение вероятных значений случайной величины χ^2 является непрерывным и асимметричным (рис. 2.8), оно зависит от числа степеней свободы k и приближается к нормальной кривой по мере увеличения числа испытаний n. Поэтому применение критерия χ^2 к оценке дискретных распределений сопряжено с некоторыми погрешностями, которые сказываются на его величине, особенно при малых выборках.

Число степеней свободы устанавливают по вторичному числу классов с учетом ограничений свободы вариации, которая в разных случаях бывает различной. Так, при оценке эмпирических распределений, следующих нормальному закону, число степеней свободы $v = n - 3$ (с учетом трех ограничений свободы вариации: n, $\bar{x}$ и s_x).

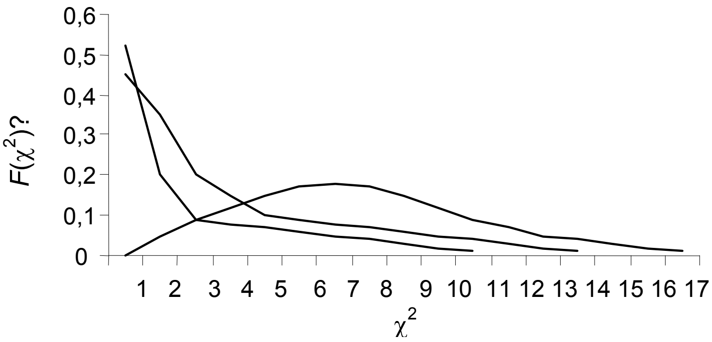


Рис. 2.8. Функция χ^2-распределения в зависимости от разных чисел степени свободы k

На величине критерия χ^2 сказывается степень точности, с какой определены теоретически вычисленные или ожидаемые частоты. Поэтому при сопоставлении эмпирических частот с вычисленными частотами последние не следует округлять до целых чисел.

Нулевая гипотеза сводится к предположению, что различия, наблюдаемые между эмпирическими и вычисленными или ожидаемыми частотами, носят исключительно случайный характер. Для проверки нулевой гипотезы нужно фактически полученную величину $\chi^2_{\text{эмп}}$ сравнить с ее критическим значением $\chi^2_{\text{кр}}$. Если $\chi^2_{\text{эмп}} \geq \chi^2_{\text{кр}}$, то нулевая гипотеза должна быть отвергнута на принятом уровне значимости с числом степеней свободы v. Критические точки $\chi^2_{\text{кр}}$ приведены в таблицах [6, 24].

Пример 2.6 [15]

В табл. 2.9 приведены эмпирические и вычисленные по нормальному закону частоты распределения длины тела у 267 мужчин. Из приведенных данных видно, что между эмпирическими и вычисленными частотами нет полного совпадения. Нужно установить, случайны или закономерны эти различия, т. е. выяснить, следует ли это распределение нормальному закону. Расчет χ^2-критерия, который оказался равным 1,47, приведен в табл. 2.9.

В данном случае число вторичных классов $N = 7$. Число степеней свободы $k = 7 - 3 = 4$. Исходя из 5%-го уровня значимости $\chi^2_{\text{кр}} = 9{,}49$. Эта величина значительно превышает $\chi^2_{\text{эмп}} = 1{,}47$, что не позволяет отвергнуть H_0-гипотезу. Следовательно, существуют достаточные основания для утверждения, что данное распределение следует нормальному закону.

Таблица 2.9

Частоты		$d = f - f'$	d^2	d^2/f'
Эмпирические f	Вычисленные f'			
1	2	3	4	5
12	11,6	0,4	0,16	0,01
31	34,3	3,3	10,89	0,32
71	67,8	3,2	10,24	0,15
82	77,6	4,4	19,36	0,25
46	51,2	5,2	27,04	0,53
19	19,5	0,5	0,25	0,01
6	5,0	1,0	1,00	0,20
$\Sigma = 267$	$\Sigma = 267$	–	–	1,47

Критерий χ^2 применяют и для оценки сходства между вариационными рядами, частоты которых распределяются в границах одних и тех же классов. В таких случаях критерий χ^2 определяют по формулам:

при $n_1 = n_2$
$$\chi^2 = 4\left(\sum_{i=1}^{k} \frac{f_1^2}{f_1 + f_2}\right) - N; \tag{2.25}$$

при $n_1 \neq n_2$
$$\chi^2 = \frac{N^2}{n_1 n_2}\left(\sum_{i=1}^{k} \frac{f_1^2}{f_1 + f_2} - \frac{n^2}{N}\right). \tag{2.26}$$

В этих формулах f_1 и f_2 – частоты сравниваемых распределений; n_1 – объем одного (любого), а n_2 – объем другого ряда распределения; $N = n_1 + n_2$. Число степеней свободы k определяют по числу классов N без единицы, т. е. $k = N - 1$. При этом частоты, которые меньше 5, не объединяют, как это принято в отношении теоретически вычисленных частот.

Как правило, статистические пакеты обязательно содержат процедуры проверки гипотез о нормальном законе распределения экспериментальных выборок. В качестве примера был выбран статистический пакет STATGRAPHICS.

Для того чтобы выполнить проверку на нормальность, выберите в главном меню пункт *Describe – Distribution Fitting – Uncensored Data* (рис. 2.9).

Далее, необходимо ввести имя признака и нажать кнопку *Ok*. Выбрав в верхней части рабочего поля кнопку с табличными опциями  (*Tabular options*) и отметив галочкой пункт *Tests for Normality*, завершить выбор нажатием кнопки *Ok*. Активизируем окно двойным нажатием левой кнопки мыши. Таким образом, мы получили возможность проверить гипотезу о нормальном распределении значений исследуемой выборки не только с помощью критерия хи-квадрат, но и критерия Шапиро–Уилкса (для малых выборок), а также с помощью оценки параметров коэффициента асимметрии и эксцесса.

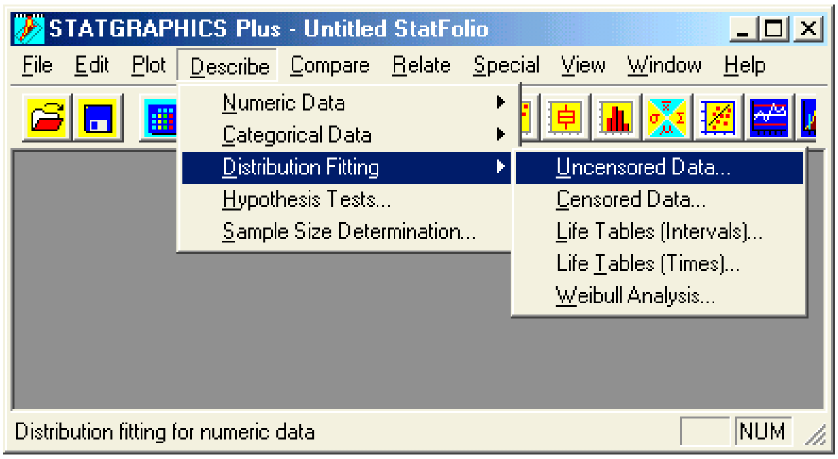


Рис. 2.9. Выбор процедуры проверки на нормальность исследуемой выборки

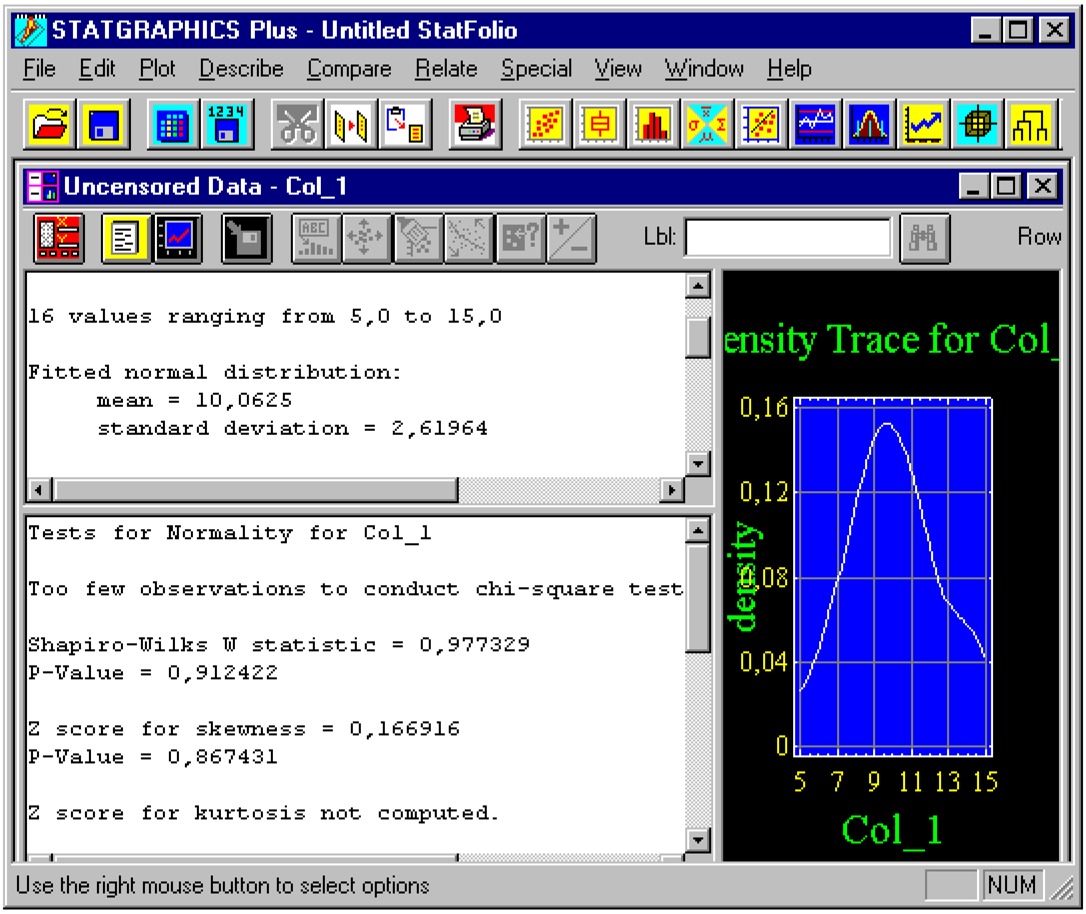
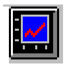


Рис. 2.10. Результаты проверки на нормальность исследуемой выборки

Для того чтобы построить график распределения исследуемой выборки требуется активизировать пиктограмму графических опций  (*Gaphics Options*) и отметить пункт *Density Trace***.**

На рис. 2.10. представлены результаты проверки на нормальность распределения значений показателя длительности попыток решения анаграмм для 16 испытуемых (см. табл. 2.8). Проверку нормальности критерием хи-квадрат выполнить не удалось, так как в выборке недостаточное количество объектов исследования. Тем не менее, проверка на нормальность была выполнена критерием Шапиро–Уилкса, который работает и на малых выборках. Также выводятся значения ***Z***-статистики асимметрии и эксцесса (для данного примера ***Z***-статистика эксцесса не вычислена) и уровни значимости, по значениям которых можно или принять гипотезу о нормальности распределения, или отклонить. В рассмотренном примере уровни значимости $\boldsymbol{p} > 0{,}05$, и, следовательно, нельзя отклонить гипотезу о нормальности распределения. Построенный график на рис. 2.10 подтверждает правильность полученных результатов.

2.4. Непараметрические критерии проверки статистических гипотез

2.4.1. Основные понятия о непараметрических критериях

Рассматривая принадлежность двух выборок к данной генеральной совокупности, можно высказать две гипотезы [15]:

- выборки принадлежат к одной генеральной совокупности, и, значит, различия между ними не являются существенными;
- выборки принадлежат к разным генеральным совокупностям.

При оценке достоверности различий выборок принято обозначать величиной *p* вероятность ошибочного принятия нулевой гипотезы, т. е. гипотезы об отсутствии существенных различий, когда в действительности такие различия есть. Если вероятность принятия нулевой гипотезы достаточно мала, делают вывод о статистической значимости различий. Для большинства медико-психологических исследований принято считать, что в случае, если *p* не превышает 0,05, наблюдаемые различия неслучайны.

В большинстве медицинских исследований для оценки существенности различий используют главным образом параметрический *t*-критерий Стьюдента, который основан на предположении, что сравниваемые выборки принадлежат нормальным распределениям. Между тем в психологических и педагогических исследованиях распределения могут значительно отличаться от нормального. В этих случаях и даже тогда, когда просто неизвестно, являются ли распределения нормальными, применение *t*-критерия, строго говоря, является необоснованным и может привести к ошибочным заключениям. Поэтому все большее распространение получают непараметрические критерии различий, не зависящие от формы распределений. Название «непараметрические» связано с тем, что эти критерии не требуют вычисления параметров известных распределений, которыми аппроксимируются исследуемые выборки, как это требуется при использовании параметрических критериев.

Основные преимущества непараметрических критериев сводится к следующему. Во-первых, при распределениях, далеких от нормального, непараметрические критерии позволяют обнаружить существенные различия тогда, когда, например, t-критерий их не выявляет. Во-вторых, при распределениях, близких к нормальному, непараметрические критерии также дают хороший результат, хотя и уступают в этом случае t-критерию.

Область применения. Рассмотрим часто встречающуюся на практике задачу сравнения двух выборочных совокупностей.

В духе основной статистической предпосылки мы будем рассматривать эти совокупности как случайные. Например, нас может интересовать сравнение двух методов обработки, т. е. двух разных действий, направленных к одной цели: двух лекарств, двух рационов питания, двух методик обучения или профессиональной подготовки и т. д.

Данные. Для исследования нужны однородные объекты, разделенные на две группы. Взаимные влияния и взаимодействия объектов должны быть исключены. Для каждого объекта регистрируется некоторая его числовая характеристика. Возникающие при этом две группы чисел можно рассматривать как две независимые выборки.

Постановка задачи. Рассмотрим вопрос о том, какие задачи наиболее часто рассматриваются при сравнении двух выборок. Обычно две выборки получаются как результаты применения различных условий эксперимента к двум группам испытуемых, однородных по своему составу. Как отмечается в [31], изменение условий эксперимента обычно сказывается, прежде всего, на изменении положения распределения измеряемой числовой характеристики на числовой прямой. Масштаб и форма распределения при малых изменениях условий эксперимента обычно остаются практически неизменными. При больших различиях в условиях эксперимента наряду с изменением положения распределения изменяется и его разброс (дисперсия). И совсем редко происходит изменение самой формы распределения. Поэтому при исследовании различий в двух выборках часто предполагают, что законы распределения двух анализируемых выборок отличаются только сдвигом [24].

Как уже отмечалось выше, правильное применение параметрических критериев для проверки статистических гипотез основано на предположении о нормальном распределении совокупностей, из которых взяты сравниваемые выборки, что не всегда имеет место, так как не все психологические признаки распределяются нормально. Немаловажным является то обстоятельство, что исследователю приходится иметь дело не только с количественными, но и с качественными признаками, многие из которых выражаются порядковыми номерами, индексами и другими условными знаками. В таких случаях необходимо использовать непараметрические критерии.

2.4.2. Оценка достоверности сдвига в значениях исследуемого признака

В психологических исследованиях часто бывает важно доказать, что в результате действия каких-либо факторов произошли достоверные изменения («сдвиги») в измеряемых показателях. Это, например, фактор времени. Сопоставление показателей, полученных у одних и тех же испытуемых по одним и тем же методикам, но в разное время, дает нам *временной сдвиг.*

Сопоставление показателей, полученных по одним и тем же методикам, но в разных условиях измерения (например, «покоя» или «стресса»), дает нам *ситуационный сдвиг.*

Иногда создаются специальные экспериментальные условия, предположительно влияющие на те или иные показатели, и сопоставить замеры, произведенные до и после экспериментального воздействия. Если сдвиги окажутся статистически достоверными, то можно утверждать, что экспериментальные воздействия были существенными или эффективными.

Существуют также так называемые *структурные сдвиги.* Мы можем сопоставлять между собой разные показатели одних и тех же испытуемых, если они измерены в одних и тех же единицах, по одной и той же шкале. Например, мы можем исследовать различия между вербальным и невербальным интеллектом, измеренным по методике Д. Векслера, по методике Амтхауэра и т. д.

Часто возникает вопрос, должны ли мы всегда производить оба замера на одной и той же выборке, или «сдвиг» можно изучать на сходных, так называемых «уравновешенных» выборках, совпадающих друг с другом по полу, возрасту, профессии и другим значимым для исследователя характеристикам.

Согласно [24], допускается сопоставление показателей разных выборок, уравновешенных по всем значимым для исследования признакам. Опыт показывает, однако, что создать «уравновешенные» подгруппы практически невозможно. Это относится и к проблеме сопоставления экспериментальной и контрольной групп: мы почти никогда не можем быть уверены, что выявленные различия объясняются действием исследуемых факторов, а не различиями между двумя выборками [24].

В табл. 2.10 приведена классификация сдвигов и указаны статистические методы, позволяющие оценить их достоверность.

Таблица 2.10 [24]

Классификация сдвигов и критериев оценки их статистической достоверности

<table>
<tr><th rowspan="2">Виды сдвигов</th><th rowspan="2">Объект сопоставлений</th><th colspan="2">Условия</th><th rowspan="2">Критерии оценки достоверности сдвигов</th></tr>
<tr><th>Кол-во замеров</th><th>Кол-во групп</th></tr>
<tr><td>1. Временные, ситуационные, умозрительные, измерительные</td><td>Одни и те же показатели, измеренные у одних и тех же испытуемых в разное время, в разных ситуациях, в разных представляемых условиях или разными способами</td><td>2</td><td>1</td><td>Критерий знаков z;
T-критерий Уилкоксона</td></tr>
<tr><td rowspan="2">2. Сдвиги под влиянием экспериментальных воздействий</td><td>Одни и те же показатели, измеренные у одних и тех же испытуемых до и после воздействия:
а) при отсутствии контрольной группы</td><td>2</td><td>1</td><td>Критерий знаков z;
T-критерий Уилкоксона</td></tr>
<tr><td>б) при наличии контрольной группы</td><td>2</td><td>2</td><td>*Вариант 1*. Сопоставление значений «до» и «после» отдельно по экспериментальной и контрольной группам: критерий знаков z; T-критерий Уилкоксона.
Вариант 2. Сопоставление сдвигов в двух группах: Q-критерий; U-критерий; φ^*-критерий Фишера</td></tr>
<tr><td>3. Структурные сдвиги</td><td>Разные показатели одних и тех же испытуемых</td><td>2</td><td>1</td><td>Критерий знаков z;
T-критерий Уилкоксона</td></tr>
</table>

2.4.3. Критерий знаков *G*

Данный критерий применяется при сравнении двух выборочных совокупностей. Результаты наблюдений выражаются не числами, а знаками плюс (+) и минус (–).

Критерий знаков позволяет установить, в какую сторону в выборке в целом изменяются значения признака при воздействии каких-либо факторов: изменяются показатели в сторону улучшения, повышения или усиления или, наоборот, в сторону ухудшения, понижения или ослабления.

Ограничения критерия. Количество наблюдений в обоих замерах – не менее 5 и не более 300.

Суть метода заключается в следующем. Если попарно сравниваемые значения двух зависимых выборок существенно не отличаются друг от друга, то число «плюсов» и «минусов» окажется совершенно одинаковым; если же заметно преобладают «плюсы» и «минусы», это будет указывать на положительное или отрицательное действие изучаемого фактора на результативный признак. Большее число «плюсов» или «минусов» берется в качестве эмпирического (или фактического) значения *G*-критерия знаков ($G_{эмп}$). При этом не давшие ни положительного, ни отрицательного результата сравнения, обозначаемые цифрой 0, в расчет не принимаются, и число парных наблюдений соответственно уменьшается.

Как и всякий другой выборочный показатель, *G*-критерий знаков является величиной случайной; он служит для проверки H_0-гипотезы, т. е. предположения о том, что совокупность или совокупности, из которых взяты сравниваемые выборки, имеют одну и ту же или одинаковые функции распределения. H_0-гипотеза отвергается, если $G_{эмп} \leq G_{кр}$ для принятого уровня значимости p и числа парных наблюдений n, взятых без нулевых разностей. Критические точки $G_{кр}$ для двух уровней значимости и числа парных наблюдений содержатся в статистических таблицах (см. [6, 24]).

Пример использования критерия знаков *G* приводится ниже.

2.4.4. *T*-критерий Уилкоксона (Вилкоксона)

Различия между элементами сравниваемых зависимых выборок с достаточной точностью оцениваются с помощью рангового *T*-критерия Уилкоксона. *T*-критерий Уилкоксона является более мощным, чем критерий знаков *G*.

T-критерий рассчитывают следующим образом:

1. Ранжируют попарные разности, как положительные, так и отрицательные, в один общий ряд. При этом нулевые разности в расчет не

принимают, а все остальные независимо от знака ранжируют так, чтобы наименьшая абсолютная разность получила первый ранг, причем одинаковым по величине разностям присваивают один и тот же ранг.

2. Находят отдельно суммы положительных и отрицательных разностей. Меньшую из двух сумм разностей, без учета ее знака, используют в качестве фактически установленной величины *T*-критерия.
3. Сравнивают эту величину $T_{эмп}$ с критическим значением $T_{кр}$ для принятого уровня значимости α и числа парных наблюдений n, которое берут без нулевых разностей. Нулевая гипотеза отвергается, если $T_{эмп} < T_{кр}$. Критические значения парного критерия Уилкоксона $T_{кр}$ содержатся в таблицах (см. [6, 24]).

Пример 2.7 [24]

12 участников комплексной программы тренинга партнерского общения, продолжавшегося 7 дней, дважды оценивали у себя уровень владения коммуникативными навыками. Первое измерение производилось в первый день тренинга, второе – в последний. Все измерения производились по 10-балльной шкале. Данные представлены в табл. 2.11.

Таблица 2.11

Оценки уровней развития коммуникативных навыков

№	1 измерение			2 измерение		
	Активное слушание	Снижение эмоционального уровня	Аргументация	Активное слушание	Снижение эмоционального уровня	Аргументация
1	6	5	5	7	6	7
2	3	1	4	5	4	5
3	4	4	5	8	7	6
4	4	4	5	6	5	5
5	6	4	4	4	5	5
6	6	5	3	8	7	6
7	3	5	2	7	8	5
8	6	5	3	5	7	5
9	6	5	5	7	6	5
10	5	6	5	7	7	6
11	6	6	3	5	4	3
12	6	3	4	7	6	5

Ощущаются ли участниками значимые сдвиги в уровне владения каждым из трех навыков после тренинга?

Поскольку данные представлены по одной экспериментальной выборке, и было совершено 2 замера, мы должны выбирать между критерием знаков и критерием Уилкоксона. Сдвиги могут быть определены количественно, но они варьируют в достаточно узком диапазоне от –2 до +4. Учитывая это, применим последовательно оба критерия.

Решим данную задачу двумя способами:

1) с использованием критерия знаков,
2) с помощью критерия Уилкоксона.

Использование критерия знаков G

Преобразуя данные табл. 2.11, составим таблицу сдвигов (табл. 2.12), которая будет полезна при использовании обоих критериев. Для этого из значения, полученного во 2-м замере, вычтем значение, полученное данным испытуемым по данной шкале в 1-м замере.

Таблица 2.12

Сдвиги в уровне владения коммуникативными навыками «после»-«до»

<table>
<tr><th rowspan="2">№</th><th colspan="3">Сдвиги</th><th colspan="3">Ранги абсолютных величин сдвигов</th></tr>
<tr><th>Активное слушание</th><th>Снижение эмоционального уровня</th><th>Аргументация</th><th>Активное слушание</th><th>Снижение эмоционального уровня</th><th>Аргументация</th></tr>
<tr><td>1</td><td>1</td><td>1</td><td>2</td><td>3</td><td>3</td><td>6,5</td></tr>
<tr><td>2</td><td>2</td><td>3</td><td>1</td><td>8</td><td>10,5</td><td>3</td></tr>
<tr><td>3</td><td>4</td><td>3</td><td>1</td><td>11,5</td><td>10,5</td><td>3</td></tr>
<tr><td>4</td><td>2</td><td>1</td><td>0</td><td>8</td><td>3</td><td>–</td></tr>
<tr><td>5</td><td>–2</td><td>1</td><td>1</td><td>8</td><td>3</td><td>3</td></tr>
<tr><td>6</td><td>2</td><td>2</td><td>3</td><td>8</td><td>7</td><td>8,5</td></tr>
<tr><td>7</td><td>4</td><td>3</td><td>3</td><td>11,5</td><td>10,5</td><td>8,5</td></tr>
<tr><td>8</td><td>–1</td><td>2</td><td>2</td><td>3</td><td>7</td><td>6,5</td></tr>
<tr><td>9</td><td>1</td><td>1</td><td>0</td><td>3</td><td>3</td><td>–</td></tr>
<tr><td>10</td><td>2</td><td>1</td><td>1</td><td>8</td><td>3</td><td>3</td></tr>
<tr><td>11</td><td>–1</td><td>–2</td><td>0</td><td>3</td><td>7</td><td>–</td></tr>
<tr><td>12</td><td>1</td><td>3</td><td>1</td><td>3</td><td>10,5</td><td>3</td></tr>
<tr><td rowspan="2">Кол-во нетипичных сдвигов</td><td rowspan="2">3</td><td rowspan="2">1</td><td rowspan="2">0</td><td colspan="3">Суммы рангов нетипичных сдвигов</td></tr>
<tr><td>14</td><td>7</td><td>0</td></tr>
<tr><td>Всего сдвигов</td><td>12</td><td>12</td><td>9</td><td>12</td><td>12</td><td>9</td></tr>
</table>

Сформулируем гипотезы:

Для шкалы «Активное слушание», $n = 12$: H_0: Преобладание типичного (положительного) сдвига в самооценках уровня владения коммуникативными навыками является случайным.

H_1: Преобладание типичного (положительного) сдвига в самооценках уровня владения коммуникативными навыками не является случайным.

Из табл. 2.11 мы видим, что положительных сдвигов по всем шкалам больше.

$$G_{\text{кр}} = \begin{cases} 2 \quad \text{при} \quad p \le 0{,}05, \\ 1 \quad \text{при} \quad p \le 0{,}01; \end{cases}$$

$$G_{\text{эмп}} = 3;$$

$$G_{\text{эмп}} > G_{\text{кр}}.$$

Ответ: H_0 принимается.

Для шкалы «Снижение напряжения», $n = 12$:

$$G_{\text{кр}} = \begin{cases} 2 \quad \text{при} \quad p \le 0{,}05, \\ 1 \quad \text{при} \quad p \le 0{,}01; \end{cases}$$

$$G_{\text{эмп}} = 1;$$

$$G_{\text{эмп}} \le G_{\text{кр}}.$$

Ответ: H_0 отвергается. Принимается H_1. Преобладание положительного сдвига в самооценке уровня владения навыком снижения напряжения не является случайным ($p \le 0{,}01$).

Для шкалы «Аргументация», $n = 9$:

$$G_{\text{кр}} = \begin{cases} 1 \quad \text{при} \quad p \le 0{,}05, \\ 0 \quad \text{при} \quad p \le 0{,}01; \end{cases}$$

$$G_{\text{эмп}} = 0;$$

$$G_{\text{эмп}} \le G_{\text{кр}}.$$

Ответ: H_0 отвергается. Принимается H_1 при уровне значимости $p \le 0{,}01$. Преобладание положительного сдвига в самооценке уровня владения навыком аргументация не является случайным.

Итак, наличие положительного сдвига по навыку активного слушания является случайным, а по навыкам владения эмоционального напряжения и аргументации не является случайным.

Использование Т-критерия Уилкоксона

Сформулируем гипотезы:

H_0: Интенсивность положительных сдвигов не превосходит интенсивности отрицательных сдвигов.

H_1: Интенсивность положительных сдвигов превосходит интенсивности отрицательных сдвигов.

В табл. 2.11 нами уже просуммированы ранги «редких», в данном случае отрицательных, сдвигов. Сопоставляем эти значения с максимальными значениями $T_{\text{кр}}$, при которых различия еще могут считаться достоверными.

Для шкалы «Активное слушание», $n = 12$:

$$T_{\text{кр}} = \begin{cases} 17 \quad \text{при} \quad p \le 0{,}05, \\ 9 \quad \text{при} \quad p \le 0{,}01; \end{cases}$$

$$T_{\text{эмп}} = 14;$$

$$T_{\text{эмп}} < T_{\text{кр}}.$$

Ответ: H_0 отклоняется. Преобладание положительных сдвигов по навыкам активного слушания неслучайно.

Для шкалы «Снижение напряжения», $n = 12$:

$$T_{\text{кр}} = \begin{cases} 17 \quad \text{при} \quad p \le 0{,}05, \\ 9 \quad \text{при} \quad p \le 0{,}01; \end{cases}$$

$$T_{\text{эмп}} = 7;$$

$$T_{\text{эмп}} < T_{\text{кр}}.$$

Ответ: H_0 отвергается. Принимается H_1 при уровне значимости $p \le 0{,}01$. Преобладание положительных сдвигов по навыку снижения напряжения не является случайным.

Для шкалы «Аргументация», $n = 9$:

$$T_{\text{кр}} = \begin{cases} 8 \quad \text{при} \quad p \le 0{,}05, \\ 3 \quad \text{при} \quad p \le 0{,}01; \end{cases}$$

$$T_{\text{эмп}} = 0;$$

$$T_{\text{эмп}} < T_{\text{кр}}.$$

Ответ: H_0 отвергается. Принимается H_1 при уровне значимости $p \le 0{,}01$. Преобладание положительных сдвигов по навыкам аргументации не является случайным.

Итак, участники ощущают значимые положительные сдвиги по всем трем группам коммуникативных навыков.

Для того чтобы проверить статистическую достоверность различий двух зависимых выборок в статистическом пакете *STATGRAPHICS* необходимо выбрать в меню *Compare – Two Samples – Paired-Sample Comparison*. Далее ввести в поля *Sample 1 (выборка 1) и Sample 2 (выборка 2)* исследуемые выборки и нажать кнопку *OK*. После этого на экране открывается окно с общей информацией. Активизировав кнопку табличных опций , необходимо отметить галочкой пункт *Hypothesis Tests* и завершить ввод информации клавишей *Enter*. Двойной щелчок правой кнопки мышки на окне *Hypothesis Tests* раскрывает окно на всю рабочую область. Данное окно отображает результаты трех тестов: *t*-теста, знакового и знакового рангового теста (*T*-критерия Уилкоксона). По подсчитанным *p*-значениям делается вывод о принятии или отклонении нулевой гипотезы. Например, для шкалы «*Активное слушание*», результаты тестов представлены на рис. 2.11.

```
sign test
---------
Null hypothesis: median = 0,0
Alternative: not equal

Number of values below hypothesized median: 9
Number of values above hypothesized median: 3

Large sample test statistic = 1,44338 (continuity correction applied)
P-Value = 0,148914

Do not reject the null hypothesis for alpha = 0,05.


signed rank test
----------------
Null hypothesis: median = 0,0
Alternative: not equal

Average rank of values below hypothesized median: 7,11111
Average rank of values above hypothesized median: 4,66667

Large sample test statistic = 1,95298 (continuity correction applied)
P-Value = 0,0508214

Do not reject the null hypothesis for alpha = 0,05.
```

Рис. 2.11. Результаты знакового и рангового знакового теста (Т-критерий Уилкоксона) (STATGRAPHICS)

Уровень значимости для знакового теста (*sing test*) равен 0,15, поскольку $p > 0,05$, то нулевая гипотеза отклоняется. Для знакового рангового теста (T-критерия Уилкоксона) уровень значимости равен 0,05 и поскольку $p \leq 0,05$, то нулевую гипотезу нельзя не учитывать.

В данном случае T-критерий доказал свою большую мощность по сравнению с критерием знаков. Он подтвердил ранее установленные различия на высоком уровне значимости ($p < 0,01$) и позволил выявить их для шкалы «Активное слушание» ($p \leq 0,05$).

Однако мы не можем интерпретировать полученные результаты в терминах эффективности тренинга по меньшей мере по двум причинам:

1) у нас не было контрольной группы, у которой измерялись бы те же показатели с тем же интервалом времени;
2) показатели самооценки после тренинга могли отражать желание испытуемых косвенно поблагодарить тренера за его работу.

Рассмотрим использование T-критерия Уилкоксона на интернет-ресурсе http://www.psychol-ok.ru/. После ввода соответствующих выборок «до» и «после» выполняем расчет.

Для шкалы «Снижение напряжения», $n = 12$:

Таблица 2.13

Вычисление сдвигов для шкалы «Снижение напряжения»

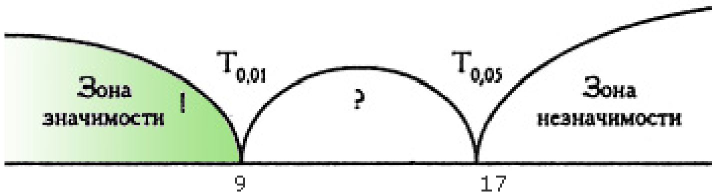
№	«До»	«После»	Сдвиг ($t_{после} - t_{до}$)	Абсолютное значене сдвига	Ранговый номер сдвига
1	5	6	1	1	3
2	1	4	3	3	10,5
3	4	7	3	3	10,5
4	4	5	1	1	3
5	4	5	1	1	3
6	5	7	2	2	7
7	5	8	3	3	10,5
8	5	7	2	2	7
9	5	6	1	1	3
10	6	7	1	1	3
11	6	4	–2	2	7
12	3	6	3	3	10,5
Сумма рангов нетипичных сдвигов:					**7**

Результат: $T_{эмп} = 7$

Критические значения T при n=12

n	$T_{Кр}$	
	0.01	0.05
12	**9**	**17**

Ось значимости:



Полученное эмпирическое значение $T_{эмп}$ находится в зоне значимости.

Рис. 2.12. Результат вычисления Т-критерия Уилкоксона для шкалы «Снижение напряжения»

Для шкалы «Активное слушание», n = 12:

Таблица 2.14

Вычисление сдвигов для шкалы «Активное слушание»

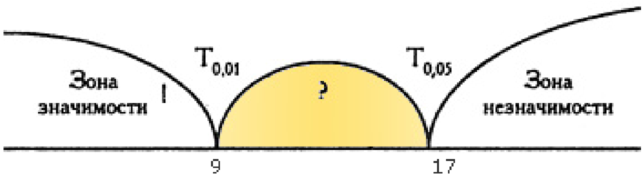
№	«До»	«После»	Сдвиг ($t_{после} - t_{до}$)	Абсолютное значене сдвига	Ранговый номер сдвига
1	6	7	1	1	3
2	3	5	2	2	8
3	4	8	4	4	11,5
4	4	6	2	2	8
5	6	4	–2	2	8
6	6	8	2	2	8
7	3	7	4	4	11,5
8	6	5	–1	1	3
9	6	7	1	1	3
10	5	7	2	2	8
11	6	5	–1	1	3
12	6	7	1	1	3
Сумма рангов нетипичных сдвигов:					**14**

Результат: $T_{Эмп}$ = 14

Критические значения T при n=12

n	$T_{Кр}$	
	0.01	0.05
12	**9**	**17**

Ось значимости:



Полученное эмпирическое значение $T_{Эмп}$ находится в зоне неопределенности.

Рис. 2.13. Результат вычисления T-критерия Уилкоксона для шкалы «Активное слушание»

2.4.5. Критерий серий (*W*-критерий)

Критерий серий предназначен для анализа двух независимых выборок. Размеры этих выборок могут различаться.

Критерий серий в некоторых случаях целесообразно использовать при связанных выборках, рассматривая при этом их как независимые. Дело в том, что связи между парами «опыт – контроль» могут оказаться слабыми, а различия между ними – сильными. Тогда, рассматривая выборки как независимые, мы можем обнаружить различия, не выявляемые критериями для связанных выборок. Это замечание особо важно для очень малых выборок, так как критерий знаков и *T*-критерий можно применять при выборках, включающих не менее 5 пар, а *W*-критерий применим уже при $m = n = 3$.

Назначение критерия – проверка гипотезы о статистической однородности двух выборок. Для критерия серий существенны не сами значения результатов наблюдений, а порядок их расположения. Различие между группами наиболее значимы, когда после упорядочения располагаются сначала все числа первого ряда, а потом второго. В этом свободном от распределения критерии равенства двух распределений используется понятие серий в объединенном вариационном ряду.

Порядок расчета

Рассматриваются две выборки $x_1, \ldots, x_m$ (выборка x) и $y_1, \ldots, y_n$ (выборка y) объемов m и n.

1. Необходимо составить общий упорядоченный ряд, расположив результаты эксперимента в порядке их возрастания.

2. Подсчитать количество нарушений расположения чисел по сравнению с их наиболее различающимся расположением. Одним нарушением (инверсией) считают такое расположение, когда перед некоторым числом первого столбца стоит число второго столбца. Если перед некоторым числом первого столбца стоят два числа второго столбца, это считают за две инверсии и т. д. Число инверсий обозначают через $W_{эмп}$ – это и будет эмпирическое значение для критерия серий.

В любом общем упорядоченном ряду инверсии можно подсчитывать двумя способами: относительно первой и относительно второй группы. Следует выбирать тот способ, который дает наименьшую сумму инверсий.

3. Величину $W_{эмп}$ сравниваем с табличным значением $W_{кр}$ (см. [6, 24]).

При $W_{эмп} \geq W_{кр}$ верно утверждение, что выборки однородны.

С увеличением числа наблюдений трудоемкость критерия серий возрастает. В этих случаях целесообразно сначала использовать критерий Q (Розенбаума).

2.4.6. *Q*-критерий Розенбаума

Q-критерий Розенбаума позволяет быстро оценить различия между двумя независимыми выборками по какому-либо признаку.

Критерий используется для оценки различий между двумя выборками по уровню какого-либо признака, количественно измеренного. Объемы выборок должны примерно совпадать. Однако если критерий Q не выявляет достоверных различий, это еще не означает, что их действительно нет.

Ограничения критерия

1. В каждой из сопоставимых выборок должно быть не менее 11 наблюдений.
2. Если в обеих выборках меньше 50 наблюдений, то абсолютная величина разности между m и n не должна быть больше 10 наблюдений. Если в каждой из выборок больше 51 наблюдения, но меньше 100, то абсолютная величина разности между m и n не должна быть больше 20 наблюдений. Если в каждой из выборок больше 100 наблюдений, то допускается, чтобы одна из выборок была больше другой не более чем в 1,5–2 раза [8].
3. Диапазоны разброса значений в двух выборках должны не совпадать между собой, в противном случае применение критерия бессмысленно.

Порядок расчета

1. Проверить, выполняются ли ограничения: $m, n \geq 11$, $m \approx n$.
2. Упорядочить значения отдельно в каждой выборке по степени возрастания признака. Считать первой ту выборку, значения в которой

предположительно выше, а второй – ту, где значения предположительно ниже.

3. Определить самое высокое (максимальное) значение в выборке 2.
4. Подсчитать количество значений в выборке 1, которые выше максимального значения в выборке 2. Обозначить полученную величину как S_1.
5. Определить самое низкое (минимальное) значение в выборке.
6. Подсчитать количество значений в выборке 2, которые ниже минимального значения в выборке 1. Обозначить полученную величину как S_2.
7. Подсчитать эмпирическое значение Q по формуле $Q = S_1 + S_2$.
8. Определить критические значения для m и n. Если $Q_{эмп} \geq Q_{0,05}$, гипотеза H_0 отвергается.
9. При m, n > 26 сопоставить полученное эмпирическое значение с $Q_{кр}$. Если $Q_{эмп} \geq Q_{кр} = 8$, H_0 отвергается.

Пример 2.8 [24]

У предполагаемых участников психологического эксперимента, моделирующего деятельность воздушного диспетчера, был измерен уровень вербального и невербального интеллекта с помощью методики Д. Векслера. Было обследовано 26 юношей в возрасте от 18 до 24 лет (средний возраст 20,5 лет). 14 из них были студентами физического факультета, а 12 – студентами психологического факультета Ленинградского университета. Показатели вербального интеллекта представлены в табл. 2.15.

Таблица 2.15

Индивидуальные значения вербального интеллекта

Студенты-физики		Студенты-психологи	
№ испытуемого	Показатель вербального интеллекта	№ испытуемого	Показатель вербального интеллекта
1	132	1	126
2	134	2	127
3	124	3	132
4	132	4	120
5	135	5	119
6	132	6	126
7	131	7	120
8	132	8	123
9	121	9	120
10	127	10	116

Окончание табл. 2.15

Студенты-физики		Студенты-психологи	
№ испытуемого	Показатель вербального интеллекта	№ испытуемого	Показатель вербального интеллекта
11	136	11	123
12	129	12	115
13	136	–	–
14	136	–	–

Можно ли утверждать, что одна из групп превосходит другую по уровню вербального интеллекта?

Упорядочим значения в обеих выборках (табл. 2.16). Как видно из табл. 2.16, показатель вербального интеллекта у физиков выше, чем у психологов. Введем соответствующие обозначения: первый ряд тот, значения которого «выше» – ряд физиков, а второй тот, чьи значения «ниже» – ряд психологов.

Сформулируем гипотезы:

H_0: Студенты-физики не превосходят студентов-психологов по уровню вербального интеллекта.

H_1: Студенты-физики превосходят студентов-психологов по уровню вербального интеллекта.

Таблица 2.16

Упорядоченные ряды индивидуальных значений [24]

Студенты-физики		Студенты-психологи
136	S_1	
136		
136		
135		
134		
132		132
132		
132		
132		
131		
129		
127		127
		126
		126
124		

Окончание табл. 2.16

Студенты-физики	Студенты-психологи
	123
	123
121	
	120
	120
	120 S_2
	119
	116
	115

По табл. 2.16 определяем количество значений первого ряда, которые больше максимального значения второго ряда: $S_1 = 5$.

Теперь определяем количество значений первого ряда, которые меньше минимального значения первого ряда: $S_2 = 6$.

Вычисляем $Q_{эмп}$ по формуле: $Q_{эмп} = S_1 + S_2 = 11$.

По таблицам (см. [6, 24]) определяем критические значения Q при $m = 14$, $n = 12$:

$$Q_{кр} = \begin{cases} 7 \quad \text{при} \quad p \le 0{,}05, \\ 9 \quad \text{при} \quad p \le 0{,}01; \end{cases}$$

Ясно, что чем больше расхождения между выборками, тем больше величина Q. Гипотеза H_0 отклоняется при $Q_{эмп} \ge Q_{кр}$, в противном случае мы будем вынуждены принять H_0.

$$Q_{эмп} > Q_{кр} \ (p \le 0{,}01).$$

Ответ: H_0 отклоняется. Принимается H_1. Студенты-физики превосходят студентов-психологов по уровню вербального интеллекта ($p \le 0{,}01$).

Поскольку уровень значимости выявленных различий достаточно высок ($p < 0{,}01$), мы могли бы на этом остановится. Можно попробовать сопоставить выборки по невербальному интеллекту (см. U-критерий Уилкоксона), так как именно он определяет уровень интеллекта в целом и степень его организованности.

2.4.7. *U*-критерий Уилкоксона (Манна–Уитни)

Область применения U-критерия Уилкоксона – анализ двух независимых выборок. Размеры этих выборок могут различаться.

Назначение критерия – проверка гипотезы о статистической однородности двух выборок. Иногда эту гипотезу называют гипотезой об отсутствии эффекта обработки (имея в виду, что одна из выборок держит характеристики объектов, подвергшихся некоему воздействию, а другая – характеристики контрольных объектов). Он основан на попарном сравнении результатов из первой и второй выборок.

Ограничения критерия

1. В каждой выборке должно быть не менее 3 наблюдений: m и $n \geq 3$; допускается, чтобы в одной выборке было 2 наблюдения, но тогда во второй их должно быть не менее 5.
2. В каждой выборке должно быть не более 60 наблюдений; m и $n \leq 60$. Однако уже при m, $n > 20$ ранжирование становиться достаточно трудоемким.

Порядок расчета

1. Расположить числовые значения сравниваемых выборок в возрастающем порядке в один общий ряд и пронумеровать члены общего ряда от 1 до $N = m + n$ (эти номера будут «рангами» членов ряда).
2. Отдельно для каждой выборки найти суммы рангов R и определить величины:

$$U_1 = R_1 - \frac{m(m+1)}{2} \text{ и } U_2 = R_2 - \frac{n(n+1)}{2}, \tag{2.27}$$

   которые отображают связь между суммами рангов первой и второй выборки.
3. В качестве U-критерия использовать меньшую величину $U_{эмп}$, которую сравнить с табличным значением $U_{кр}$. Условием для сохранения принятой H_0-гипотезы служит неравенство $U_{эмп} > U_{кр}$. Критические точки U-критерия $U_{кр}$ для m и n и принимаемого уровня значимости α содержатся в статистических таблицах (см. [6, 24]).

Пример 2.9 [24]

Вернемся к результатам обследования студентов физического и психологического факультетов Ленинградского университета с помощью методики Д. Векслера для измерения вербального и невербального интеллекта. С помощью Q-критерия Розенбаума мы в предыдущем примере смогли с высоким уровнем значимости определить, что уровень вербального интеллекта в выборке студентов физического факультета выше. Попытаемся установить теперь, воспроизводится ли этот результат при сопоставлении выборок по уровню невербального интеллекта. Данные приведены в табл. 2.17.

Таблица 2.17

Индивидуальные значения невербального (НИ)

1	111	1	113
Студенты-физики	Студенты-психологи	2	107
№ испытуемого	Показатель невербального интеллекта	№ испытуемого	Показатель невербального интеллекта
4	90	4	122
5	115	5	117
6	107	6	112
7	106	7	105
8	107	8	108
9	95	9	111
10	116	10	114
11	127	11	102
12	115	12	104
13	102	–	–
14	99	–	–

Таблица 2.18

Подчсет ранговых ранговых сумм [24]

Студенты-физики (m = 14)		Студенты-психологи (n = 12)	
Показатель невербального интеллекта	Ранг	Показатель невербального интеллекта	Ранг
127	26		
		123	25
		122	24
		117	23
116	22		
115	20,5		
115	20,5	114	19
		113	18
		112	17
111	15,5	111	15,5
		108	14
107	11,5	107	11,5
107	11,5		
107	11,5		
106	9		

Окончание табл. 2.18

Студенты-физики (m = 14)			Студенты-психологи (n = 12)	
Показатель невербального интеллекта		Ранг	Показатель невербального интеллекта	Ранг
			105	8
104		6,5	104	6,5
102		4,5	102	4,5
99		3		
95		2		
90		1		
Суммы	1501	165	1338	186
Средние	107,2	–	111,5	–

В соответствии со следующим шагом алгоритма определяем величины U_1 и U_2:

$$U_1 = 186 - \frac{12(12+1)}{2} = 108;$$

$$U_2 = 165 - \frac{14(14+1)}{2} = 60.$$

Для сопоставления с критическим значением выбираем меньшую величину $U_{эмп} = 60$.

По статистическим таблицам (см. [6, 24]) определим критическое значение для $m = 14$, $n = 12$:

$$U_{кр} = \begin{cases} 51 & \text{при } p \le 0{,}05, \\ 38 & \text{при } p \le 0{,}01; \end{cases}$$

Ответ: H_0 принимается. Группа студентов-психологов не превосходит группы студентов-физиков по уровню невербального интеллекта.

Чтобы провести анализ однородности двух независимых выборок критерием Манна–Уитни в статистическом пакете STATGRAPHICS, необходимо в меню выбрать *Compare – Two Samples – Two-Sample Comparison*, задать необходимые параметры в диалоговом окне *Two-Sample Comparison*. После того, как раскроется окно с общей информацией, необходимо активизировать диалоговое окно *Табличные Опции* (*Tabular Options*), нажав кнопку . Далее отметьте пункт *Comparison of Medians*, в котором будут отображаться результаты проведенного анализа однородности выборок критерием Манна–Уитни. Необходимо отметить, что проверяется не только нулевая гипотеза, но и три альтернативные, которые позволяют определить, какая из двух медиан больше, и неравенство медиан (рис. 2.14).

```
Comparison of Medians
---------------------

Median of sample 1: 107,0

Median of sample 2: 111,5

Mann-Whitney (Wilcoxon) W tests to compare mediar

   Null hypothesis: median1 = median2

   Average rank of sample 1: 11,7857
   Average rank of sample 2: 15,5

   (1) Alt. hypothesis: median1 NE median2
          W = 108,0   P-value = 0,22566

   (2) Alt. hypothesis: median1 > median2
          W = 108,0   P-value = 0,88717

   (3) Alt. hypothesis: median1 < median2
          W = 108,0   P-value = 0,11283
```

Рис. 2.14. Результаты анализа однородности выборок критерием Манна–Уитни

Из представленной сводки мы получаем сведения о вычисленных значениях для каждой выборки: медианы, среднего ранга выборки, W-статистики и соответствующие им уровни значимости. Полученные уровни значимости для всех трех альтернативных гипотез больше 0,05, следовательно, нулевая гипотеза отклоняется, т. е. различия между медианами статистически не значимы. Вывод: группа студентов-психологов не превосходит группы студентов-физиков по уровню невербального интеллекта.

Обратим внимание на то, что для данного случая Q-критерий Розенбаума неприменим, так как размах вариативности в группе физиков шире, чем в группе психологов: и самое высокое, и самое низкое значение невербального интеллекта приходятся на группу физиков (см. табл. 2.18).

Проверим реализацию критерия Манна–Уитни на интернет-ресурсе http://www.psychol-ok.ru/. Выбираем пункт «Расчет U-критерия

Манна–Уитни» и вводим значения двух выборок, нажимаем «шаг 2». Результат выполнения:

Таблица 2.19

Таблица выборок и рангов

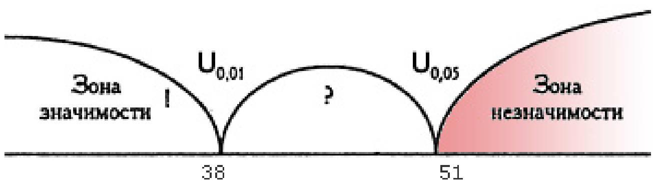
№	Выборка 1	Ранг 1	Выборка 2	Ранг 2
1	127	26	123	25
2	116	22	122	24
3	115	20,5	117	23
4	115	20,5	114	19
5	111	15,5	113	18
6	107	11,5	114	17
7	107	11,5	111	15,5
8	107	11,5	108	14
9	106	9	107	11,5
10	104	6,5	105	8
11	102	4,5	104	6,5
12	99	3	102	4,5
13	95	2		
14	90	1		
Суммы:		165		186

Результат: $U_{Эмп} = 60$

Критические значения

$U_{Кр}$	
p≤0.01	p≤0.05
38	**51**

Ось значимости:



Полученное эмпирическое значение $U_{эмп}$(60) находится в зоне незначимости.

Рис. 2.15. Результат вычисления Uэмп

2.4.8. Выявление различий в уровне исследуемого признака

Очень часто перед исследователем в психологии стоит задача выявления различий между двумя и более выборками испытуемых. Это может быть, например, задача определения психологических особенностей хронически больных детей по сравнению со здоровыми.

По выявленным в исследовании статически достоверным различиям формируется «групповой профиль» или «усредненный портрет» человека той или иной профессии, статуса, соматического заболевания и др. [24].

В последние годы все чаще встает задача выявления психологического портрета специалиста новых профессий: «успешного менеджера», «успешного политика», «успешного коммерческого директора» и др. Такого рода исследования не всегда подразумевают участие двух или более выборок. Иногда обследуется одна, но достаточно представительная выборка численностью не менее 60 человек, а затем внутри этой выборки выделяются группы более и менее успешных специалистов, и их данные по исследованным переменным сопоставляются между собой. В самом простом случае критерием для разделения выборки на «успешных» и «неуспешных» будет средняя величина по показателю успешности.

Чем меньше испытуемых оказывается в группах, тем меньше у нас возможностей для выявления достоверных различий, так как критические значения большинства критериев при малых n строже, чем при больших n.

Как отмечено в [24], при нестрогом разделении испытуемых на группы мы теряем в точности, а при строгом – в количестве испытуемых.

При решении задачи выявления различий в уровневых показателях следует помнить, что «усредненный профиль успешного специалиста» должен рассматриваться скорее как исследовательский результат, позволяющий сформулировать гипотезы для дальнейших исследований, а не как основание для профессионального отбора.

Сопоставление уровневых показателей в разных выборках может быть необходимой частью комплексных диагностических, учебных, психокоррекционных и иных программ. Оно помогает нам обратить внимание на те особенности обследованных выборок, которые должны быть учтены и использованы при адаптации психокоррекционных программ к данной группе в процессе их конкретного воплощения.

Критерии, которые рассматриваются в данной главе, предполагают, что мы сопоставляем так называемые независимые выборки, т. е. две или более выборки, состоящие из разных испытуемых. В противоположность этому, если мы обследуем одну и ту же выборку испытуемых, несколько раз подвергая ее аналогичным измерениям («замерам»), то перед нами – так называемые связанные, или зависимые, выборки данных.

Решение о выборе того или иного критерия принимается на основе того, сколько выборок сопоставляется и каков их объем.

2.4.9. *H*-критерий Крускала–Уоллиса

Критерий предназначен для оценки различий одновременно между тремя, четырьмя и так далее выборками по уровню какого-либо признака.

Он позволяет установить, что уровень признака изменяется при переходе от группы к группе, но не указывает на направление этих изменений.

Гипотезы для критерия Крускала–Уоллиса формулируются следующим образом:

H_0: Между выборками 1, 2, 3 и так далее существуют лишь случайные различия по уровню исследуемого признака.

H_1: Между выборками 1, 2, 3 и так далее существую неслучайные различия по уровню исследуемого признака.

H-критерий иногда рассматривается как непараметрический аналог метода дисперсионного однофакторного анализа для несвязных выборок [31].

Данный критерий является продолжением критерия U на большее, чем две сопоставляемые выборки. Все индивидуальные значения ранжируются так, как если бы это была одна большая выборка. Затем все индивидуальные значения возвращаются в свои первоначальные выборки, и мы подсчитываем суммы полученных ими рангов отдельно по каждой выборке. Если различия между выборками случайны, суммы рангов не будут различаться сколько-нибудь существенно, так как высокие и низкие ранги равномерно распределятся между выборками. Но если в одной из выборок будут преобладать низкие значения рангов, в другой – высокие, а в третьей – средние, то критерий H позволит установить эти различия.

Теперь познакомимся с алгоритмом подсчета H-критерия Крускала–Уоллиса.

1. Необходимо составить общий упорядоченный ряд. Результаты эксперимента расположены в порядке их возрастания.
2. Проранжировать значения, приписывая меньшему значению меньший ранг. Общее количество рангов равно количеству испытуемых в выборке.
3. Подсчитать сумму рангов отдельно по каждой группе. Проверить совпадение общей суммы рангов с расчетной.
4. Подсчитать значение H-критерия по формуле

$$H = \left[\frac{12}{N(N+1)} \sum_{j=1}^{c} \frac{T_j^2}{n} \right] - 3(N+1), \qquad (2.28)$$

где N – общее количество испытуемых в объединенной выборке; c – количество групп (выборок); n – количество испытуемых в каждой группе; T_j – суммы рангов по каждой группе.

5. При количестве групп $c = 3$ и n_1, n_2, $n_3 \leq 5$ определить критическое значение и соответствующий им уровень значимости (см. [6, 31]). Если $H_{эмп}$ равна или превышает критическое значение при уровне значимости $p = 0,05$ (обозначается $H_{0,05}$), H_0 отвергается.

При количестве групп $c > 3$ или количестве испытуемых n_1, n_2, $n_3 > 5$ критические значения определятся как критические значения χ^2 (см. [6, 31]). Если $H_{эмп}$ равен или превышает критическое значение χ^2, H_0 отвергается.

Ограничения критерия

1. При сопоставлении трех выборок допускается, чтобы в одной из них $n = 3$, а в других $n = 2$. Но при таких численных составах выборок мы сможем установить различия на низшем уровне значимости (р ≤ 0,05).

Для того чтобы оказалось возможным диагностировать различия на более высоком уровне значимости ($p \leq 0,01$), необходимо, чтобы в каждой выборке было не менее 3 наблюдений, или чтобы по крайней мере в одной из них было 4 наблюдения, а в других – по 2; при этом неважно, в какой именно выборке сколько испытуемых, а важно соотношение 4 : 2 : 2.

2. При большом количестве выборок и испытуемых в каждой выборке необходимо пользоваться таблицей критических значений критерия χ^2, поскольку критерий Крускала–Уоллиса асимптотически приближается к распределению χ^2 [31].

Количество степеней свободы при этом определяется по формуле $\nu = c - 1$, где c – количество выборок.

Пример 2.10 [24]

В эксперименте по исследованию интеллектуальной настойчивости 22 испытуемым предъявлялись сначала разрешимые четырехбуквенные, пятибуквенные и шестибуквенные анаграммы, а затем неразрешимые анаграммы, время работы над которыми не ограничивалось. Эксперимент проводился индивидуально с каждым испытуемым. Использовалось четыре комплекта анаграмм. У исследователя возникло впечатление, что над некоторыми неразрешимыми анаграммами испытуемые продолжали работать дольше, чем над другими, и, возможно, необходимо будет делать поправку на то, какая именно неразрешимая анаграмма предъявлялась тому или иному испытуемому. Показатели длительности попыток в решении неразрешимых анаграмм представлены в табл. 2.20. Все испытуемые были юношами-студентами технического вуза в возрасте от 20 до 22 лет.

Таблица 2.20

Показатели длительности попыток решения 4 неразрешимых анаграмм в секундах (n = 22) [31]

	Анаграмма ФОЛИТОН $n_1 = 4$	Анаграмма КАМУСТО $n_2 = 8$	Анаграмма СНЕРАКО $n_3 = 6$	Анаграмма ГРУТОСИЛ $n_4 = 4$
1	145	145	128	60
2	194	210	283	2361
3	731	236	469	2416
4	1200	385	482	3600
5	–	720	1678	–
6	–	848	2081	–
7	–	905	–	–
8	–	1080	–	–
Суммы	2270	4549	5121	8437
Средние	568	566	854	2109

Можно ли утверждать, что длительность попыток решения каждой из четырех неразрешимых анаграмм примерно одинакова?

Сформулируем гипотезы:

H_0: 4 группы испытуемых, получившие разные неразрешимые анаграммы, не различаются по длительности попыток их решения.

H_1: 4 группы испытуемых, получившие разные неразрешимые анаграммы, различаются по длительности попыток их решения.

Воспользуемся алгоритмом подсчета H-критерия Крускала–Уоллиса. Результаты работы алгоритма представлены в табл. 2.21.

Общая сумма рангов = 38,5 + 82,5 + 68 + 64 = 253. Расчетная сумма рангов:

$$\sum R_i = \frac{22(22+1)}{2} = 253.$$

Равенство реальной и расчетной сумм соблюдено.

Теперь определяем эмпирическое значение H-критерия:

$$H_{\text{эмп}} = \left[\frac{12}{22(22+1)}\left(\frac{38{,}5^2}{4} + \frac{82{,}5^2}{8} + \frac{68^2}{6} + \frac{64^2}{4}\right)\right] - 3(22+1) = 2{,}48.$$

Таблица 2.21

Подсчет ранговых сумм по группам испытуемых, работавших над четырьмя неразрешимыми анаграммами [24]

Анаграмма ФОЛИТОН $n_1 = 4$		Анаграмма КАМУСТО $n_2 = 8$		Анаграмма СНЕРАКО $n_3 = 6$		Анаграмма ГРУТОСИЛ $n_4 = 4$	
Длительность	Ранг	Длительность	Ранг	Длительность	Ранг	Длительность	Ранг
						60	1
				128	2		
145	3,5	145	3,5				
194	5						
		210	6				
		236	7				
				283	8		
		385	9				
				469	10		
				482	11		
		720	12				
731	13						
		848	14				
		905	15				
		1080	16				
1200	17						
				1678	18		
				2081	19		
						2361	20
						2416	21
						3600	22
Суммы	38,5		8205		68		64
Средние	9,6		10,3		11,3		16,0

Поскольку таблицы критических значений критерия H предусмотрены только для количества групп $c = 3$, а в данном случае у нас 4 группы, придется сопоставлять полученное эмпирическое значение H с критическими значениями χ^2. Для этого вначале определим количество степеней свободы ν для $c = 4$: $\nu = c - 1 = 4 - 1 = 3$.

Определим критические значения (см. [6, 31]) для $\nu = 3$:

$$\chi^2_{\text{кр}} = \begin{cases} 7{,}815 \ \ (p \le 0{,}05), \\ 11{,}345 \ \ (p \le 0{,}01); \end{cases}$$

$$H_{\text{эмп}} = 2{,}48;$$

$$H_{\text{эмп}} < \chi^2_{\text{кр}}.$$

Ответ: H_0 принимается: 4 группы испытуемых, получившие разные неразрешимые анаграммы, не различаются по длительности попыток их решения.

Оценим различия одновременно между четырьмя выборками по уровню признака в статистическом пакете *STATISTICA*. Загрузим модуль *Nonparametric Statistics*. В окне диалога непараметрической статистики (рис. 2.16) выберем пункт *Kruskal-Wallis ANOVA*, *median test.*

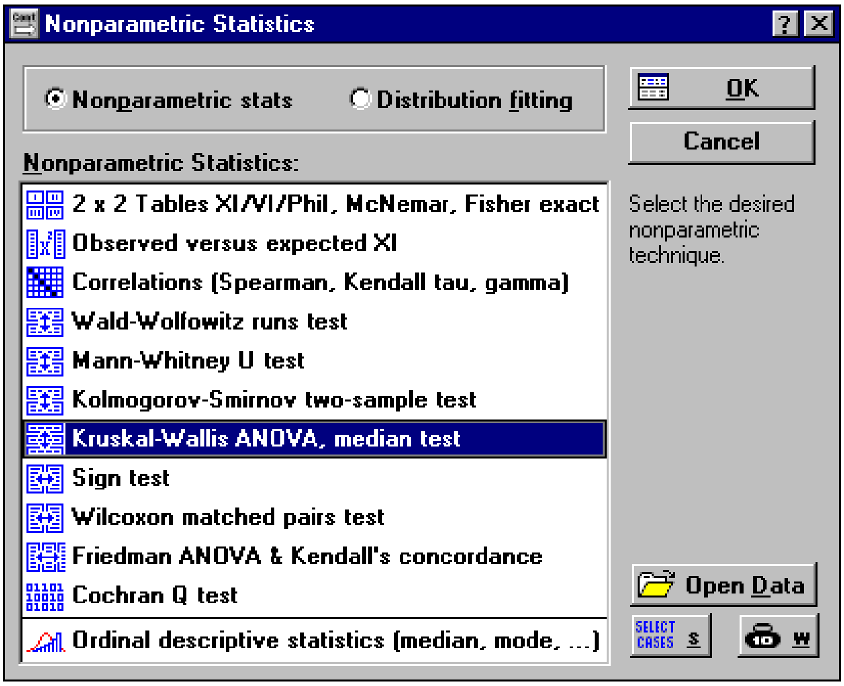


Рис. 2.16. Окно непараметрической статистики (STATISTICA)

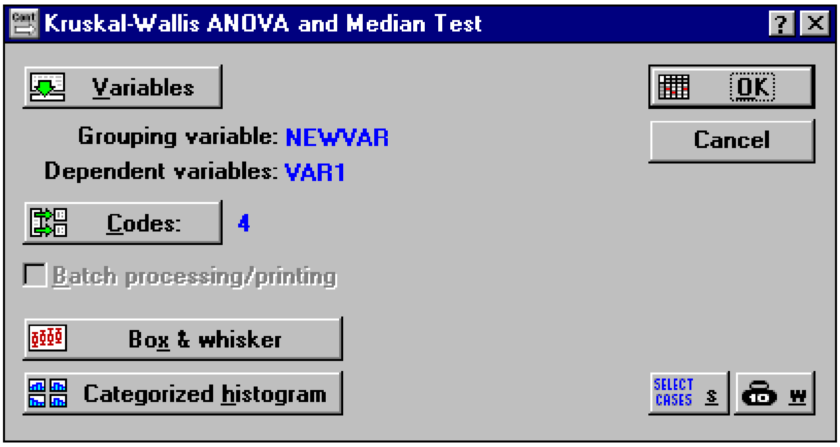


Рис. 2.17. Окно ввода параметров анализа различий групп (STATISTICA)

Далее в окне ввода параметров анализа 

необходимо ввести группирующую переменную и зависимую переменную (в нашем примере группирующая переменная – это номер анаграммы, а зависимая все показатели длительности решения анаграммы). Завершить ввод нажатием кнопки **OK** (рис. 2.17).

На экране появится окно вывода результатов, где вверху будут значения: хи-квадрат, степени свободы и *p*-значение.

```
Chi-Square = 1,666667, df = 3, p = ,6444
```

Полученный уровень значимости равен 0,644, что позволяет сделать вывод о том, что четыре группы испытуемых, получившие разные неразрешимые анаграммы, не различаются по длительности попыток их решения.

2.5. Многофункциональный статистический критерий φ^*

2.5.1. Понятие многофункциональных критериев

Многофункциональные статистические критерии – это критерии, которые могут использоваться по отношению к самым разнообразным данным, выборкам и задачам [24].

Многофункциональные критерии позволяют решать задачи сопоставления уровней исследуемого признака, сдвигов в значениях исследуемого признака и сравнения распределений.

К числу многофункциональных критериев в полной мере относится критерий φ^* Фишера (угловое преобразование Фишера).

Многофункциональные критерии построены на сопоставлении долей, выраженных в долях единицы или в процентах. Суть критериев состоит в определении того, какая доля наблюдений (реакций, выборов, испытуемых) в данной выборке характеризуется интересующим исследователя эффектом и какая доля этим эффектом не характеризуется.

Таким эффектом может быть [24]:

а) определенное *значение* качественно определяемого признака – например: выражение согласия с каким-либо предложением; выбор правой дорожки из двух симметричных дорожек; принадлежность к определенному полу; присутствие фигуры отца в раннем воспоминании и др.;

б) определенный *уровень* количественно измеряемого признака, например получение оценки, превосходящей проходной балл; решение задачи менее чем за 20 с; факт работы в команде, по численности превышающей 4 человека; выбор дистанции в разговоре, превышающей 50 см, и др.;

в) определенное *соотношение* значений или уровней исследуемого признака, например более частый выбор альтернатив А и Б по сравнению с альтернативами В и Г; преимущественное проявление крайних значений признака как самых высоких, так и самых низких; преобладание положительных сдвигов над отрицательными и др.

Критерий φ^* применяется в тех случаях, когда имеется *две* выборки испытуемых.

2.5.2. Критерий φ^* – угловое преобразование Фишера

Критерий Фишера предназначен для сопоставления двух выборок по частоте встречаемости интересующего исследователя эффекта.

Критерий оценивает достоверность различий между процентными долями двух выборок, в которых зарегистрирован интересующий нас эффект.

Суть углового преобразования Фишера состоит в переводе процентных долей в величины центрального угла, который измеряется в радианах. Большей процентной доле будет соответствовать больший угол φ, а меньшей доле – меньший угол, но соотношения здесь не линейные:

$$\varphi = 2\arcsin(\sqrt{P}), \tag{2.29}$$

где P – процентная доля, выраженная в долях единицы (см. рис. 2.18).

При увеличении расхождения между углами φ_1 и φ_2 и увеличении численности выборок значение критерия возрастает. Чем больше величина φ^*, тем более вероятно, что различия достоверны.

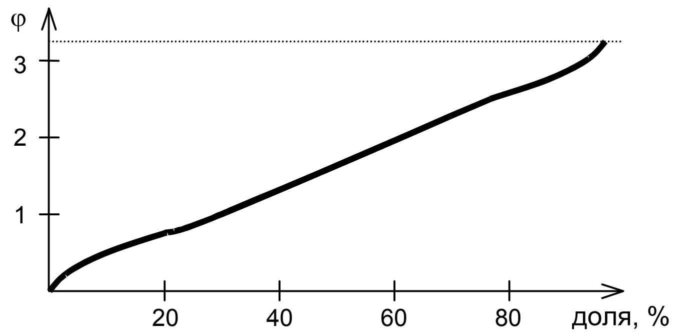


Рис. 2.18. График зависимости угла φ от процентной доли [24]

Гипотезы для данного критерия формулируются следующим образом:

H_0: Доля лиц, у которых проявляется исследуемый эффект, в выборке 1 не больше, чем в выборке 2.

H_1: Доля лиц, у которых проявляется исследуемый эффект, в выборке 1 больше, чем в выборке 2.

Метод углового преобразования несколько более абстрактен, чем остальные критерии.

Формула, которой придерживается автор [9] при подсчете значений φ, предполагает, что 100 % составляют угол $\varphi = 3{,}142$, т. е. округленную величину $\pi = 3{,}14159$. Это позволяет нам представить сопоставляемые выборки в виде двух полукругов, каждый из которых символизирует 100 % численности своей выборки. Процентные доли испытуемых с «эффектом» будут представлены как секторы, образованные центральными углами φ. На рис. 2.19 представлены два полукруга, иллюстрирующие пример 2.11. В первой выборке 60 % испытуемых решили задачу. Этой процентной доле соответствует угол $\varphi = 1{,}772$. Во второй выборке 40 % испытуемых решили задачу. Этой процентной доле соответствует угол $\varphi = 1{,}369$.

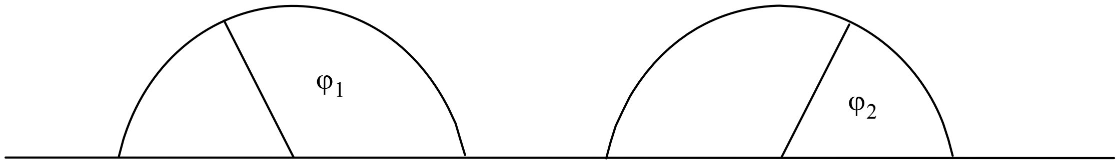


Рис. 2.19. Графическое представление углов, образованных процентными долями испытуемых, решивших задачу в группе 1 (слева) и в группе 2 (справа); отсчет углов идет справа налево [24]

Критерий φ^* позволяет определить, действительно ли один из углов статистически достоверно превосходит другой при данных объемах выборок.

Ограничения критерия φ^*

1. Ни одна из сопоставляемых долей не должна быть равной нулю. Формально нет препятствий для применения метода φ в случаях, когда доля наблюдений в одной из выборок равна 0. Однако в этих случаях результат может оказаться неоправданно завышенным [9].
2. Верхний предел в критерии φ отсутствует – выборки могут быть сколь угодно большими. Нижний предел – два наблюдения в одной из выборок. Однако должны соблюдаться следующие соотношения в численности двух выборок:
   а) если в одной выборке всего 2 наблюдения, то во второй должно быть не менее 30: $n_1 = 2 \rightarrow n_2 \geq 30$;
   б) если в одной из выборок всего три наблюдения, то во второй должно быть не менее семи: $n_1 = 3 \rightarrow n_2 \geq 7$;
   в) если в одной из выборок всего четыре наблюдения, то во второй должно быть не менее пяти: $n_1 = 4 \rightarrow n_2 \geq 5$;
   г) при n_1, $n_2 \geq 5$ возможны любые сопоставления.

В принципе, возможно и сопоставление выборок, не отвечающих этому условию, например, с соотношением $n_1 = 2$, $n_2 = 15$, но в этих случаях не удастся выявить достоверных различий.

Другиx ограничений у критерия φ^* нет.

Рассмотрим два примера из [24], иллюстрирующих возможности критерия φ^*.

Пример 2.11

Сопоставление выборок по качественно определяемому признаку *[24]. В данном варианте использования критерия сравнивают процент испытуемых в одной выборке, характеризующихся каким-либо качеством, с процентом испытуемых в другой выборке, характеризующихся тем же качеством.*

Допустим, нас интересует, различаются ли две группы студентов по успешности решения новой экспериментальной задачи. В первой группе из 20 человек с нею справились 12 человек, а во второй выборке из 25 человек – 10. В первом случае процентная доля решивших задачу составит 12/20·100 % = 60 %, а во второй 10/25·100 % = 40 %. Достоверно ли различаются эти процентные доли при данных n_1 и n_2?

Казалось бы, и «на глаз» можно определить, что 60 % значительно выше 40 %. Однако на самом деле эти различия при данных n_1, n_2 недостоверны.

Проверим это. Поскольку нас интересует факт решения задачи, будем считать «эффектом» успех в решении экспериментальной задачи, а отсутствием эффекта – неудачу в ее решении.

Сформулируем гипотезы:

H_0: Доля лиц, справившихся с задачей, в первой группе не больше, чем во второй группе.

H_1: Доля лиц, справившихся с задачей, в первой группе больше, чем во второй группе.

Теперь построим так называемую четырехклеточную, или четырехпольную, таблицу, которая фактически представляет собой таблицу эмпирических частот по двум значениям признака «есть эффект» – «нет эффекта».

В четырехклеточной таблице, как правило, сверху размечаются столбцы «Есть эффект» и «Нет эффекта», а слева – строки «1 группа» и «2 группа». Участвуют в сопоставлениях, собственно, только поля (ячейки) А и В, т. е. процентные доли по столбцу «Есть эффект» (табл. 2.22).

По таблицам (см. [6, 24]) определяем критические величины φ, соответствующие процентным долям в каждой из групп.

$$\varphi_{1(60\ \%)} = 1{,}772;$$
$$\varphi_{2(40\ \%)} = 1{,}369.$$

Теперь подсчитаем эмпирическое значение φ^* по формуле

$$\varphi^* = (\varphi_1 - \varphi_2)\sqrt{\frac{n_1 n}{n_1 + n_2}}\,, \qquad (2.30)$$

где φ_1 – угол, соответствующий большей процентной доле; φ_2 – угол, соответствующий меньшей процентной доле; n_1 – количество наблюдений в выборке 1; n_2 – количество наблюдений в выборке 2.

Таблица 2.22

Четырехклеточная таблица для расчета критерия при сопоставлении двух групп испытуемых по процентной доле решивших задачу

	«Есть эффект»: задача решена			«Нет эффекта»: задача не решена			Суммы
	Количество испытуемых			Количество испытуемых			
1 группа	12	(60 %)	А	8	(40 %)	Б	20
2 группа	10	(40 %)	В	15	(60 %)	Г	25
Суммы	22	–	–	23	–	–	45

В данном случае

$$\varphi* = (1,772 - 1,369) \cdot \sqrt{\frac{20 \cdot 25}{20 + 25}} = 0,403 \cdot \sqrt{11,11} = 1,34.$$

Величина $\varphi*_{\text{эмп}}$ = 1,34 соответствует уровню значимости:

$$p = 0,09.$$

Можно установить и критические значения $\varphi*$, соответствующие принятым в психологии уровням статистической значимости:

$$\varphi*_{\text{кр}} = \begin{cases} 1,64 \ \ (p \le 0,05), \\ 2,31 \ \ (p \le 0,01); \end{cases}$$

$$\varphi_{\text{эмп}} = 1,34;$$

$$\varphi_{\text{эмп}} \le \varphi_{\text{кр}}.$$

Ответ: H_0 принимается: доля лиц, справившихся с задачей, в первой группе не больше, чем во второй группе.

Ниже будет рассмотрен пример вычисления критерия Фишера в прикладном статистическом пакете *STATISTICA*. В переключателе модулей выберите модуль непараметрической статистики *Nonparametric Statistics*. В диалоговом окне непараметрической статистики выберите пункт *2x2 Tables XI/VI/Phil*, *McNemar*, *Fisher exact*. В раскрывшуюся таблицу введите полученные частоты и нажмите кнопку **OK**. В окне вывода (рис. 2.20) выводится четырехпольная таблица и результаты подсчитанных статистик. Для критерия Фишера уровень значимости двухсторонний равен 0,24, а односторонний 0,15. Гипотеза о том, что в первой группе доля лиц, справившихся с задачей, не больше, чем во второй группе, принимается.

Как отмечает автор [24], «можно лишь посочувствовать исследователю, который считает существенными различия в 20 % и даже в 10 %, не проверив их достоверность с помощью критерия φ^*». В данном случае, например, достоверными были бы только различия не менее чем в 24,3 %.

2 x 2 Table (new.sta)

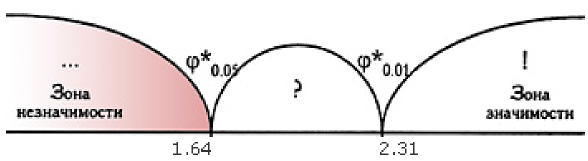
Continue...	Column 1	Column 2	Row Totals
Frequencies, row 1	12	8	20
Percent of total	26,667%	17,778%	44,444%
Frequencies, row 2	10	15	25
Percent of total	22,222%	33,333%	55,556%
Column totals	22	23	45
Percent of total	48,889%	51,111%	
Chi-square (df=1)	1,78	p= ,1823	
V-square (df=1)	1,74	p= ,1873	
Yates corrected Chi-square	1,07	p= ,3013	
Phi-square	,03953		
Fisher exact p, one-tailed		p= ,1507	
two-tailed		p= ,2362	
McNemar Chi-square (A/D)	,15	p= ,7003	
Chi-square (B/C)	,06	p= ,8137	

Рис. 2.20. Окно вывода результатов анализа четырехпольной таблицы

Также рассмотрим решение данного примера, воспользовавшись интернет-ресурсом http://www.psychol-ok.ru/. После ввода данных из табл. 2.22 и выполнения вычисления получили:

Группы	"Есть эффект": задача решена	"Нет эффекта": задача не решена	Суммы
	Количество испытуемых	Количество испытуемых	
1 группа	12 (60%)	8 (40%)	20 (100%)
2 группа	10 (40%)	15 (60%)	25 (100%)

Ось значимости:



Ответ: $\varphi^*_{эмп}$ = 1.343

Полученное эмпирическое значение φ* находится в зоне незначимости. H_1 отвергается

Рис. 2.21. Результат вычисления углового преобразования Фишера

Пример 2.12

Сопоставление двух выборок по количественно измеряемому признаку [24]. В данном варианте использования критерия сравнивают процент испытуемых в одной выборке, которые достигают определенного уровня значения признака, с процентом испытуемых, достигающих этого уровня в другой выборке.

Из 70 юношей – учащихся ПТУ в возрасте от 14 до 16 лет было отобрано по результатам обследования по Фрайбургскому личностному опроснику 10 испытуемых с высоким показателем по шкале агрессивности и 11 испытуемых с низким показателем по шкале агрессивности. Необходимо определить, различаются ли группы агрессивных и неагрессивных юношей по показателю расстояния, которое они спонтанно выбирают в разговоре с сокурсником. Данные представлены в табл. 2.22. Можно заметить, что агрессивные юноши чаще выбирают расстояние в 50 см или даже меньше, в то время как неагрессивные юноши чаще выбирают расстояние, превышающее 50 см.

Таблица 2.22

Показатели расстояния (в см), выбираемого агрессивными и неагрессивными юношами в разговоре с сокурсником [24]

<table>
<tr><td rowspan="2"></td><td colspan="2">Группа 1: юноши с высокими показателями по шкале агрессивности ($n_1 = 10$)</td><td colspan="2">Группа 2: юноши с низкими значениями по шкале агрессивности ($n_2 = 11$)</td></tr>
<tr><td>d (см)</td><td>% доля</td><td>d (см)</td><td>% доля</td></tr>
<tr><td>«Есть эффект»
$d < 50$ см</td><td>30
40
50
50
50
50
50</td><td>70</td><td>40
45</td><td>18,2</td></tr>
<tr><td>«Нет эффекта»
$d > 50$ см</td><td>70
80
90</td><td>30</td><td>65
75
75
75
75
100
100
100
100</td><td>81,8</td></tr>
<tr><td>Суммы</td><td>560</td><td>100</td><td>850</td><td>100</td></tr>
<tr><td>Средние</td><td>56,0</td><td></td><td>77,3</td><td></td></tr>
</table>

Теперь мы можем рассматривать расстояние в 50 см как критическое и считать, что если выбранное испытуемым расстояние меньше или равно 50 см, то «эффект есть», а если выбранное расстояние больше 50 см, то «эффекта нет». Мы видим, что в группе агрессивных юношей эффект наблюдается в 7 из 10, т. е. в 70 % случаев, а в группе неагрессивных юношей – в 2 из 11, т. е. в 18,2 % случаев. Эти процентные доли можно сопоставить по методу φ*, чтобы установить достоверность различий между ними.

Сформулируем гипотезы:

H_0: Доля лиц, которые выбирают дистанцию $d < 50$ см, в группе агрессивных юношей не больше, чем в группе неагрессивных юношей.

H_1: Доля лиц, которые выбирают дистанцию $d < 50$ см, в группе агрессивных юношей больше, чем в группе неагрессивных юношей. Теперь построим так называемую четырехклеточную таблицу (табл. 2.23).

Таблица 2.23

Четырехклеточная таблица для расчета критерия φ при сопоставлении групп агрессивных (n_1 = 10) и неагрессивных юношей (n_2 = 11) [31]*

Группы	«Есть эффект»: $d < 50$			«Нет эффекта»: $d > 50$			Суммы
	Количество испытуемых	% доля		Количество испытуемых	% доля		
1 группа – агрессивные юноши	7	70	А	3	30	Б	10
2 группа – неагрессивные юноши	2	18,2	В	9	81,8	Г	11
Сумма	9	–	–	12	–	–	21

По таблице определяем величины, соответствующие процентным долям «эффекта» в каждой из групп:

$$\varphi_{1(70\,\%)} = 1{,}982;$$
$$\varphi_{2(18{,}2\,\%)} = 0{,}881;$$
$$\varphi* = (1{,}982 - 0{,}881) \cdot \sqrt{\frac{10 \cdot 11}{10 + 11}} = 1{,}101 \cdot \sqrt{5{,}238} = 2{,}520.$$

Критические значения φ* нам уже известны:

$$\varphi*_{\text{кр}} = \begin{cases} 1{,}64 \ \ (p \le 0{,}05), \\ 2{,}31 \ \ (p \le 0{,}01); \end{cases}$$

$$\varphi_{эмп} = 2{,}52;$$
$$\varphi_{эмп} > \varphi_{кр}, \ (p < 0{,}01).$$

Ответ: H_0 отвергается. Принимается H_1. Доля лиц, которые выбирают дистанцию в беседе меньшую или равную 50 см, в группе агрессивных юношей больше, чем в группе неагрессивных юношей ($p < 0{,}01$).

Рассмотрим другой вариант вычисления критерия Фишера в прикладном статистическом пакете *STATISTICA*. В переключателе модулей выберите модуль непараметрической статистики *Nonparametric Statistics*. Загрузите файл с данными и подготовьте их к дальнейшему анализу. Объедините оба признака (группу агрессивных и неагрессивных юношей). Создайте два новых признака: 1) номер группы (агрессивные и неагрессивные); 2) выбираемое расстояние (до 50 см. и выше). Составьте четырехпольную таблицу следующим образом: вызовите переключатель модулей и выберите модуль основных статистик *Basic Statistic and Tables* в меню выберите пункт **Tables and banners** . Откроется окно *Specify Tables*. На экране появится следующее окно, в котором можно задать до шести списков группирующих переменных. Введите первой группирующую переменную номер группы, а второй расстояние. После нажатия кнопки OK, раскрывается окно с таблицами кросстабуляции *Crosstabulation Tables Results* (рис. 2.22). Щелкнув по кнопке **Review summary tables**, вы увидите итоговую таблицу (рис. 2.23), соответствующую табл. 2.23.

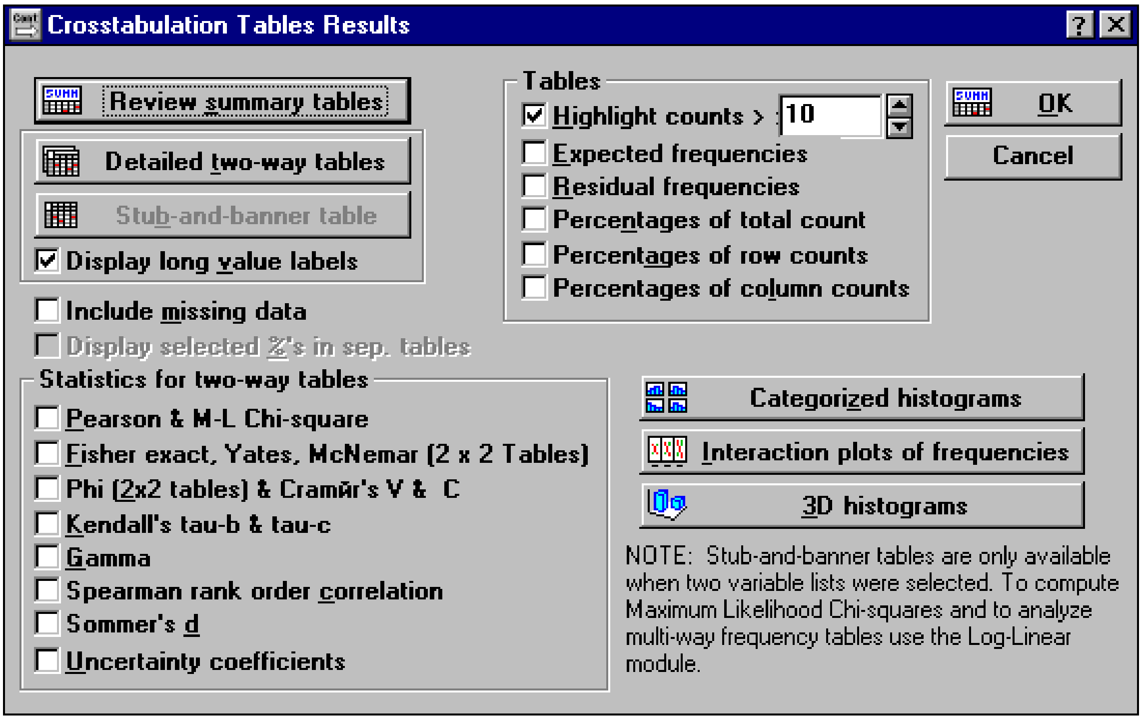
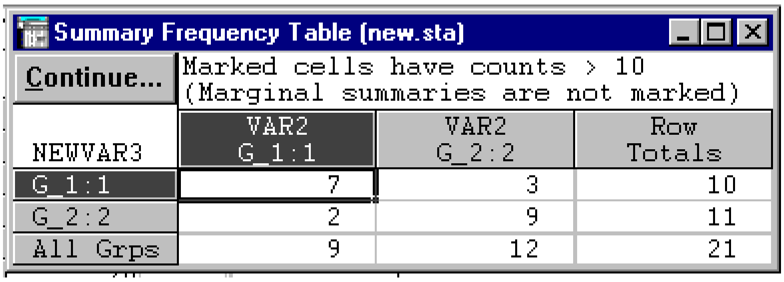


Рис. 2.22. Диалоговое окно Таблицы результатов кросстабуляции



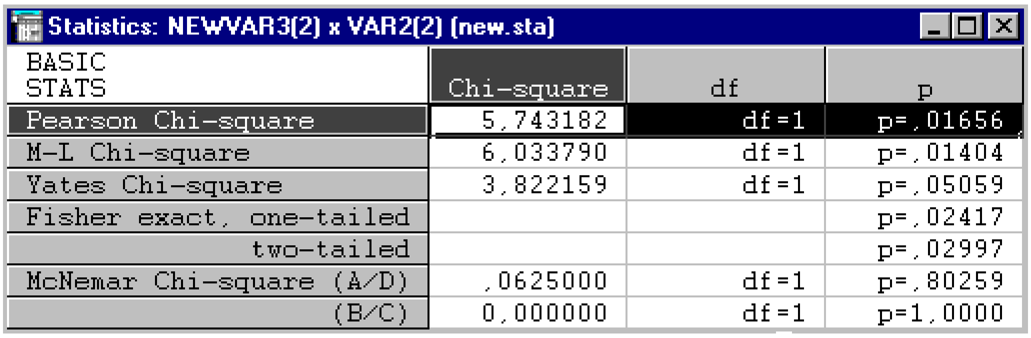
Summary Frequency Table (new.sta)

Continue... Marked cells have counts > 10 (Marginal summaries are not marked)

NEWVAR3	VAR2 G_1:1	VAR2 G_2:2	Row Totals
G_1:1	7	3	10
G_2:2	2	9	11
All Grps	9	12	21

Рис. 2.23. Таблица кросстабуляции (сопряженности)

Для продолжения анализа в верхнем левом углу окна с четырехпольной таблицей нужно нажать кнопку Continue... и вы вновь окажетесь в окне *Crosstabulation Tables Results.* Теперь обратимся к группе опций *Statistics for two-way tables.* Выберете в этой группе опцию *Fisher exact*, *Yates*, *McNemar (2 x 2 Tables)* и щелкните на кнопку Detailed two-way tables. Вы увидите на экране таблицу с подсчитанными статистиками (рис. 2.24). Для критерия Фишера уровень значимости двухсторонний равен 0,03, а односторонний равен 0,02. Оба значения меньше 0,05, что позволяет отвергнуть нулевую гипотезу.



Statistics: NEWVAR3(2) x VAR2(2) (new.sta)

BASIC STATS	Chi-square	df	p
Pearson Chi-square	5,743182	df=1	p=,01656
M-L Chi-square	6,033790	df=1	p=,01404
Yates Chi-square	3,822159	df=1	p=,05059
Fisher exact, one-tailed			p=,02417
two-tailed			p=,02997
McNemar Chi-square (A/D)	,0625000	df=1	p=,80259
(B/C)	0,000000	df=1	p=1,0000

Рис. 2.24. Таблица значений статистик критериев

На основании полученного результата мы можем сделать заключение, что более агрессивные юноши чаще выбирают расстояние менее полуметра, в то время как неагрессивные юноши чаще выбирают большее, чем полметра, расстояние. Мы видим, что агрессивные юноши общаются фактически на границе интимной (0…46 см) и личной зоны (от 46 см). Однако, как справедливо отмечено в [31], интимное расстояние между партнерами является прерогативой не только близких добрых отношений, но и рукопашного боя.

Глава 3
ИССЛЕДОВАНИЕ ВЗАИМОСВЯЗИ МЕЖДУ ПРИЗНАКАМИ

3.1. Параметрические меры связи

3.1.1. Коэффициент корреляции Пирсона

Исследователей часто интересует, как связаны между собой две переменные в данной группе лиц (классы, школы, нации и т. д.). Например, имеют ли ученики, научившиеся читать раньше других, тенденцию к более высокой успеваемости в шестом классе? Наблюдаются ли в больших классах меньшие успехи в приобретении знаний за семестр, чем в небольших классах? Связана ли средняя продолжительность работы педагогов в школе непосредственно со средней заработной платой? Очевидно, для ответа на такие вопросы мы должны провести наблюдения по каждой переменной для группы объектов (типичных представителей, которыми могут быть классы, школы, районы и т. д.). Данные, собранные для ответа на один из подобных вопросов, могут выглядеть, как на приведенной ниже табл. 3.1.

Таблица 3.1

№ учащегося	1	2	3	4	5	6	7	8	9	10	11	12
Оценка *IQ* Стенфорда–Бине (переменная *x*)	120	112	110	120	103	126	113	114	106	108	128	109
Необработанная оценка теста успеваемости по химии (переменная *y*)	31	25	19	24	17	28	18	20	16	15	27	19

В этом примере переменными, которые изучались у 12 школьников, были оценки *IQ*, определенные с помощью шкалы интеллекта Стенфорда–Бине в шестом классе, и успеваемость по химии в средней школе, оцененная на основе теста, состоящего из 35 вопросов.

Связь между двумя переменными можно выразить графически *диаграммой рассеивания.* Диаграмма рассеивания для данных из табл. 3.1 показана на рис. 3.1.

На диаграмме рассеивания каждый ученик изображается точкой. Точка, или метка, располагается в месте пересечения прямых линий, проведенных через оценку *IQ* перпендикулярно оси x и через оценку теста по химии перпендикулярно оси y для каждого ученика. Диаграмма на рис. 3.1 показывает слабую положительную связь x и y. Однако мы пока не имеем обобщенной меры этой связи.

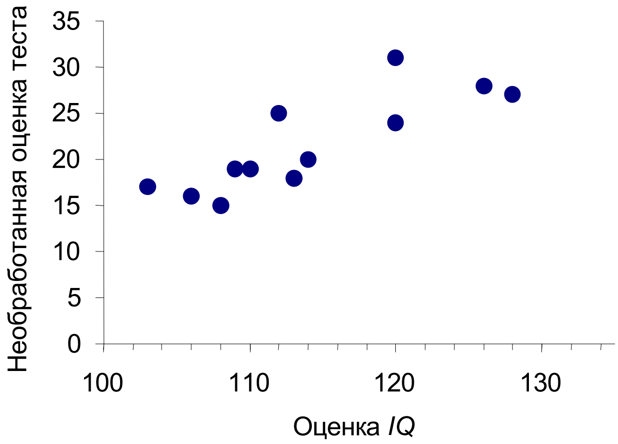


Рис. 3.1. Диаграмма рассеивания, показывающая связь IQ (переменная x) с успеваемостью по химии (переменная y) для 12 школьников [6]

Надо поставить общий вопрос о более точном смысле термина «связь». Существует ли соответствие большого значения x большим или малым значениям тех же объектов по y или систематического распределения по парам с большими и малыми значениями не наблюдается?

Положение объекта относительно остальных в выборке по x и y, определяемое средними двух распределений, проявляется в величинах и знаках отклонений $(x_i - \bar{x})$ и $(y_i - \bar{y})$ соответственно. Если объект имеет высокий уровень по обеим переменным, как, например, учащийся 11 в вышеприведенном примере (табл. 3.1), то произведение $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ будет большим и положительным. Аналогично, если он относительно низок как по x, так и по y, то $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ для него также будет большим и *положительным* (поскольку произведение двух отрицательных чисел положительно). Если x и y в основном связаны *прямо* (большие значения с большими, а малые – с малыми), то большинство произведений $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ будет положительным: следовательно, сумма этих произведений для всех объектов [т. е. $\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})$] будет большой и положительной.

Если x и y имеют *обратную* связь (большое x встречается с малым y и наоборот), то многие объекты с положительными значениями $(x_i - \bar{x})$ будут тяготеть к отрицательным значениям $(y_i - \bar{y})$, а отрицательные $(x_i - \bar{x})$ –

к положительным ($y_i - \bar{y}$). В этом случае произведения ($x_i - \bar{x}$) · ($y_i - \bar{y}$) будут, как правило, отрицательными. Следовательно, $\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})$ будет *отрицательной*, когда x и y связаны обратной зависимостью.

Если x и y не имеют систематической связи (большие x сочетаются с малыми y столь же часто, как и с большими y, и то же самое справедливо для малых x), то среди объектов с большими положительными значениями ($x_i - \bar{x}$) у некоторых ($y_i - \bar{y}$) будут положительные, а у других – отрицательные. При образовании произведений ($x_i - \bar{x}$) · ($y_i - \bar{y}$) одни сомножители станут положительными, а другие – отрицательными. Сумма произведений $\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})$ должна приблизительно балансировать положительные и отрицательные члены и поэтому должна быть довольно близкой к нулю.

Таким образом, мы имеем величину $\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})$, которая велика и положительна, когда x и y сильно связаны прямой зависимостью, близка к нулю в случае отсутствия связи между x и y и велика и отрицательна, когда x и y сильно связаны обратной зависимостью. Однако эта сумма произведений отклонений все еще не является адекватной обобщенной мерой связи. Прежде всего, ее величина зависит от числа пар значений, участвующих в подсчете. Так как мы можем пожелать сравнить степень связи между x и y в двух выборках разного объема, то надо уметь измерять связь независимо от объема групп. Простое усреднение позволяет достигнуть этого. Два средних значения для выборок разного объема сравниваются в терминах центров группирования данных, а простые суммы для двух выборок не сопоставляются. Вот почему мы берем среднее, если хотим, чтобы статистика не зависела от объема выборки. Однако по той же причине, по которой s_x^2 получилась в результате деления суммы квадратов отклонений на $n - 1$, а не на n, нам следует разделить $\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})$ на $n - 1$.

Величина $\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})/(n-1)$ является мерой связи x и y и называется *ковариацией* x и y. Ковариация x и y обозначается через s_{xy}:

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}. \tag{3.1}$$

Заметим, что ковариация x с самим собой – это просто дисперсия x:

$$s_{xx} = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{n-1} = \frac{\sum (x_i - \bar{x})^2}{n-1} = s_x^2. \quad (3.2)$$

Ковариация является вполне удовлетворительной мерой связи во многих задачах физики и техники. И она представляет собой адекватную меру в той же степени, в какой шкала (среднее и дисперсия) переменных не является произвольной и имеет некоторый смысл. Многие переменные, с которыми мы имеем дело, измеряются в произвольных шкалах: среднее и дисперсию можно сделать любыми, какими вздумается, поскольку нас обычно интересует только взаимоположение объектов в группе. Это, в частности, верно для обработки психологических и педагогических данных.

Чтобы избавить меру связи от влияния стандартных отклонений двух групп значений, s_{xy} делят на произведение s_x и s_y. Полученная в результате мера связи x и y называется *коэффициентом корреляции* Пирсона и обозначается r_{xy}:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (3.3)$$

или

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y}) / (n-1)}{\sqrt{\sum (x_i - \bar{x})^2 / (n-1)} \cdot \sqrt{\sum (y_i - \bar{y})^2 / (n-1)}}. \quad (3.4)$$

Величину $1/(n-1)$ можно выделить в качестве сомножителя из двух членов знаменателя уравнения (3.4) ($1/\sqrt{(n-1)}$ из каждого члена) и сократить с $1/(n-1)$ в числителе. Тогда формула для вычисления коэффициента корреляции имеет вид:

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right] \cdot \left[\sum_{i=1}^{n} (y_i - \bar{y})^2\right]}}. \quad (3.5)$$

Простые преобразования приводят к следующей формуле для r_{xy}:

$$r_{xy} = \frac{\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)/n}{\sqrt{\left[\sum x_i^2 - \left(\sum x_i\right)^2 / n\right] \cdot \left[\sum y_i^2 - \left(\sum y_i\right)^2 / n\right]}}, \quad (3.6)$$

которую можно еще более упростить, *образуя расчетную формулу*

$$r_{xy} = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right] \cdot \left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}}. \quad (3.7)$$

Пример 3.1

Рассмотрим способ вычисления коэффициента корреляции r_{xy} в прикладном статистическом пакете SPSS на примере данных теста общих и специальных способностей. Исследователь изучает связь двух типов умственных способностей учеников неполной средней школы: абстрактное мышление и вербальное мышление. Разработаны два теста: для измерения склонности к абстрактному мышлению (x) и к вербальному мышлению (y). Оба теста были предложены 40 школьникам младшего класса средней школы в одном городе с 30 000 жителей. Результаты 40 учащихся представлены в табл. 3.2. В каждом тесте было 50 вопросов, а результатом служило число правильных ответов.

Таблица 3.2 [6]

Исходные оценки по тесту, состоящему из 50 вопросов и измеряющему способность к абстрактному и вербальному мышлению у 40 школьников неполной средней школы

№ ученика	x Абстрактное мышление	y Вербальное мышление	№ ученика	x Абстрактное мышление	y Вербальное мышление
1	19	17	21	38	30
2	32	7	22	25	18
3	33	17	23	35	26
4	44	28	24	22	17
5	28	27	25	40	17
6	35	31	26	42	26
7	39	20	27	41	16
8	39	17	28	41	37
9	44	35	29	37	26
10	44	43	30	30	21
11	24	10	31	31	16
12	37	28	32	41	37
13	29	13	33	42	37
14	40	43	34	24	14
15	42	45	35	43	41
16	32	24	36	36	19
17	48	45	37	39	18
18	43	26	38	39	39
19	33	16	39	39	37
20	47	26	40	48	47

Для того чтобы построить диаграмму рассеивания исследуемых данных (рис. 3.2), представленных в табл. 3.2 воспользуемся возможностями *SPSS*. Выберем из меню *Graphs* команду *Scaterr*… На экране появится окно, где вам будет предложено определить тип выводимого графика. Выберите *Sample* и нажмите кнопку *Define*. Распределите признаки по оси *x* и *y*. Пользователь может ввести заголовки или комментарии, для этого нажмите кнопку *Titles* (рис. 3.2). Завершите ввод информации нажатием кнопки *OK*. На кране активизируется окно вывода, в котором будет отображена диаграмма рассеивания (рис. 3.3). Для того чтобы ввести точное название осей нужно двойным щелчком правой кнопки мыши зайти в окно редактирования графика. Щелчок мыши по названию оси позволит изменить уже существующие данные (например, сокращенное название переменной). Изменить шрифт надписи можно будет после нажатия кнопки T на панели инструментов. После внесения всех необходимых изменений закройте окно редактирования графика.

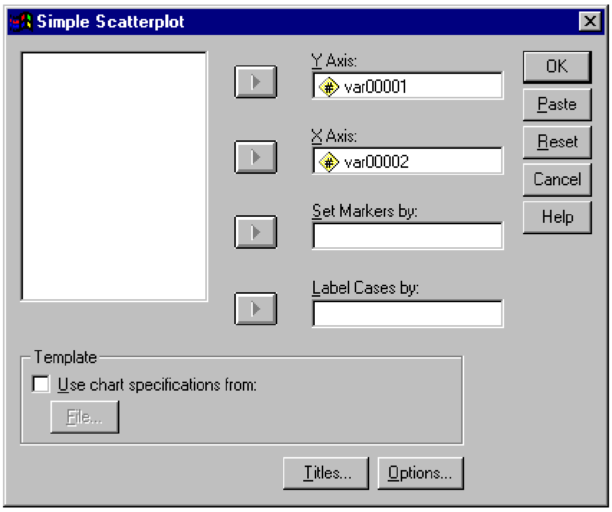


Рис. 3.2. Окно ввода параметров диаграммы рассеивания (SPSS)

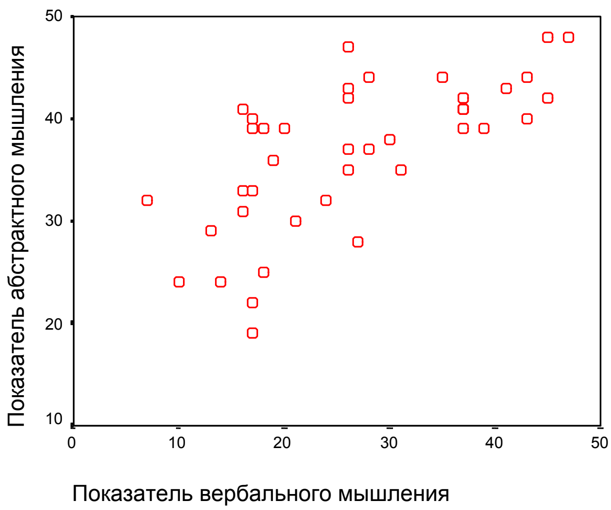


Рис. 3.3. Диаграмма рассеивания 40 пар результатов теста из табл. 3.2

Продолжим работу с пакетом *SPSS*, для того чтобы получить значения пирсоновского коэффициента корреляции. Из меню *Analyze* выберите команду *Correlate – Bivariate* и задайте исследуемые признаки. Вы можете в появившемся диалоговом окне *Bivariate Correlations* (рис. 3.4) задать вывод ранговых коэффициентов корреляции Спирмена или тау-Кендалла. Также имеется возможность задать параметры вычисления двухстороннего или одностороннего уровня значимости и выделение статистически значимых коэффициентов корреляции ($p \leq 0{,}05$).

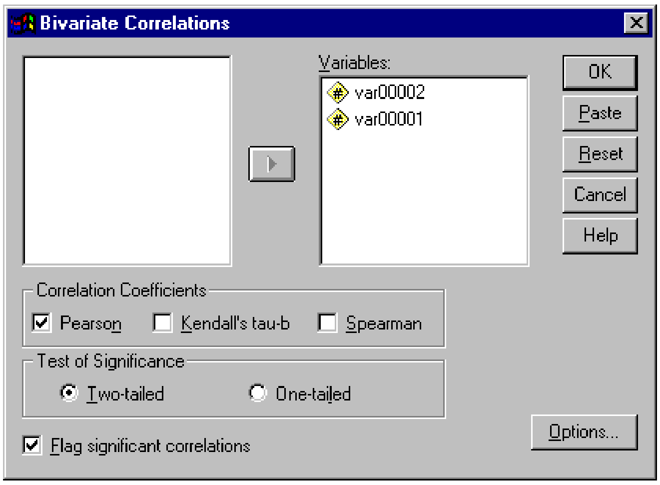


Рис. 3.4. Диалоговое окно ввода параметров для расчета коэффициента корреляции

Завершив ввод параметров вывода нажатием кнопки *OK*, на экране появится окно вывода с таблицей значений полученных коэффициентов корреляции (табл. 3.3). Для нашего примера коэффициент корреляция Пирсона равен 0,67 с нулевым уровнем значимости, что указывает на существующую статистически значимую взаимосвязь между способностями к абстрактному и вербальному мышлению измеренных у школьников младших классов.

Таблица 3.3

Корреляционная матрица по данным табл. 3.2

		абстрактное мышление	вербальное мышление
абстрактное мышление	Pearson Correlation	1,000	0,674
	Sig. (2-tailed)	0	0,000
	N	40	40
вербальное мышление	Pearson Correlation	0,674	1,000
	Sig. (2-tailed)	0,000	0
	N	40	40

3.1.2. Интерпретация коэффициентов корреляции

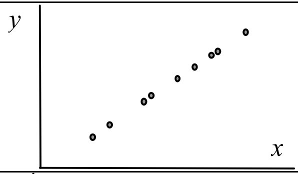
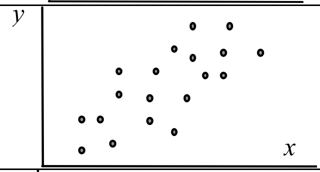
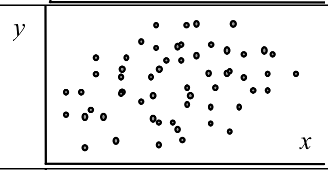
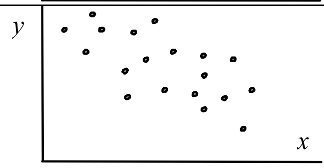
Область изменения коэффициента корреляции

Хотя это и затруднительно доказать, но r_{xy} никогда не может принять значение ни меньше –1, ни больше +1.

В табл. 3.4 перечислены различные значения r_{xy} с иллюстрациями типа *линейной* связи, которая существует между x и y для данных значений r_{xy}^2.

Таблица 3.4

Интерпретация значений r_{xy}

Величина r_{xy}	Описание линейной связи	Диаграмма рассеивания
+1	Сильная положительная связь	
Около +0,5	Слабая положительная связь	
0	Нет линейной связи, т. е. ковариация x и y = 0	
Около –0,5	Слабая отрицательная связь	
–1	Сильная отрицательная связь	

После приобретения опыта работы с реальными данными у исследователя развивается «чувство» степени связи, характеризующейся любым частным значением r, а также способность мысленно связывать диаграмму рассеивания точек и соответствующую ей приблизительную величину r. Некоторые авторы применяют описательные прилагательные к значениям r, такие, например, как «высокие» для $r = 0{,}80$ или «низкие» для $r = 0{,}20$. Однако такие выводы некорректны, поскольку статистическая достоверность корреляционной связи между исследуемыми переменными при одном и том же значении коэффициента корреляции зависит от объема выборки. Подробно данный вопрос рассматривается в [6].

Для того чтобы определить, является ли корреляционная связь между исследуемыми переменными статистически значимой, нужно вычисленное значение r_{xy} сравнить с критическим значением (см. [6, 24]) – $r_{кр}$. Если окажется, что $r_{xy} \geq r_{кр}$, то коэффициент корреляции является «значимым», а корреляционная связь – статистически достоверной.

Причинность и корреляция

Наличие корреляции двух переменных отнюдь не означает, что между ними существует причинная связь. Несмотря на то, что сосуществование (корреляцию) событий можно использовать для выявления причинных связей наряду с другими методологическими подходами, монопольное применение корреляции к анализу причинности рискованно и может вводить в заблуждение. Во-первых, даже в тех случаях, когда можно предположить существование причинной связи между двумя переменными, которые коррелированны, r_{xy} сам по себе ничего не говорит о том, вызывает ли x появление y или y вызывает появление x. Во-вторых, часто наблюдаемая связь существует благодаря другим переменным, а не двум рассматриваемым. В-третьих, взаимосвязи переменных в педагогике и общественных науках почти всегда слишком сложны, чтобы их объяснением могла служить единственная причина. Успеваемость в школе – результат многочисленных влияний, да и сама по себе она является сложным понятием, которое нельзя описать адекватно при помощи какого бы то ни было одного измерения.

Гласс Дж. и Стенли Дж. в своей монографии [6] подробно рассматривают некоторые проблемы, возникающие при попытке выявить причинные связи с помощью корреляции. Так, например, справедливо, что в США существует положительная корреляция между средним заработком преподавателей в школах и процентом выпускников, поступивших в колледж. Значит ли это, что высокооплачиваемое школьное преподавание *вызывает* появление лучше подготовленных абитуриентов колледжа? Увеличится ли процент выпускников, поступивших в колледж, если повысить плату преподавателям? Конечно, утвердительные ответы на эти вопросы не объяснить одной ассоциативной связью. Связь между двумя факторами не проста, кроме того, еще не упоминалась одна существенная переменная, которая характеризует финансовые и экономические условия жизни общества и определяет его возможность нести расходы, как по оплате преподавателей, так и по обучению в колледжах. Наряду с этим экономическая и финансовая обстановка отчасти зависит от интеллектуальных возможностей населения – другой переменной, вносящей вклад и в более высокую оплату педагогов, и в повышенную посещаемость колледжей молодежью.

Установлено, что процент «исключенных» из школ отрицательно коррелирует с числом учебников, приходящихся на ученика в библиотеках этих школ. Но здравый смысл подсказывает нам, что нагромождение книг в библиотеке не больше повлияет на число исключенных, чем наем ленивого служащего на магическое увеличение школьной библиотеки.

Многие исследователи не останавливаются на том ложном выводе, что корреляция свидетельствует, *на первый взгляд*, о причинной зависимости, а выводят также и другое заключение. Они приписывают причинной связи определенное направление. Например, предположим, что в большой группе учащихся коэффициент корреляции между тревожностью (x) и результатом теста *IQ* (y) равен –0,60. Означает ли это, что большое волнение привело к тому, что учащиеся плохо выдержали испытание, а более спокойные ученики, не травмированные страхом, оказались в состоянии успешно проявить свои способности? Этот вывод склонны делать некоторые исследователи. Но разве не столь же правдоподобно считать, что сам этот тест есть фактор, вызывающий беспокойство? Не могли ли тупые ученики бояться испытания их интеллекта, а способные найти эксперимент приятным и не вызывающим беспокойства? В данном случае вопрос в том, можно ли сказать, что x вызывает y или что y вызывает x? Обычный коэффициент корреляции между x и y не может дать ответ на этот вопрос. Без экспериментальной проверки связи сами по себе часто трудно интерпретировать. Искусный экспериментальный подход к той же самой задаче предполагал бы формирование группы тревожных учеников и сравнение их оценок с оценками контрольной группы.

Хотя корреляция прямо не указывает на причинную связь, она может служить ключом к разгадке причин. При благоприятных условиях на ее основе можно сформулировать гипотезы, проверяемые экспериментально, когда возможен контроль других влияний, помимо тех немногочисленных, которые подлежат исследованию.

Иногда отсутствие корреляции может иметь более глубокое воздействие на нашу гипотезу о причинной связи, чем наличие сильной корреляции. Нулевая корреляция двух переменных может свидетельствовать о том, что никакого влияния одной переменной на другую не существует, при условии, что мы доверяем результатам измерений и что коэффициент корреляции r Пирсона, измеряющий только частный тип связи, подходит для измерения более общего типа связи, называемой «причинной». Но все это мало помогает: требуются методы обнаружения причинных связей, а не методы иллюстрации беспричинных явлений.

Идентичные группы с различными средними

Существенная корреляция между двумя переменными – это факт, который в разных ситуациях можно объяснить по-разному. Некоторые корреляции – результат измерения причины и ее действия, например, когда x – пища, съеденная за месяц, а y – вес, приобретенный за то же время. Другие корреляции возникают при измерениях двух переменных с общей причиной или влиянием, например, когда x – успеваемость по английскому языку, а y – по общественным наукам. Иногда возникают иные корреляции, когда объединяются две различные группы, в каждой из которых x и y не имеют связи. Предположим, что девочки проявляют большую тревожность, чем мальчики, при проверке, например, по шкале выраженной тревожности Тейлора. Хорошо известно, что девочки, как правило, имеют более высокие оценки по английскому языку, по сравнению с мальчиками, особенно в средних классах. Диаграмма рассеивания тревоги и успеваемости по английскому для 15 мальчиков и 15 девочек могла быть подобна той, которая представлена на рис. 3.5.

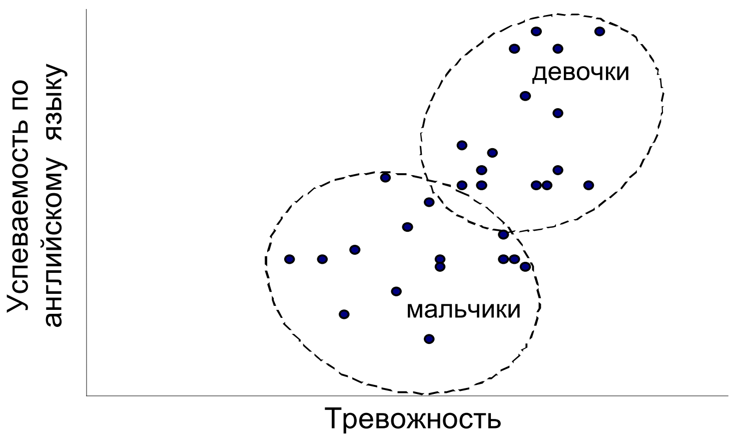


Рис. 3.5. Диаграмма рассеивания оценок тревоги и успеваемости по английскому языку для 15 мальчиков и 15 девочек [6]

На рис. 3.5 видна довольно сильная положительная связь между тревогой и успехами в английском, когда объединяются оценки мальчиков и девочек. Свидетельствует ли это о том, что тревожность (напряжение) заставляет учащегося усерднее трудиться и тем самым стимулирует большие достижения? Вовсе нет. Если бы это было так, то почему никому не удалось установить какую-либо связь между двумя переменными отдельно для мальчиков и девочек?

На рис. 3.5 видно, что ненулевые корреляции могут получиться в тех случаях, когда объединяются отдельные группы, например мальчики и девочки с различными средними. В результате такого объединения могут наблюдаться либо положительные, либо отрицательные связи.

Идентификация подгрупп с различными средними по x и y не исключает возможности корреляции x и y. Однако она допускает более рациональное объяснение того, почему r_{xy} существенно отличается от нуля.

Дополнительные замечания об интерпретации r_{xy}

Из всех способов, которыми могут быть связаны измерения двух переменных, r_{xy} оценивает только один. Величина r_{xy} представляет собой меру степени *линейной связи x и y*. Если x и y жестко линейно связаны, то точки диаграммы рассеивания будут расположены на одной прямой, как это показано в табл. 3.4. Если мы разбросаем точки на таком графике над и под прямой случайным образом и приблизительно на одинаковые расстояния, то получим различные степени линейных связей между x и y. Если точки на диаграмме рассеивания ориентируются – хотя и отклоняются случайным образом – относительно *кривой*, связь x и y может быть существенно *криволинейной.*

Из того, что r_{xy} измеряет только линейную связь между x и y, следует, что различные виды нелинейных связей x и y могут дать такие значения $r_{xy,}$ которые подозрительно близки к нулю, если интерпретировать их без учета диаграммы рассеивания.

Если известно, что x и y в общем тесно связаны линейно, то смысл r_{xy} совершенно ясен. Однако если x и y имеют некую нелинейную связь, то близкие к нулю значения r_{xy} могут быть получены даже несмотря на то, что x и y сильно связаны. Рис. 3.6 содержит две разные диаграммы рассеивания, каждая из которых имеет близкие к нулю коэффициенты корреляции.

Хотя обе диаграммы рассеивания A и B на рис. 3.6 имеют нулевые коэффициенты корреляции, в B есть существенная связь между x и y, а в A нет никакой систематической связи между ними. Одной иллюстрации на рис. 3.6, по-видимому, достаточно для предупреждения против опрометчивого вывода о том, что две переменные не *связаны* только потому, что $r_{xy} = 0$.

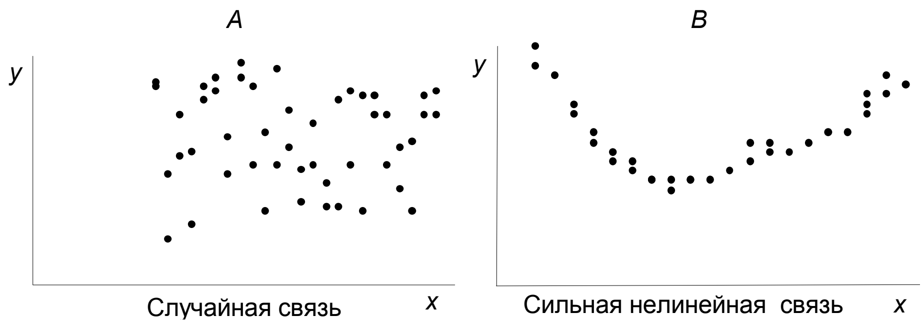


Рис. 3.6. Два примера близкой к нулю линейной корреляции

Оценки педагогических и психологических тестов часто дают «потолочные» или «подвальные» эффекты у нетипичных групп, т. е. испытания могут быть слишком легкими или слишком трудными, ибо многие получают максимальную или минимальную оценку. Диаграмма рассеивания оценок теста *A*, который характеризуется «потолочным эффектом», и теста *B* с «подвальным эффектом» могла бы быть подобна диаграмме рис. 3.7.

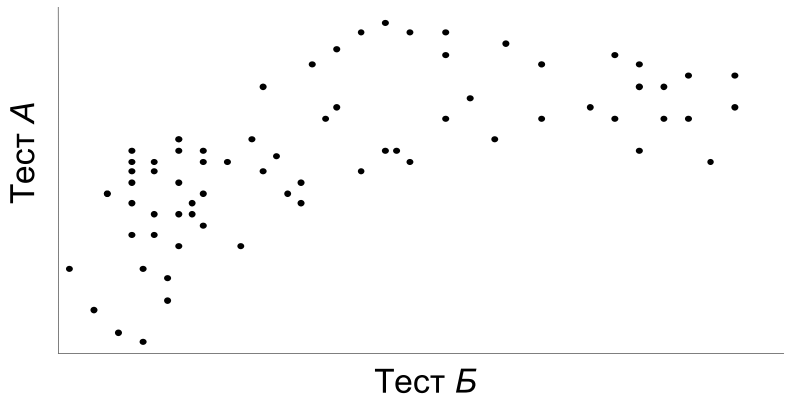


Рис. 3.7. Диаграмма рассеивания оценок для теста A (слишком простого) и теста B (слишком трудного для испытываемой группы)

Величина r_{AB} для данных рис. 3.7 невелика; вероятно, она приблизительно равна 0,30. Оказывается, что в области, для которой оба теста эквивалентны по трудности, они связаны более сильно. Считают, что если бы тест *A* был более трудным, а тест *B* – более легким без радикального изменения их содержания, то величина r_{AB} увеличилась бы. Диаграмма рассеивания для подобных измененных тестов, возможно, обладала бы меньшей нелинейностью, чем имеющаяся.

Этот пример показывает другой важный момент: степень связи между любыми двумя переменными независимо от того, как эта связь выражена – зависит от характера измерения переменных. Например, мы обычно считаем, что характеристики *веса* и *роста* довольно сильно связаны между собой у взрослых людей; но нетрудно представить себе весьма плохие способы измерения этих переменных, например измерение с помощью субъективных суждений четырехлетних детей, оценки веса и роста которых не показали бы почти никакой корреляции.

Рассмотрим на примере этих же выборок (*Таблица 3.2 [6]*) реализацию подсчета коэффициента корреляции Пирсона на интернет-ресурсе http://www.psychol-ok.ru/ . После ввода всех необходимых данных и выполнения получили результат:

Таблица 3.5

Было выполнено:
1) расчет теоретической частоты ($f_Т$);
2) подсчитана разность между эмпирической и теоретической частотой по каждому разряду;
3) определено число степеней свободы. Внесена поправка на «непрерывность» (если $\nu = 1$);
4) полученные разности возведены в квадрат;
5) полученные квадраты разностей разделены на теоретическую частоту (последний столбец);
6) полученная сумма является $x^2_{эмп}$.

№	Эмпирическая частота	Теоретическая частота	$(f_Э \cdot f_Т)$	$(f_Э \cdot f_Т)^2$	$(f_Э \cdot f_Т)^2 / f_Т$
1	19	20,91	–1,91	3,65	0,175
2	17	15,09	1,91	3,65	,242
3	32	22,65	9,35	87,42	3,86
4	7	16,35	–9,35	87,42	5,347
5	33	29,04	3,96	15,68	0,54
6	17	20,96	–3,96	15,68	0,748
7	44	41,82	2,18	4,75	0,114
8	28	30,18	–2,18	4,75	0,157
9	28	31,95	–3,95	15,6	0,488
10	27	23,05	3,95	15,6	0,677
11	35	38,34	–3,34	11,16	0,291
12	31	27,66	3,34	11,16	0,403
13	39	34,27	4,73	22,37	0,653
14	20	24,73	–4,73	22,37	0,905
15	39	32,53	6,47	41,86	1,287
16	17	23,47	–6,47	41,86	1,784
17	44	45,89	–1,89	3,57	0,078
18	35	33,11	1,89	3,57	0,108
19	44	50,54	–6,54	42,77	0,846
20	43	36,46	6,54	42,77	1,173
21	24	19,75	4,25	18,06	0,914
22	10	14,25	–4,25	18,06	1,267
23	37	37,76	–0,76	0,58	0,015
24	28	27,24	0,76	0,58	0,021
25	29	24,4	4,6	21,16	0,867
26	13	17,6	–4,6	21,16	1,202
27	40	48,21	–8,21	67,4	1,398
28	43	34,79	8,21	67,4	1,937
29	42	50,54	–8,54	72,93	1,443

Продолжение табл. 3.5

№	Эмпирическая частота	Теоретическая частота	$(f_Э \cdot f_Т)$	$(f_Э \cdot f_Т)^2$	$(f_Э \cdot f_Т)^2 / f_Т$
30	45	36,46	8,54	72,93	2
31	32	32,53	–0,53	0,28	0,009
32	24	23,47	0,53	0,28	0,012
33	48	54,02	–6,02	36,24	0,671
34	45	38,98	6,02	36,24	0,93
35	43	40,08	2,92	8,53	0,213
36	26	28,92	–2,92	8,53	0,295
37	33	28,46	4,54	20,61	0,724
38	16	20,54	–4,54	20,61	1,003
39	47	42,4	4,6	21,16	0,499
40	26	30,6	–4,6	21,16	0,692
41	38	39,5	–1,5	2,25	0,057
42	30	28,5	1,5	2,25	0,079
43	25	24,98	0,02	0	0
44	18	18,02	–0,02	0	0
45	35	35,43	–0,43	0,18	0,005
46	26	25,57	0,43	0,18	0,007
47	22	22,65	–0,65	0,42	0,019
48	17	16,35	0,65	0,42	0,026
49	40	33,11	6,89	47,47	1,434
50	17	23,89	–6,89	47,47	1,987
51	42	39,5	2,5	6,25	0,158
52	26	28,5	–2,5	6,25	0,219
53	41	33,11	7,89	62,25	1,88
54	16	23,89	–7,89	62,25	2,606
55	41	45,31	–4,31	18,58	0,41
56	37	32,69	4,31	18,58	0,568
57	37	36,6	0,4	0,16	0,004
58	26	26,4	–0,4	0,16	0,006
59	30	29,63	0,37	0,14	0,005
60	21	21,37	–0,37	0,14	0,007
61	31	27,3	3,7	13,69	0,501
62	16	19,7	–3,7	13,69	0,695
63	41	45,31	–4,31	18,58	0,41
64	37	32,69	4,31	18,58	0,568
65	42	45,89	–3,89	15,13	0,33
66	37	33,11	3,89	15,13	0,457
67	24	22,07	1,93	3,72	0,169
68	14	15,93	–1,93	3,72	0,234

Окончание табл. 3.5

№	Эмпирическая частота	Теоретическая частота	$(f_Э \cdot f_Т)$	$(f_Э \cdot f_Т)^2$	$(f_Э \cdot f_Т)^2 / f_Т$
69	43	48,79	–5,79	33,52	0,687
70	41	35,21	5,79	33,52	0,952
71	36	31,95	4,05	16,4	0,513
72	19	23,05	–4,05	16,4	0,711
73	39	33,11	5,89	34,69	1,048
74	18	23,89	–5,89	34,69	1,452
75	39	45,31	–6,31	39,82	0,879
76	39	32,69	6,31	39,82	1,218
77	39	44,15	–5,15	26,52	0,601
78	37	31,85	5,15	26,52	0,833
79	48	55,18	–7,18	51,55	0,934
80	47	39,82	7,18	51,55	1,295
Суммы	2522	2522	–	–	**59,952**

Результат: $\chi^2_{Эмп} = 59.952$

Критические значения χ^2 при v=39

v	p	
	0.05	0.01
39	**54.572**	**62.428**

Различия между двумя распределениями могут считаться достоверными, если $\chi^2_{Эмп}$ достигает или превышает $\chi^2_{0.05}$, и тем более достоверным, если $\chi^2_{Эмп}$ достигает или превышает $\chi^2_{0.01}$.

Ответ: $\chi^2_{Эмп}$ равно критическому значению или превышает его, расхождения между распределениями статистически достоверны (гипотеза H_1).

Рис. 3.8. Результаты вычисления и выводы

3.2. Непараметрические меры связи

3.2.1. Коэффициент корреляции рангов

Из непараметрических показателей связи наиболее широкое применение нашел коэффициент корреляции рангов, предложенный Спирменом [15]:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)}, \qquad (3.8)$$

где $d_i = R_x - R_y$ – разность между рангами признаков x и y; n – число парных членов ряда или объем выборки.

В основу конструкции этого показателя положены весьма простые соображения. Ранжируя попарно связанные значения признаков, можно видеть, как они распределяются относительно друг друга. Если возрастающим значениям одного признака x соответствуют возрастающие значения другого y, то между ними существует положительная связь. Если же при возрастании значений одного признака значения другого последовательно уменьшаются, это указывает на наличие отрицательной связи между ними. При отсутствии корреляции ранжированным значениям одного признака будут соответствовать самые различные значения другого.

Обозначив ранжированные значения признаков порядковыми числами 1, 2, 3, 4, …, нетрудно определить ранги этих значений и по их разностям судить о степени зависимости одного признака от изменений другого. Очевидно, при полной связи ранги коррелируемых признаков совпадут и разность между ними будет равна нулю. В таких случаях коэффициент корреляции рангов окажется равным единице. Если же признаки варьируют независимо друг от друга, то величина $\frac{6\sum d^2}{n(n^2-1)} = 1$ и коэффициент корреляции рангов будет равен нулю. Таким образом, коэффициент корреляции рангов выражается в долях единицы и может принимать значения от –1 до + 1, т. е. сопровождается положительным или отрицательным знаком.

Значимость этого показателя оценивают путем сравнения выборочного коэффициента r_s с критической точкой r_{st}, которую можно определить по формуле

$$r_{st} = \frac{1}{\sqrt{n-1}}\left(1 - \frac{m}{n-1}\right), \tag{3.9}$$

где n – объем выборки; t и m – величины, связанные, с уровнем значимости α следующим образом: для $\alpha = 5$ %, $t = 1{,}96$ и $m = 0{,}16$; для $\alpha = 1$ %, $t = 2{,}58$ и $m = 0{,}69$. Нулевую гипотезу отвергают, если эмпирически найденная величина $r_{эмп}$ превзойдет или окажется равной критическому значению $r_{кр}$ для принятого уровня значимости α и объема выборки n. Чтобы каждый раз не рассчитывать критические точки $r_{кр,}$ составлена специальная таблица (см. [6, 24]).

Приведенный способ оценки значимости выборочного $r_{эмп}$ не единственный. При $n \geq 10$ значимость эмпирического коэффициента корреляции рангов можно оценить с помощью t-критерия Стьюдента, т. е. по отношению этого показателя к своей статистической ошибке $t_{эмп} = \left| r_s \sqrt{\frac{n-2}{1-r_s^2}} \right| \geq t_{кр}$ для $n - 2$ и принятого уровня значимости α.

Рассчитывая коэффициент корреляции рангов, следует иметь в виду, что на его значении сказывается наличие групп с одинаковыми рангами, и тем сильнее, чем больше таких групп среди сопряженных значений признаков x и y. Чтобы получить более точное значение $r_{эмп}$, нужно при наличии указанных групп вносить поправку в формулу (3.8). Эту поправку, обозначаемую буквой T, прибавляют к числителю формулы, т. е.

$$r_s^* = 1 - \frac{6\sum d^2 + T}{n(n^2-1)}, \tag{3.8а}$$

где $T = V_x + V_y$, а V_x – поправка для одного признака (ряд x), V_y – для другого (ряд y). Для определения V_x и V_y составлена специальная таблица (табл. 3.6), в которой l обозначает число групп с одинаковыми рангами (столбцы), а t – число рангов в этих группах (строки).

Таблица 3.6

<table>
<tr><th>Число групп t</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th></tr>
<tr><td>2</td><td>0,5</td><td>1,0</td><td>1,5</td><td rowspan="4">2,0
8,0
20,0 40,0</td><td rowspan="4">2,5
10,0
25,0 50,0</td><td rowspan="4">3,0
12,0
30,0 60,0</td><td rowspan="4">3,5
14,0
35,0 70,0</td></tr>
<tr><td>3</td><td>2,0</td><td>4,0</td><td>6,0</td></tr>
<tr><td>4</td><td>5,0</td><td>10,0</td><td>15,0</td></tr>
<tr><td>5</td><td>10,0</td><td>20,0</td><td>30,0</td></tr>
</table>

Пример 3.2

Рассмотрим способ вычисления рангового коэффициента корреляции Спирмена в прикладном статистическом пакете STATISTICA (для работы в пакете SPSS порядок действий описан в примере 3.1). Воспользуемся данными табл. 3.7 и вычислим коэффициент корреляции рангов между шкалами тревожности и агрессивности теста MMPI.

Для того чтобы получить значения рангового коэффициента корреляции, активизируйте модуль *Nonparametric Statistics*. Загрузите файл данных. Выберите пункт *Correlations (Spearman*, *Kendall tau*, *gamma*) из предложенного меню и завершите выбор нажатием кнопки *OK*. В открывшемся диалоговом окне задайте параметры корреляционного анализа. Для того чтобы выбрать исследуемые признаки, нажмите кнопку 

. Чтобы задать вычисляемый непараметрический коэффициент корреляции, воспользуйтесь всплывающим меню 

. Имеется возможность вывести графики (гистограммы и диаграмму рассеивания), для этого нажмите кнопку 

. Всплывающее меню 

позволяет задать режим вывода информации: детальный отчет, только коэффициент корреляции или корреляционную матрицу. Мы остановимся на *Detailed report*. Задав необходимые параметры, нажмите кнопку *OK*. На экране откроется окно с результатами корреляционного анализа (рис. 3.9).

Таблица 3.7

№ исследований	Шкалы	
	Агрессивность *x*	Тревожность *y*
1	10	65
2	17	11
3	27	44
4	34	36
5	45	45
6	61	71
7	61	61
8	65	98
9	65	89
10	76	57
11	80	61
12	105	79

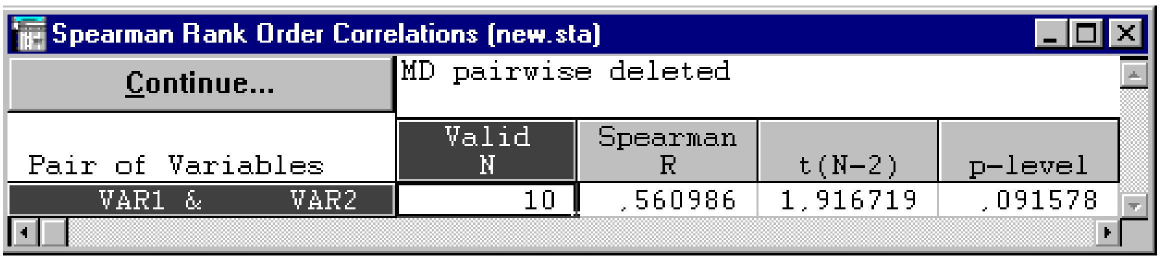


Рис. 3.9. Результаты корреляционного анализа

Итак, коэффициент корреляции Спирмена равен 0,56 с уровнем значимости 0,09. Различие между шкалами тревожности и агрессивности теста MMPI в данной группе испытуемых статистически не значимо, другими словами, существующая разница в оценках не существенна.

3.2.2. Коэффициент ассоциации

Тесноту связи между качественными признаками y и x, группируемые в четырехпольную корреляционную таблицу, измеряют с помощью коэффициента ассоциации, предложенного К. Пирсоном в 1901 г. В простейшем виде формула, по которой рассчитывают этот показатель, обозначаемый символом r_A, выглядит следующим образом [17]:

$$r_A = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \qquad (3.10)$$

Здесь a, b, c и d – численности коррелируемых групп (вариант), распределяемых по клеткам четырехпольной таблицы.

Коэффициент ассоциации, как и пирсоновский коэффициент корреляции, изменяется от –1 до + 1. Значимость выборочного коэфициента ассоциации оценивают по величине критерия Пирсона χ^2. Нулевая гипотеза сводится к предположению, что в генеральной совокупности этот показатель r_A равен нулю. H_0-гипотезу отвергают, если $\chi^2 = nr_A^2 \geq \chi^2_{\text{кр}}$ для принятого уровня значимости (α) и числа степени свободы $k = (2 - 1)(2 - 1) = 1$.

Значимость r_A можно проверить и с помощью t-критерия Стьюдента. Нулевую гипотезу отвергают, если

$$t_{\text{ф}} = \frac{r_A\sqrt{n-2}}{\sqrt{1-r_A^2}} \geq t_{\text{кр}}$$

для принятого уровня значимости (α) и числа степеней свободы $k = n - 2$.

Так как распределение вероятных значений критерия χ^2 является непрерывным, а качественные признаки дискретны, то их числовые значения не распределяются непрерывно. Учитывая эту особенность, в формулу (3.10) принято вносить поправку Йейтса на непрерывность вариации, равную половине объема выборки. Эту поправку вычитают из разности ($ad - bc$), и формула (3.10) принимает следующий вид:

$$r_A = \frac{|ad - bc| - 0{,}5n}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \qquad (3.11)$$

Пример 3.3

Два теста были предложены 100 лицам. Каждому человеку приписывалась 1, если он правильно отвечал на вопрос, и 0 в случае неверного ответа. В табл. 3.8 представлены числа лиц с каждым сочетанием верных и неверных ответов на два вопроса. Число людей, ответивших неправильно на оба вопроса, оказалось равным 64; 5 человек ответило верно на второй и неверно на первый вопрос.

Таблица 3.8

	№ 1: неверно	№ 1: верно	
№ 2: верно	$a = 5$	$b = 25$	$a + b = 30$
№ 2: неверно	$c = 64$	$d = 6$	$c + d = 70$
	$a + c = 69$	$b + d = 31$	$n = 100$

Подставляя известные значения в формулу (4.4), находим:

$$r_A = \frac{|5 \cdot 6 - 25 \cdot 64| - 0{,}5 \cdot 100}{\sqrt{69 \cdot 31 \cdot 30 \cdot 70}} = \frac{1550}{2119{,}41} = 0{,}731.$$

Полученная величина указывает на наличие тесной связи между ответами по двум тестам. Значимость этого показателя ($\chi^2_{\text{кр}} = 100(0{,}731)^2 = 53{,}44$) значительно превышает критический уровень $\chi^2_{\text{кр}} = 10{,}83$ для $\alpha = 0{,}1$ % и $\nu = 1$.

3.2.3. Коэффициент взаимной сопряженности

Для определения степени сопряженности между качественными признаками с числами вариантов, большими двух, служит коэффициент взаимной сопряженности или полихорический показатель связи, предложенный К. Пирсоном:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{\varphi^2}{\varphi^2 + 1}}, \tag{3.12}$$

где $\varphi^2 = \left(\sum_{i-1}^{n} \frac{f^2_{xy}}{\sum f_x \sum f_y} \right) - 1$ – величина, в которой f_{xy} обозначает частоты в клетках многопольной корреляционной таблицы, а Σf_x и Σf_y – суммы частот по строкам и столбцам той же таблицы; $N = \Sigma f_x + \Sigma f_y$ – общая сумма частот, или объем выборки.

Пирсоновский коэффициент взаимной сопряженности (C) имеет один существенный недостаток: его значение значительно зависит от количества вариант коррелируемых качественных признаков.

Учитывая этот недостаток, А.А. Чупров внес поправки в формулу (3.12) и получил следующее выражение [17]:

$$K^2 = \frac{\chi^2}{N\sqrt{(n_x - 1)(n_y - 1)}} = \frac{\varphi^2}{\sqrt{(n_x - 1)(n_y - 1)}}. \tag{3.13}$$

Здесь K – коэффициент взаимной сопряженности Чупрова; n_x и n_y – численность групп по строкам и столбцам многопольной таблицы; N –

объем выборки. Остальные символы объяснены выше. Нулевую гипотезу отвергают, если $\chi^2_{\text{эмп}} = N\varphi^2 \geq \chi^2_{\text{кр}}$ для принятого уровня значимости и числа степеней свободы.

Пример 3.4 [15]

Изучали зависимость между цветом волос и цветом глаз у человека. Результаты наблюдений сведены в табл. 3.9.

Таблица 3.9

Цвет глаз	Цвет волос			Всего
	Блондины	Шатены	Рыжие	
Голубые	170	80	5	255
Серые	70	152	8	230
Карие	68	340	7	415
Всего	308	572	20	900

Определим коэффициент взаимной сопряженности между этими признаками, предварительно рассчитав величину φ^2:

$$\varphi^2 = \frac{170^2}{255 \cdot 308} + \frac{80^2}{255 \cdot 572} + \frac{5^2}{255 \cdot 20} + \frac{70^2}{230 \cdot 308} + \frac{152^2}{230 \cdot 572} + \frac{8^2}{230 \cdot 20} +$$

$$+ \frac{68^2}{415 \cdot 308} + \frac{340^2}{415 \cdot 572} + \frac{7^2}{415 \cdot 20} - 1 = 1{,}205 - 1 = 0{,}205.$$

Подставляем известные значения в формулу (3.13):

$$K = \sqrt{\frac{0{,}205}{\sqrt{(3-1)(3-1)}}} = \sqrt{\frac{0{,}205}{4}} = 0{,}226.$$

Найденная величина $K = 0{,}226$ указывает на наличие слабой связи между цветом глаз и цветом волос у человека. Критерий $\chi^2_{\text{эмп}} = N\varphi^2 =$ $= 900 \cdot 0{,}205 = 184{,}5 > \chi^2_{\text{кр}} = 18{,}47$ для $\alpha = 0{,}1$ % и $\nu = (3 - 1)(3 - 1) = 4$. Так как $\chi^2_{\text{эмп}} > \chi^2_{\text{кр}}$ нулевая гипотеза отвергается на весьма высоком уровне значимости ($p < 0{,}001$).

Необходимо помнить, что правильное применение критерия χ^2 основано на требовании, чтобы в клетках корреляционной таблицы содержалось не менее пяти вариантов и общее число наблюдений не было меньше 50. Несоблюдение этих требований не гарантирует получение достаточно точных оценок генерального параметра ρ_{χ^2}, а следовательно, и правильных выводов, которые делают на основании выборочных показателей.

3.3. Выбор меры связи

При решении задачи оценки взаимосвязи между исследуемыми переменными нередко возникает вопрос: какому показателю следует отдать предпочтение? Ответ на этот вопрос не может быть однозначным. Дело в том, что параметрический пирсоновский коэффициент корреляции достаточно точно характеризует линейную связь, когда коррелируемые признаки x и y имеют нормальное или логнормальное распределение, т. е. такое, при котором не сама случайная величина, а логарифмы ее значений распределяются нормально. Коэффициент корреляции рангов характеризует корреляционную связь независимо от закона распределения. И все же, если коррелируемые признаки распределяются нормально, предпочтение следует отдавать пирсоновскому коэффициенту корреляции как более мощному показателю связи между переменными y и x по сравнению с коэффициентом Спирмена. В тех случаях, когда коррелируемые признаки не распределяются нормально, необходимо исследовать непараметрические показатели связи.

Коэффициент ранговой корреляции Спирмена и другие непараметрические показатели независимы от закона распределения, и в этом их большая ценность. Они позволяют измерять тесноту сопряженности между такими признаками, которые не поддаются непосредственному измерению, но могут быть выражены баллами или другими условными единицами, позволяющими ранжировать выборку. Ценность коэффициента корреляции рангов заключается также в том, что он позволяет быстро оценивать взаимосвязь между признаками независимо от закона распределения.

Для анализа сопряженности между различными показателями жизнедеятельности в ряде случаев целесообразно использовать не корреляционный анализ, а так называемый дискретно-динамический метод [26, 36], являющийся одним из вариантов дисперсионного анализа. В соответствии с ним обследуемую группу людей ранжируют по одному из параметров (например, рост, уровень тревожности, балльная самооценка настроения и др.) и разделяют на три статистически значимых подгруппы (не менее чем по 5–6 человек) с максимальными, средними и минимальными значениями данного параметра. Затем для каждой группы обследуемых, сформированной по признакам относительной однородности одного такого показателя (выделяемого в качестве базового), вычисляют средние арифметические и ошибки средних арифметических других (вспомогательных) показателей. При сопоставлении параметров каждый из них последовательно рассматривается как базовый и затем как вспомогательный. Наличие статистически значимых (например, для $p < 0{,}05$) различий между значениями вспомогательных показателей, соответствующих полярным различиям базовых параметров, свидетельствует о наличии зависимости между ними. Мера линейности этой связи определяется по результатам вычислений, при которых вспомогательный параметр рассматривается как базовый, а базовый – как вспомогательный. В отличие от корреляционного анализа при таком методе математической обработки результатов исследования не утрачиваются численные значения анализируемых параметров и одновременно устанавливается характер связи между ними.

В табл. 3.10 приведены сведения о том, какой из коэффициентов корреляции следует использовать в зависимости от того, в каких шкалах измерены исследуемые переменные.

Таблица 3.10

Выбор меры связи

Тип шкалы		Метод корреляционного анализа	Примечания	Литература
Интервальная или отношений	Интервальная или отношений	Коэффициент корреляции Пирсона	Обе переменные x и y должны быть нормально распределенными	[6], [15], [27]
Интервальная или отношений	Интервальная или отношений	1. Ранговые коэффициенты корреляции (Спирмена и Кендалла) 2. Коэффициент Фехнера	Хотя бы одна из переменных x или y не подчиняется нормальному закону распределения	[6], [15]
Интервальная или отношений	Порядковая	Ранговые коэффициенты корреляции (Спирмена и Кендалла)	–	[6], [15]
Порядковая	Порядковая	Ранговые коэффициенты корреляции (Спирмена и Кендалла)	–	[6], [15]

Окончание табл. 3.8

Тип шкалы		Метод корреляционного анализа	Примечания	Литература
Интервальная или отношений	Дихотомическая	Биссериальный коэффициент корреляции	–	[15]
Ранговая	Дихотомическая	Рангово-биссериальный коэффициент корреляции	–	[6]
Дихотомическая	Дихотомическая	1. Коэффициент ассоциации. 2. Коэффициент корреляции знаков	–	[6], [15], [26]
Номинальная	Номинальная	Коэффициент взаимной сопряженности	–	[6], [15], [26]

Глава 4
ОЦЕНКА ВЛИЯНИЯ ФАКТОРОВ НА ОЦЕНИВАЕМЫЙ ПРИЗНАК. ДИСПЕРСИОННЫЙ АНАЛИЗ

4.1. Понятие дисперсионного анализа

Дисперсионный анализ – это анализ изменчивости признака под влиянием каких-либо контролируемых переменных факторов. В зарубежной литературе дисперсионный анализ часто обозначается как *ANOVA*, что переводится как анализ вариативности (*Analysis of Variance*). Автором метода является Р.А. Фишер.

Задача дисперсионного анализа состоит в том, чтобы из общей вариативности (изменчивости) признака вычленить три вида вариативности [31]:

а) вариативность, обусловленную *действием каждой* из исследуемых независимых переменных;
б) вариативность, обусловленную *взаимодействием* исследуемых независимых переменных;
в) *случайную* вариативность, обусловленную всеми другими неизвестными переменными.

Вариативность, обусловленная действием исследуемых переменных и их взаимодействием, соотносится со случайной вариативностью. Показателем этого соотношения является критерий F Фишера.

В формулу расчета критерия F входят оценки дисперсий, т. е. параметров распределения признака, поэтому критерий F является параметрическим критерием.

Чем в большей степени вариативность признака обусловлена исследуемыми переменными (факторами) или их взаимодействием, тем выше эмпирические значения критерия F.

В дисперсионном анализе исследователь исходит из предположения, что одни переменные могут рассматриваться как причины, а другие – как следствия. Переменные первого рода считаются факторами, а переменные второго рода – результативными признаками. В этом отличие дисперсионного анализа от прямолинейного корреляционного анализа, в котором мы исходим из предположения, что изменения одного признака просто сопровождаются определенными изменениями другого.

В дисперсионном анализе возможны два принципиальных пути разделения всех исследуемых переменных на независимые переменные (факторы) и зависимые переменные (результативные признаки). Ниже приведены несколько примеров из монографии Сидоренко Е.В. [24], иллюстрирующих оба подхода.

Первый путь состоит в том, что мы совершаем какие-либо воздействия на испытуемых или учитываем какие-либо, не зависящие от нас воздействия на них, и именно эти воздействия считаем независимыми переменными, или факторами, а исследуемые признаки рассматриваем как зависимые переменные, или результативные признаки. Например, возраст испытуемых или способ предъявления им информации считаем факторами, а обучаемость или эффективность выполнения задания – результативными признаками.

Второй путь предполагает, что мы, не совершая никаких воздействий, считаем, что при разных уровнях развития одних психологических признаков другие проявляются тоже по-разному. По тем или иным причинам мы решаем, что одни признаки могут рассматриваться скорее как факторы, а другие – как результат действия этих факторов. Например, уровень интеллекта или мотивации достижения начинаем считать факторами, а профессиональную компетентность или социометрический статус – результативными признаками.

Второй путь весьма уязвим для критики. Допустим, мы предположили, что настойчивость – значимый фактор учебной успешности студентов. Мы принимаем настойчивость за воздействующую переменную (фактор), а учебную успешность – за результативный признак. Против этого могут быть выдвинуты сразу же два возражения. Во-первых, успех может стимулировать настойчивость; во-вторых, как, собственно, измерялась настойчивость? Если она измерялась с помощью метода экспертных оценок, а экспертами были соученики или преподаватели, которым известна учебная успешность испытуемых, то не исключено, что это оценка настойчивости будет зависеть от известных экспертам показателей успешности, а не наоборот.

На рис. 4.1 представлены два варианта диаграмм рассеивания показателей учебной успешности в зависимости от уровня развития кратковременной памяти. Из рис. 4.1, *а* мы видим, что при низком уровне развития кратковременной памяти оценки по английскому языку, похоже, несколько ниже, чем при среднем, а при высоком уровне выше, чем при среднем. Похоже, что кратковременная память может рассматриваться как фактор успешности овладения английским языком. С другой стороны рис. 4.1, *б* свидетельствует о том, что успешность в чистописании вряд ли так же определенно зависит от уровня развития кратковременной памяти.

О том, верны ли наши предположения, мы сможем судить только после вычисления эмпирических значений критерия F.

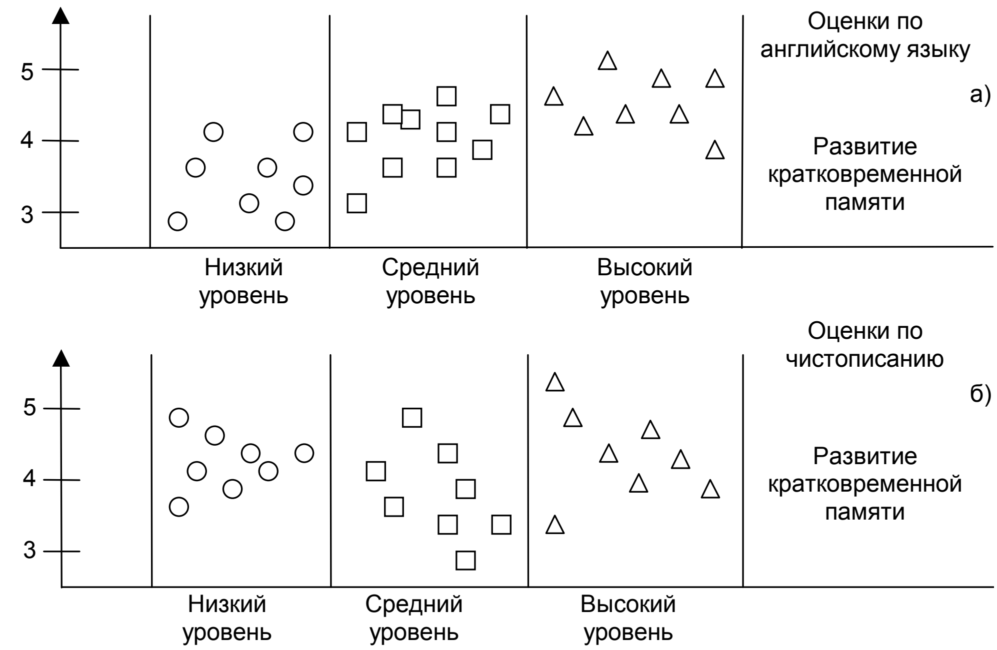


Рис. 4.1. Диаграммы рассеивания индивидуальных средних оценок по английскому языку а) и чистописанию б) у учеников с низким, средним и высоким уровнями развития кратковременной памяти [31]

Низкий, средний и высокий уровни развития кратковременной памяти можно рассматривать как градации фактора кратковременной памяти.

Нулевая гипотеза (H_0) в дисперсионном анализе формулируется следующим образом: средние величины исследуемого результативного признака при всех градациях фактора одинаковы.

Альтернативная гипотеза (H_1): средние величины результативного признака при разных градациях исследуемого фактора различны.

В зарубежных руководствах чаще говорят о переменных, действующих в разных условиях, а не о факторах и их градациях [24].

Схема дисперсионного анализа применима и в тех случаях, когда градации фактора представляют собой номинативную шкалу, т. е. отличаются лишь качественно. Например, градациями фактора могут быть: параллельные формы экспериментальных заданий; цвет окраски стимулов; жанр музыкальных произведений, сопровождающих процесс работы; традиционные или специально подобранные православные тексты в сеансах аутогенной тренировки; разные формы заболевания; разные экспериментаторы; разные психотерапевты и т. д. [24].

Если градации фактора различаются лишь качественно, их иногда называют *условиями* действия фактора или переменной. Например, действие аутогенной тренировки при условии использования текстов православных молитв или эффективность психокоррекционных воздействий при разных формах хронических заболеваний у детей [24].

Следует отметить, что понятие «фактор» используется еще в одном статистическом методе – факторном анализе, где в отличие от дисперсионного анализа слово «фактор» обозначает новую «обобщенную» переменную (подробно метод факторного анализа рассмотрен в главе 5).

Экспериментальные данные, представленные по градациям фактора, называются дисперсионным комплексом; данные, относящиеся к отдельным градациям, – ячейками комплекса.

Дисперсионный анализ позволяет нам констатировать изменение признака, но при этом не указывает *направление* этих изменений.

Подобного рода задачи, как мы помним, позволяют решать непараметрические методы сравнения выборок или условий измерения, а именно H-критерий Крускала–Уоллиса, изложенный в главе 2. Однако это касается только тех задач, в которых исследуется действие *одного* фактора, или *одной* переменной. Задачи однофакторного дисперсионного анализа, действительно, могут эффективным образом решаться с помощью непараметрических методов. Метод дисперсионного анализа становится незаменимым, только когда мы исследуем одновременное действие *двух* (или более) факторов, поскольку он позволяет выявить *взаимодействие* факторов в их влиянии на один и тот же результативный признак.

В данном учебном пособии рассматривается только однофакторный дисперсионный анализ. Читателям, желающим познакомиться с методом двухфакторного дисперсионного анализа, можно порекомендовать познакомиться с соответствующими разделами с [6], [15], [26].

Рассмотрим простейший случай дисперсионного анализа, когда исследуется действие только *одной* переменной (одного фактора). Исследователя интересует, как изменяется определенный признак в разных условиях действия этой переменной. Например, как изменяется время решения задачи при разных условиях мотивации испытуемых (низкой, средней, высокой) или при разных способах предъявления задачи (устно, письменно, в виде текста с графиками и иллюстрациями), в разных условиях работы с задачей (в одиночестве, в одной комнате с экспериментатором, в одной комнате с экспериментатором и другими испытуемыми) и т. п. В первом случае переменной, влияние которой исследуется, является мотивация, во втором – степень наглядности, в третьем – фактор публичности.

4.2. Подготовка данных к дисперсионному анализу

1. Создание комплексов. Лучше всего для каждого испытуемого создать отдельную карточку, куда были бы занесены данные по всем исследованным признакам. Дело в том, что в процессе анализа у исследователя могут измениться гипотезы. Потребуется создавать, быть может, не один, а множество дисперсионных комплексов, различающихся как по факторам, так и по результативным признакам. Карточки помогут нам быстро создавать новые дисперсионные комплексы. Благодаря карточкам мы сразу увидим, равномерно ли распределяются данные по градациям в случае, если за фактор мы решили принять один из исследованных психологических признаков. С помощью карточек мы можем помочь себе выделить три, четыре или более градаций этого фактора, например уровни мотивации, настойчивости, креативности и др.

2. Уравновешивание комплексов. Комплекс, в котором каждая ячейка представлена одинаковым количеством наблюдений, называется равномерным. Равномерность комплекса позволяет нам обойти требование равенства дисперсий в каждой из ячеек комплекса.

Равномерные комплексы позволяют также избежать значительных трудностей, которые неизбежно возникают при обсчете неравномерных, или неортогональных, комплексов. В настоящем руководстве приведены алгоритмы расчета лишь для равномерных комплексов. С методами обсчета неравномерных комплексов можно ознакомиться в [26].

В случае если в разных градациях комплекса оказалось неравное количество наблюдений, необходимо отсеять некоторые из них. Если в комплексе со связанными выборками кто-либо из испытуемых не был подвергнут одному из условий действия переменной (градаций фактора), то его данные исключаются. Если же комплекс включает независимые выборки, каждая из которых была подвергнута определенному условию воздействия (градации фактора), то «лишние» испытуемые в какой-либо из ячеек комплекса отсеиваются путем случайного выбора необходимого количества карточек.

3. Проверка нормальности распределения результативного признака. Дисперсионный анализ относится к группе параметрических методов, поэтому его следует применять только тогда, когда известно или доказано, что распределение признака является нормальным [25]. Строго говоря, перед тем как применять дисперсионный анализ, мы должны убедиться в нормальности распределения результативного признака. Нормальность распределения результативного признака можно проверить путем расчета показателей асимметрии и эксцесса и сопоставления их с критическими значениями (подробно см. главу 2).

При использовании трудоемких в вычислительном плане методов, к которым относится и дисперсионный анализ, наиболее эффективен компьютерный анализ данных. Так, например, однофакторный и двухфакторный дисперсионный анализ экспериментальных данных может быть проведен с помощью прикладных статистических пакетов *STATISTICA*, *SPSS* и *STATGRAPHICS.*

В последующих параграфах рассмотрен метод однофакторного анализа в двух вариантах:

а) для дисперсионных комплексов, представляющих данные одной и той же выборки испытуемых, подвергнутой влиянию разных условий (разных градаций фактора) (см. п. 4.3);
б) для дисперсионных комплексов, в которых влиянию разных условий (градаций фактора) были подвергнуты разные выборки испытуемых (см. п. 4.4).

Первый вариант называется однофакторным дисперсионным анализом для связанных выборок, второй – для несвязанных выборок.

Все предложенные алгоритмы расчетов предназначены для равномерных комплексов, где в каждой ячейке представлено одинаковое число наблюдений.

4.3. Однофакторный дисперсионный анализ для несвязанных выборок

Метод однофакторного дисперсионного анализа применяется в тех случаях, когда исследуются изменения результативного признака под влиянием изменяющихся условий или градаций какого-либо фактора. В данном варианте метода влиянию каждой из градаций фактора подвергаются *разные* выборки испытуемых. Градаций фактора должно быть не менее трех.

Непараметрическим вариантом этого вида анализа является H-критерий Крускала–Уоллиса (гл. 2).

Работу начинаем с того, что представляем полученные данные в виде столбцов индивидуальных значений. Каждый из столбцов соответствует тому или иному из изучаемых условий (см. пример в табл. 4.1).

После этого нам нужно просуммировать индивидуальные значения по столбцам и суммы возвести в квадрат.

Суть метода состоит в том, чтобы сопоставить сумму этих возведенных в квадрат сумм с суммой квадратов всех значений, полученных во всем эксперименте.

Гипотезы в однофакторном дисперсионном анализе формулируются следующим образом:

H_0: Различия между градациями фактора (разными условиями) являются не более выраженными, чем случайные различия внутри каждой группы.

H_1: Различия между градациями фактора (разными условиями) являются более выраженными, чем случайные различия внутри каждой группы.

Ограничения метода

1. Однофакторный дисперсионный анализ требует не менее трех градаций фактора и не менее двух испытуемых в каждой градации.

2. Должно соблюдаться правило равенства дисперсий в каждой ячейке дисперсионного комплекса. Условие равенства дисперсий выполняется при использовании предлагаемой схемы расчета за счет выравнивания количества наблюдений в каждом из условий (градаций).
3. Результативный признак должен быть нормально распределен в исследуемой выборке.

Таблица 4.1

Расчет основных величин

Обозначение	Расшифровка обозначения
T_c	сумма индив. значений по каждому из условий
$\sum(T_c^2)$	сумма квадратов суммарных значений по каждому из условий
c	количество условий (градаций фактора)
n	количество испытуемых в каждой группе (в каждом из условий)
N	общее количество индивидуальных значений
$\left(\sum x_i\right)^2$	квадрат общей суммы индивидуальных значений
$\frac{\left(\sum x_i\right)^2}{N}$	константа, которую нужно вычесть из каждой суммы квадратов
x_i	каждое индивидуальное значение
$\sum(x_i)^2$	сумма квадратов индивидуальных значений

Таблица 4.2

Последовательность операций

Операция	Формула расчета
1. Подсчитать $SS_{\text{эмп}}$	$SS_{\text{эмп}} = \frac{1}{n}\sum T_c^2 - \frac{1}{N}\left(\sum x_i\right)^2$
2. Подсчитать $SS_{\text{общ}}$	$SS_{\text{общ}} = \sum x_i^2 - \frac{1}{N}\left(\sum x_i\right)^2$
3. Подсчитать случайную (остаточную) величину $SS_{\text{сл}}$	$SS_{\text{сл}} = SS_{\text{общ}} - SS_{\text{эмп}}$
4. Определить число степеней свободы	$df_{\text{эмп}} = c - 1$ $df_{\text{общ}} = N - 1$ $df_{\text{сл}} = df_{\text{общ}} - df_{\text{эмп}}$
5. Разделить каждую SS на соответствующее число степеней свободы	$MS_{\text{эмп}} = SS_{\text{эмп}} / df_{\text{эмп}}$ $MS_{\text{общ}} = SS_{\text{общ}} / df_{\text{общ}}$ $MS_{\text{сл}} = SS_{\text{сл}} / df_{\text{сл}}$

Окончание табл. 4.2

Операция	Формула расчета
6. Подсчитать значение $F_{эмп}$	$F_{эмп} = MS_{эмп} / MS_{сл}$
7. Определить критические значения F по табл. XVII приложения	Для $df_1 = 2$ $df_2 = 15$
8. Сопоставить эмпирическое и критические значения F	При $F_{эмп} \geq F_{кр}$ H_0 отклоняется

Пример 4.1

Три различные группы из шести испытуемых получили списки из десяти слов. Первой группе слова предъявлялись с низкой скоростью – одно слово в пять секунд, второй группе со средней скоростью – одно слово в две секунды и третьей группе с большой скоростью – одно слово в секунду. Было предсказано, что показатели воспроизведения будут зависеть от скорости предъявления слов. Результаты представлены в табл. 4.3.

Поскольку сопоставляются разные группы, любые различия в показателях между разными условиями предъявления слов – это в то же время различия между группами испытуемых. Однако всякие различия между испытуемыми *внутри* каждой группы объясняются какими-то другими, не относящимися к делу переменными, будь то индивидуальные различия между отдельными испытуемыми или неконтролируемые факторы, заставляющие их реагировать различным образом.

Таблица 4.3

Количество воспроизведенных слов [20]

№ испытуемого	Группа 1: низкая скорость	Группа 2: средняя скорость	Группа 3: высокая скорость
1	8	7	4
2	7	8	5
3	9	5	3
4	5	4	6
5	6	6	2
6	8	7	4

Сформулируем гипотезу:

H_0: Различия в объеме воспроизведения слов *между* группами являются более выраженными, чем случайные различия *внутри* каждой группы.

Воспользуемся статистическим пакетом *STATISTICA*, для того чтобы провести однофакторный дисперсионный анализ изменения объема воспроизведения слов при разной скорости их предъявления. Исходными данными будут две переменные: первая это количество воспроизведенных слов, а вторая – номер группы или градации фактора. В модуле *ANOVA/MANOVA* в меню *Analysis* вызовем стартовую панель *Startup Panel* и зададим исходные данные, нажав кнопку 

. После того как выбор сделан, щелкните **OK**. Откроется окно с результатами однофакторного дисперсионного анализа (рис. 4.2).

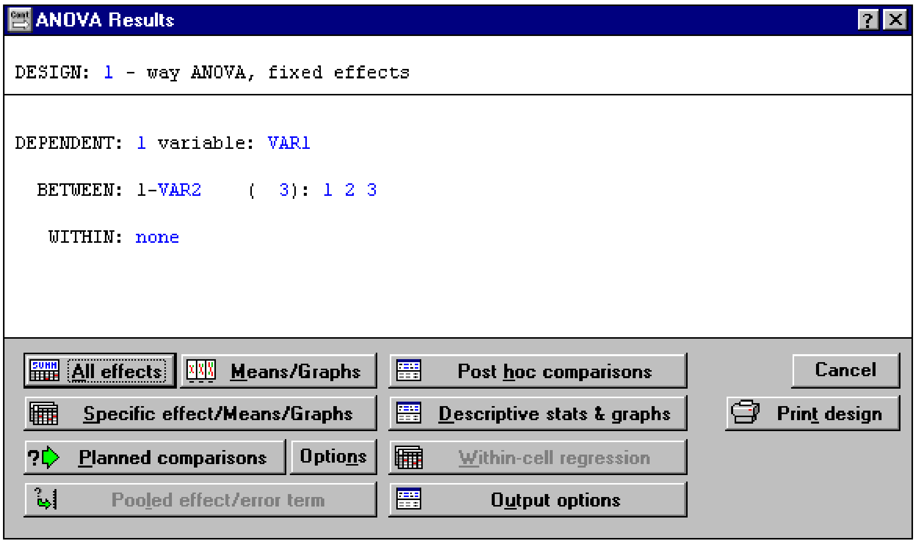


Рис. 4.2. Окно с результатами однофакторного дисперсионного анализа

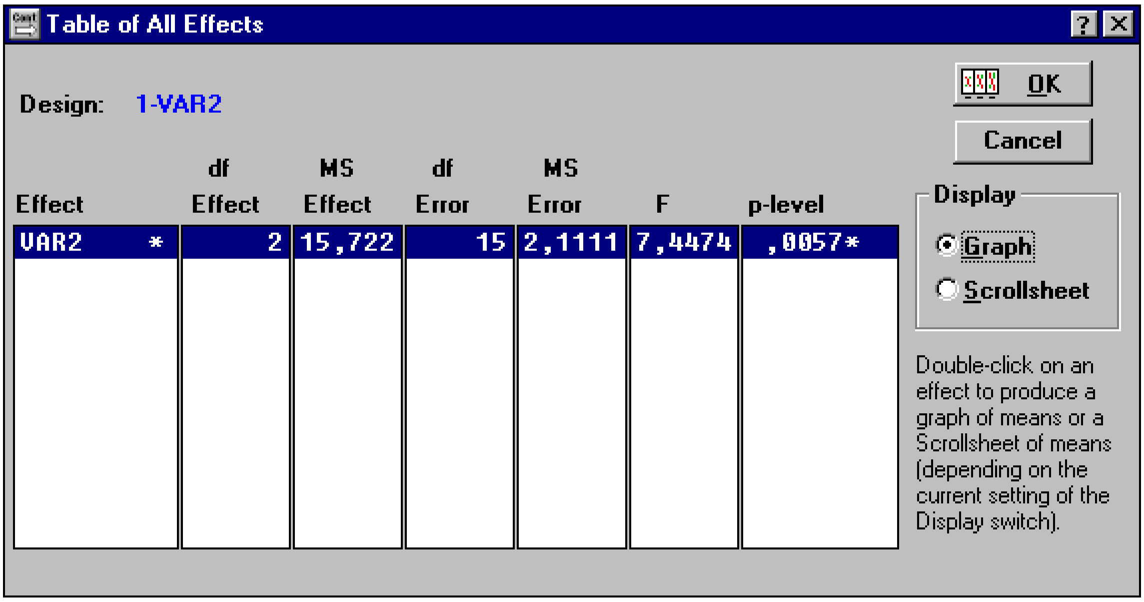


Рис. 4.3. Окно результатов

Инициируйте кнопку 

. Откроется окно (рис. 4.3) с подсчитанными статистиками критерия F и соответствующий ему уровень значимости. Для нашего примера уровень значимости равен 0,0057, что свидетельствует о статистически значимом различии между группами.

Чтобы получить кривую изменения объема воспроизведения слов при разной скорости их предъявления (рис. 4.4) в открывшемся окне в блок *Display* отметьте параметр *Graph* и нажмите кнопку *OK*. Если выделить параметр Scrollsheet, на экран будет выведена таблица со средними значениями каждой группы.

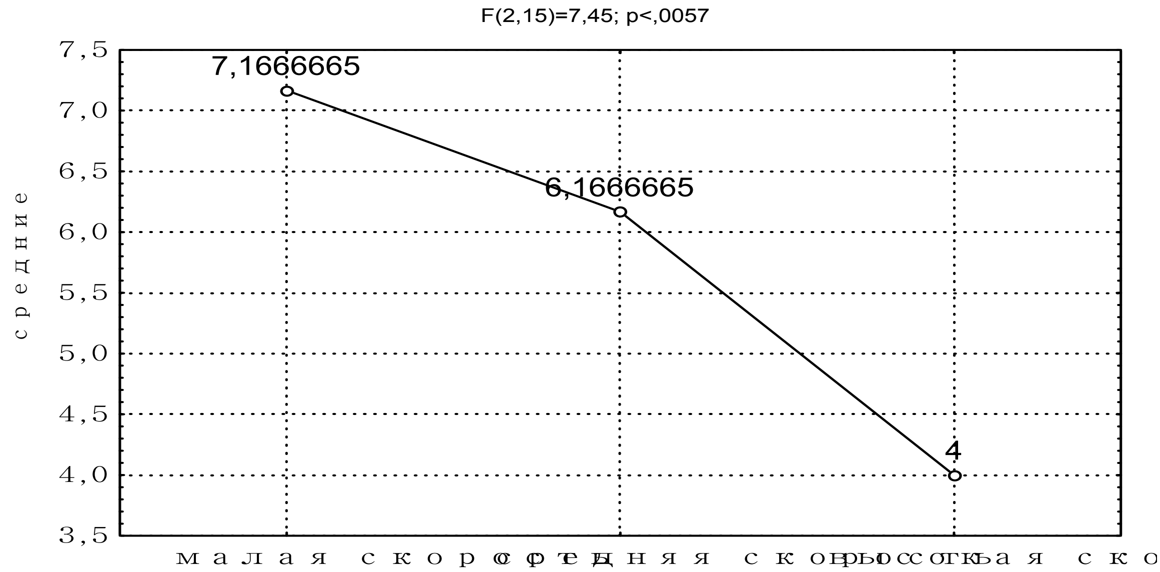
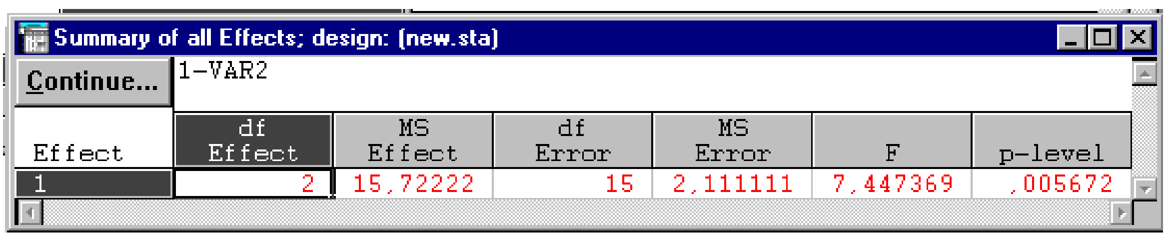


Рис. 4.4. Ломанная изменения объема воспроизведения при повышении скорости предъявления слов

Метод дисперсионного анализа позволяет определить, что перевешивает – тенденция, выраженная этой кривой, или вариативность признака внутри групп, которая на графике схематически можно отобразить в виде диапазонов изменения признака от минимального значения к максимальному значению в каждой группе.



Summary of all Effects; design: (new.sta)

Continue... 1-VAR2

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	15,72222	15	2,111111	7,447369	,005672

Рис. 4.5. Таблица результатов дисперсионного анализа

Чтобы получить сводную таблицу вычисленных статистик, нажмите кнопку 

, вернитесь в окно *ANOVA Results* и нажмите кнопку . На экране откроется таблица со статистиками (рис. 4.5). В ней приведены следующие величины: степень свободы для признака (*df Effect*), усредненная вариативность признака (*MS Effect)*, обусловленная действием исследуемого фактора, степень свободы для неучтенного фактора (*df Error*), усредненная вариативность, обусловленная неучтенными факторами (*Ms Error*), *F* – статистика и соответствующий ей уровень значимости (*p-level*).

Вывод: Принимается H_0. Различия в объеме воспроизведения слов между группами являются более выраженными, чем случайные различия внутри каждой группы ($p < 0,01$). Итак, скорость предъявления слов влияет на объем их воспроизведения.

Предположим, что показатель воспроизведения при самой высокой скорости предъявления слов (условие 3) гораздо ниже соответствующих показателей при средней и низкой скорости. Для этого продолжим анализ в диалоговом окне результатов ANOVA (рис. 4.2), нажав кнопку , что позволит сравнить разницу средних. В следующем окне нажмите кнопку **OK**. Далее раскроется еще одно окно (рис. 4.6), где выберем кнопку , завершив выбор нажатием кнопки **OK**. Данный тест в качестве результата выдает таблицу с матрицей уровней значимости, согласно которым можно делать вывод о статистически значимой разности средних.

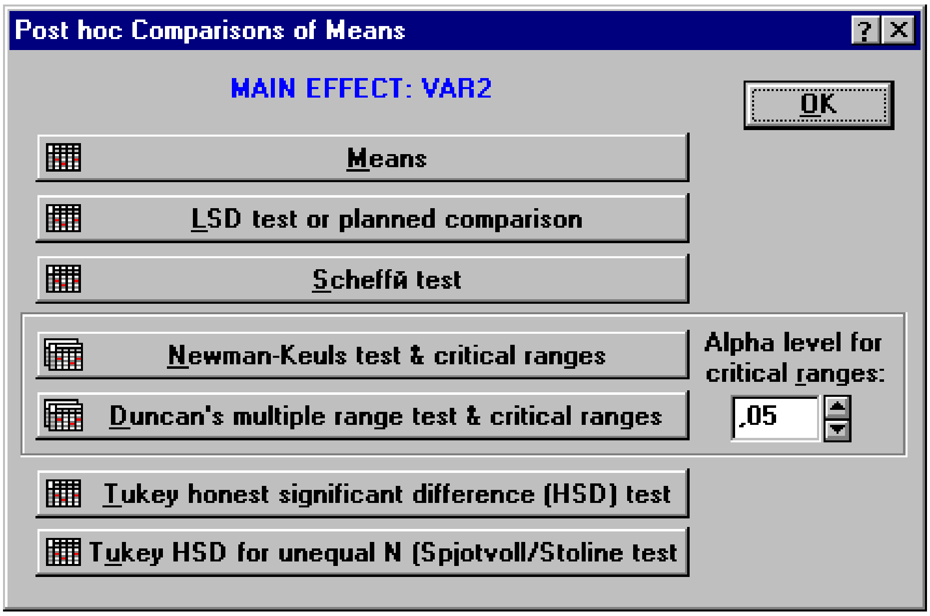


Рис. 4.6. Окно углубленного анализа результатов ANOVA

Полученные уровни значимости (р < 0,05) показывают значимое различие средних: между первой и третьей и второй и третьей группой.

4.4. Однофакторный дисперсионный анализ для связанных выборок

Метод дисперсионного анализа для связанных выборок применяется в тех случаях, когда исследуется влияние разных градаций фактора или разных (не менее трех) условий на одну *и ту же* выборку испытуемых.

Непараметрический вариант этого вида анализа – критерий Фридмана χ_r^2, подробно изложенный в [20].

В данном случае различия между испытуемыми – возможный самостоятельный источник различий. Фактор индивидуальных различий может оказаться более значимым, чем фактор изменения экспериментальных условий. Поэтому в данном случае учитывается еще одна величина – сумма квадратов сумм индивидуальных значений испытуемых.

На рис. 4.7 представлена кривая изменения времени решения анаграмм разной длины: четырехбуквенной, пятибуквенной и шестибуквенной (см. пример 4.2). Это пример типичной задачи, которую можно решать методом однофакторного дисперсионного анализа. Однофакторный дисперсионный анализ для связанных выборок в данном случае позволяет определить, что перевешивает – тенденция, выраженная этой кривой, или индивидуальные различия, диапазон которых представлен на графике в виде вертикальных линий – от минимального до максимального значения.

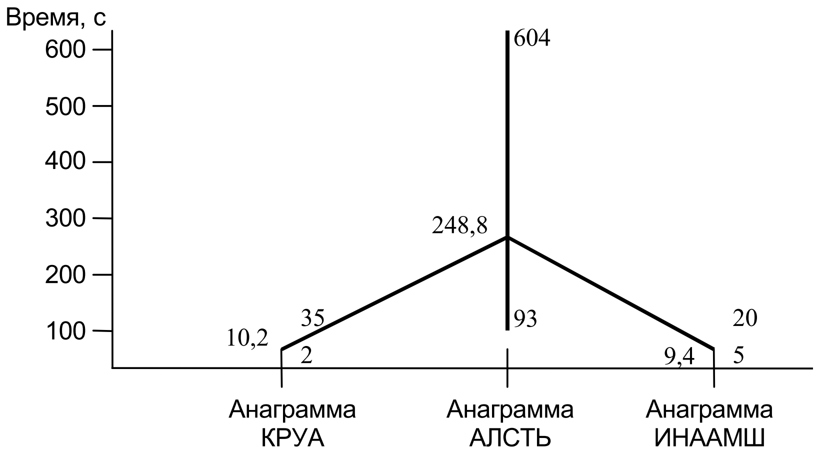


Рис. 4.7. Изменение времени работы над разными анаграммами у пяти испытуемых: вертикальными линиями отображены диапазоны изменчивости признака в разных условиях от минимального значения (снизу) до максимального значения (сверху) [20]

Ограничения метода

1. Дисперсионный анализ для связанных выборок требует не менее *трех* градаций фактора и не менее *двух* испытуемых, подвергшихся воздействию каждой из градаций фактора.

2. Результативный признак должен быть нормально распределен в исследуемой выборке.

Таблица 4.4

Расчет основных величин

Обозначение	Расшифровка обозначения
T_c	суммы индивидуальных значений по каждому из условий
$\sum(T_c^2)$	сумма квадратов суммарных значений по каждому из условий
c	количество условий (градаций фактора)
n	количество испытуемых в каждой группе (в каждом из условий)
N	общее количество индивидуальных значений
$\left(\sum x_i\right)^2$	квадрат общей суммы индивидуальных значений
$\frac{\left(\sum x_i\right)^2}{N}$	константа, которую нужно вычесть из каждой суммы квадратов
x_i	каждое индивидуальное значение
$\sum(x_i)^2$	сумма квадратов индивидуальных значений

Таблица 4.5

Последовательность операций

Операция	Формула расчета
1. Подсчитать $SS_{эмп}$	$SS_{эмп} = \frac{1}{n}\sum T_c^2 - \frac{1}{N}\left(\sum x_i\right)^2$
1. Подсчитать $SS_{исп}$	$SS_{исп} = \frac{1}{c}\sum T_u^2 - \frac{1}{N}\left(\sum x_i\right)^2$
2. Подсчитать $SS_{общ}$	$SS_{общ} = \sum x_i^2 - \frac{1}{N}\left(\sum x_i\right)^2$
3. Подсчитать случайную (остаточную) величину $SS_{сл}$	$SS_{сл} = SS_{общ} - SS_{эмп} - SS_{исп}$
4. Определить число степеней свободы	$df_{эмп} = c - 1$; $df_{исп} = n - 1$ $df_{общ} = N - 1$; $df_{сл} = df_{общ} - df_{эмп} - df_{исп}$
5. Разделить каждую SS на соответствующее число степеней свободы	$MS_{эмп} = SS_{эмп} / df_{эмп}$ $MS_{исп} = SS_{исп} / df_{исп}$ $MS_{сл} = SS_{сл} / df_{сл}$
6. Подсчитать значение $F_{эмп}$	$F_{эмп} = MS_{эмп} / MS_{сл}$; $F_{исп} = MS_{исп} / MS_{сл}$
7. Определить критические значения F по табл. XVII приложения и сопоставить эмпирическое и критические значения F	При $F_{факт} < F_{кр}$ H_0 принимается При $F_{исп} \geq F_{кр}$ H_0 отклоняется

Пример 4.2

Группа из пяти испытуемых была обследована с помощью трех экспериментальных заданий, направленных на изучение интеллектуальной настойчивости [20]. Каждому испытуемому индивидуально предъявлялись последовательно три одинаковые анаграммы: четырехбуквенная, пятибуквенная и шестибуквенная (табл. 4.6). Можно ли считать, что фактор длины анаграммы влияет на длительность попыток ее решения?

Таблица 4.6

Длительность попыток решения анаграмм (с)

№ испытуемого	Условие 1: четырехбуквенная анаграмма	Условие 2: пятибуквенная анаграмма	Условие 3: шестибуквенная анаграмма
1	5	235	7
2	7	604	20
3	2	93	5
4	2	171	8
5	35	141	7

В данном случае возможны два варианта гипотез.

Вариант А:

H_0(А): Различия в длительности попыток решения анаграмм разной длины являются более выраженными, чем различия, обусловленные случайными причинами.

Вариант Б:

H_0(Б): Индивидуальные различия между испытуемыми являются более выраженными, чем различия, обусловленные случайными причинами.

Проверим условие нормальности по величине показателей асимметрии и эксцесса:

$$As = 2{,}18;\ m_A = 0{,}632;$$

$$Ex = 4{,}17;\ m_{Ex} = 1{,}264.$$

$$t_{As} = \frac{|2{,}18|}{0{,}632} = 3{,}45;\ t_{Ex} = \frac{|4{,}17|}{1{,}264} = 3{,}30.$$

Таким образом, распределение показателей пяти человек, составляющих дисперсионный комплекс, несколько отличается от нормального: $t_{As} > 3$; $t_{Ex} > 3$. Однако в целом по выборке, из которой были отобраны эти пять человек, распределение нормальное: $n = 22$; $As = 1{,}26$; $m_{As} = 0{,}522$; $t_{As} = 2{,}41 < 3$; $Ex = 2{,}29$; $m_{Ex} = 1{,}044$; $t_{Ex} = 2{,}19 < 3$.

Согласно [20] можно удовлетвориться тем, что в выборке в целом результативный признак распределен нормально. Случайно отобранные пять человек распределением своих оценок демонстрируют некоторое отклонение. Однако если бы мы выбирали испытуемых таким образом, чтобы распределение их оценок подчинялось нормальному закону, это нарушило бы правило рандомизации – случайности отбора объектов без учета значений результативного признака при отборе.

Далее рассчитаем статистики критерия *F*. Для этого воспользуемся возможностями пакета *STATISTICA*. Исходными данными будут три переменные: первая – это длительность попыток решения анаграмм, вторая – номер группы, а третья – номер испытуемого. В модуле *ANOVA/MANOVA* в меню *Analysis* вызовем стартовую панель *Startup Panel* и зададим исходные данные, нажав кнопку 

. В качестве факторов укажем номер группы и номер испытуемого. Для задания определяющего фактора нажмите кнопку 

. В качестве первого определяющего фактора в смешанной модели выберите номер группы и нажмите *OK*. Завершите ввод данных в стартовой панели нажатием кнопки *OK*. Откроется окно с результатами однофакторного дисперсионного анализа (рис. 4.8).

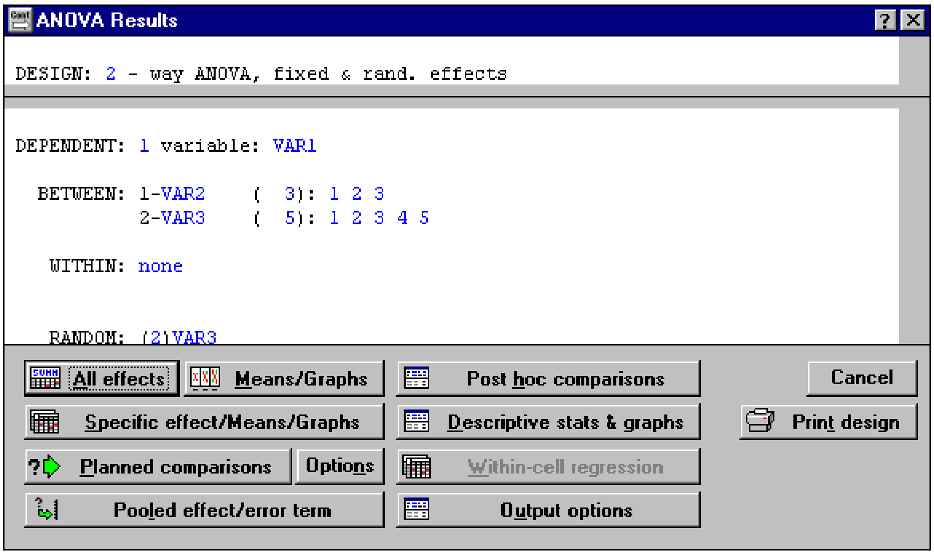
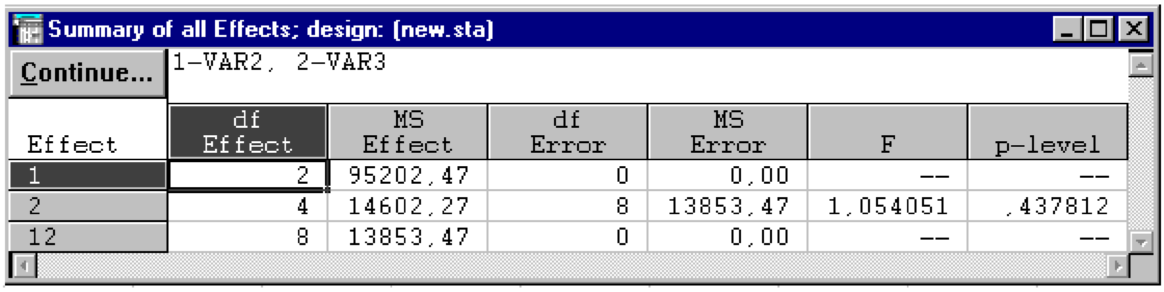


Рис. 4.8. Окно результатов однофакторного дисперсионного анализа для связанных выборок

Чтобы получить сводную таблицу вычисленных статистик нажмите кнопку 

. На экране откроется таблица со статистиками (рис. 4.9). В ней приведены следующие величины: степень свободы для признака (*df Effect*), усредненная вариативность признака (*MS Effect*), обусловленная действием исследуемого фактора, степень свободы для неучтенного фактора (*df Error*), усредненная вариативность, обусловленная неучтенными факторами (*Ms Error*), *F*-статистика и соответствующий ей уровень значимости (*p-level*).



Summary of all Effects; design: (new.sta)

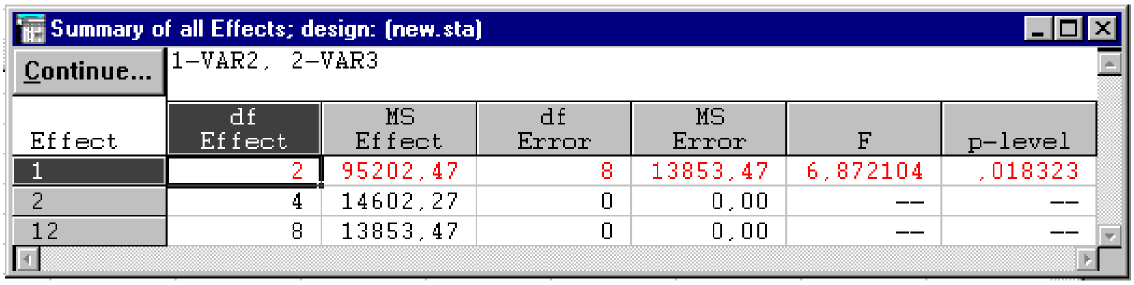
Continue... 1-VAR2, 2-VAR3

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	95202,47	0	0,00	--	--
2	4	14602,27	8	13853,47	1,054051	,437812
12	8	13853,47	0	0,00	--	--

Рис. 4.9. Окно результатов

Табл. 4.9 состоит из трех строк. Первая строка содержит информацию об усредненной вариативности признака. Вторая строка – усредненная вариативность индивидуальных различий между испытуемыми, и третья строка – усредненная вариативность обусловленная неучтенными факторами. Полученный уровень значимости равен 0,44, что предполагает отклонение гипотезы $H_{0(Б).}$ Индивидуальные различия между испытуемыми являются не более выраженными, чем различия, обусловленные случайными причинами.

Далее, нажмите кнопку **Continue...**. Откроется окно результатов дисперсионного анализа (рис. 4.8). Нажмите кнопку **Cancel**, для того чтобы задать следующий определяющий фактор. Нажмите кнопку **Random factors:** и в качестве второго определяющего фактора в смешанной модели выберите номер испытуемого. В окне с результатами однофакторного дисперсионного анализа (рис. 4.8) еще раз выведите окно с подсчитанными статистиками, нажав кнопку **All effects** (рис. 4.10).



Summary of all Effects; design: (new.sta)

Continue... 1-VAR2, 2-VAR3

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	95202,47	8	13853,47	6,872104	,018323
2	4	14602,27	0	0,00	--	--
12	8	13853,47	0	0,00	--	--

Рис. 4.10. Окно результатов

Полученный уровень значимости равен 0,02 ($p < 0{,}05$), из чего следует что $H_{0(А)}$ принимается. Различия в длительности попыток решения анаграмм разной длины являются более выраженными, чем различия, обусловленные случайными причинами.

Однако, судя по рис. 4.7, мы не можем утверждать, что срабатывает фактор длины анаграммы. Более значимыми оказываются качественные, а не количественные различия между анаграммами.

Глава 5
МНОГОМЕРНЫЕ СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ В ЗАДАЧАХ ПРИКЛАДНОГО ИССЛЕДОВАНИЯ

5.1. Основные положения факторного анализа

С начала XX в. интенсивно развивается особая область статистических исследований, называемая факторным анализом. Развитие этого направления началось в психологии. Авторами основных концепций факторного анализа являются главным образом американские и английские ученые (Ч. Спирмен, Л.Л. Тэрстоун, Г.Х. Томсон, С.Л. Барт, Р.Б. Кеттел).

Основное предположение факторного анализа можно сформулировать следующим образом: явления в определенной области исследований, несмотря на свою разнородность и изменчивость, могут быть описаны относительно небольшим числом функциональных единиц, параметров или факторов. Факторный анализ не признает произвольных решений о важности тех или иных переменных для данной области исследований. Более того, он не ограничивается утверждением, что изменение одной переменной связано или не связано с изменением другой, а идет дальше, пытаясь определить меру этой связи. При этом самое главное заключается в том, что он не ограничивается сопоставлением изменений, лежащих на поверхности явлений, а стремится обнаружить основные влияния, лежащие в основе этих изменений [21].

Исследование начинается со сбора наблюдений о варьировании некоторого набора переменных. Далее рассчитываются все возможные корреляции между наблюденными переменными для определения, существует ли между ними взаимосвязь и какова ее мера. Используя корреляционный анализ, исследователь пытается выявить взаимосвязь исследуемых признаков, что, в свою очередь, дает ему возможность выделить полный и безызбыточный набор признаков, путем объединения сильно коррелирующих признаков.

На основе полученных коэффициентов корреляции и проводится факторный анализ, который позволяет выявить новые латентные переменные, которые являются линейными комбинациями прежних и передают большую часть информации, заключенной в первоначальных наблюдениях.

Во всяком случае, нужно осознать, что с одной стороны мы имеем дело с определенной концепцией, которая стремится объяснить найденные корреляции при помощи общих факторов, а с другой стороны необходимо учитывать, что возможности достаточно точного и однозначного выделения этих факторов с помощью математики ограничены.

При решении задач факторного анализа исследователь обычно делает три шага. Их можно обозначить как:

1) подготовка соответствующей корреляционной матрицы;
2) выделение первоначальных (ортогональных) факторов;
3) вращение матрицы первоначальных факторов с целью получения окончательного решения.

Основная модель факторного анализа записывается следующей системой равенств [13]:

$$x_i = \sum_{j=1}^{m} l_{ij} f_j + \varepsilon_i; \ i = \overline{1, p}; \ m \le p. \tag{5.1}$$

То есть полагается, что значения каждого признака x_i могут быть выражены суммой простых факторов f_j, количество которых меньше числа исходных признаков, и остаточным членом ε_i с дисперсией $\sigma^2(\varepsilon_i)$, действующей только на x_i, который называют *специфическим фактором*.

Коэффициенты l_{ij} называются *нагрузкой* i-й переменной на j-й фактор или нагрузкой j-го фактора на i-ю переменную. В самой простой модели факторного анализа считается, что факторы f_i взаимно независимы и их дисперсии равны единице, а случайные величины ε_i тоже независимы друг от друга и от какого-либо фактора f_j. Максимально возможное количество факторов m при заданном числе признаков p определяется неравенством

$$(p + m) \le (p - m)^2, \tag{5.2}$$

которое должно выполняться, чтобы задача не вырождалась в тривиальную. Данное неравенство получается на основании подсчета степеней свободы, имеющихся в задаче [17].

Сумму квадратов нагрузок в выражении (5.1) называют *общностью* соответствующего признака x_i, и чем больше это значение, тем лучше описывается признак x_i факторами f_j. Общность есть часть дисперсии признака, которую объясняют факторы. В свою очередь, ε_i^2 показывает, какая часть дисперсии исходного признака остается необъясненной (специфичный фактор, соответствующий только одной определенной переменной, и дисперсии, обусловленной ошибкой) при используемом наборе факторов, и данную величину называют *характерностью признака*. Таким образом,

$$\textbf{дисперсия признака = общность } (\sum_{j=1}^{m} l_{ij}^2) + \textbf{ характерность } (\varepsilon_i^2). \tag{5.3}$$

Основное соотношение факторного анализа показывает, что коэффициент корреляции любых двух признаков x_i и x_j можно выразить суммой произведения нагрузок некоррелированных факторов:

$$r_{ij} = r(x_i, x_j) = l_{i1}l_{j1} + l_{i2}l_{j2} + \ldots + l_{im}l_{jm}. \quad (5.4)$$

Процесс выделения факторов начинается с составления матрицы коэффициентов корреляции. Цель состоит в переходе от редуцированной матрицы корреляций (элементами главной диагонали являются общности) к редуцированной факторной матрице, которая позволит определить: 1) сколько общих факторов необходимо для отражения всех корреляций между переменными и 2) каковы нагрузки каждого фактора для разных переменных.

Задачу факторного анализа нельзя решить однозначно. Равенства (5.1) не поддаются непосредственной проверке, так как p исходных признаков задается через $(p + m)$ других переменных – простых и специфических факторов. Поэтому представление корреляционной матрицы факторами, или, как говорят, факторизацию, можно произвести бесконечно большим числом способов. Если удалось произвести факторизацию корреляционной матрицы с помощью некоторой матрицы факторных нагрузок F, то любое линейное ортогональное преобразование F (ортогональное вращение) приведет к такой же факторизации [12].

Существующие программы вычисления нагрузок начинают работать с $m = 1$ (однофакторная модель) [12]. Затем проверяется, насколько корреляционная матрица, восстановленная по однофакторной модели в соответствии с основным соотношением факторного анализа (5.4), отличается от корреляционной матрицы исходных данных. Если однофакторная модель признается неудовлетворительной, то испытывается модель с $m = 2$ и так далее до тех пор, пока при некотором m не будет достигнута адекватность или число факторов в модели не превысит максимально допустимое. В последнем случае говорят, что адекватной модели факторного анализа не существует [12].

Если адекватная факторная модель существует, то производится вращение полученной системы общих факторов, так как значения факторных нагрузок и нагрузок на факторы есть лишь одно из возможных решений модели (5.1). Вращение факторов может производиться разными способами. Наиболее часто это вращение осуществляется таким образом, чтобы как можно большее число факторных нагрузок стало нулями, и каждый фактор по возможности описывал группу сильно коррелированных признаков. Также можно вращать факторы до тех пор, пока не получатся результаты, поддающиеся содержательной интерпретации. Можно, например, потребовать, чтобы один фактор был нагружен преимущественно признаками одного типа, а другой – признаками другого типа. Или, скажем, можно потребовать, чтобы исчезли какие-то трудно интерпретируемые нагрузки с отрицательными знаками. Нередко исследователи идут дальше и рассматривают прямоугольную систему факторов как частный случай косоугольной, т. е. ради содержания жертвуют условием некоррелированности факторов.

В завершение всей процедуры факторного анализа с помощью математических преобразований выражают факторы f_j через исходные признаки, т. е. получают в явном виде параметры линейной диагностической модели.

Если при факторном анализе признаков ищутся группы близких (коррелированных) признаков на основе корреляционной матрицы, то для транспонированных данных (таблица экспериментальных данных поворачивается на 90°) аналогом корреляционной матрицы является матрица, описывающая попарные коэффициенты корреляции (сходства) объектов. Она вводится в алгоритм формального факторного анализа, и в результате получаются факторы, описывающие уже не группы коррелированных признаков, а группы сходных объектов [2]. Особенности данной процедуры подробно рассмотрены в [1].

Известно большое количество методов факторного анализа (ротаций, максимального правдоподобия и др.). Нередко в одном и том же пакете программ анализа данных реализовано сразу несколько версий таких методов, и у исследователей возникает правомерный вопрос о том, какой из них лучше. В.В. Александров в [2] доказывает, что практически все методы дают весьма близкие результаты. Подобные выводы сделаны и одним из основоположников современного факторного анализа Г. Харманом: «Ни в одной из работ не было показано, что какой-либо один метод приближается к «истинным» значениям общностей лучше, чем другие методы… Выбор среди группы методов «наилучшего» производится в основном с точки зрения вычислительных удобств, а также склонностей и привязанностей исследователя, которому тот или иной метод казался более адекватным его представлениям об общности» [30].

У факторного анализа есть много сторонников и много оппонентов. Но, как справедливо заметил В.В. Налимов: «…У психологов и социологов не оставалось других путей, и они изучили факторный анализ со всей обстоятельностью» [24]. Для более подробного ознакомления с факторным анализом и его методами может быть рекомендована литература [1, 12, 13, 17, 20, 21, 29, 30].

5.2. Компьютерная обработка данных при проведении факторного анализа

В последние десятилетия интерес к факторному анализу вспыхнул с новой силой. Это обусловлено бурным развитием вычислительной техники, без применения которой использование метода факторного анализа является слишком трудоемким процессом (подробно процедура «ручного» расчета факторов изложена в [13, 21]). На сегодняшний день метод факторного анализа входит в состав практически всех статистических пакетов (например, *SPSS*, *STATGRAPHICS*, *STATISTICA*).

В табл. 5.1 указаны способы активизации факторного анализа.

Таблица 5.1

Способы активизации факторного анализа в статистических пакетах

Пакет	Последовательность действий
SPSS	*Analyze – Data Reduction – Factor…*
STATGRAPHICS	*Special – Multivariate Methods – Factor Analysis…*
STATISTICA	*Модуль Factor Analysis*

В каждом модуле факторного анализа возможны следующие типы исходных данных:

- файл, где по строкам записаны значения переменных;
- файл с корреляционной матрицей.

Также пользователю предоставляется возможность задать способ обработки пропущенных значений. Здесь возможны разные варианты. Первый вариант состоит в том, что в исходной матрице данных игнорируются все строки, в которых имеется хотя бы одно пропущенное значение. Второй вариант – парный способ исключения пропущенных значений, т. е. игнорируются пропущенные случаи не для всех переменных, а лишь для выбранной пары (например, при вычислении корреляционной матрицы). В данном варианте остается больше наблюдений для обработки. И третий вариант – подстановка среднего вместо пропущенных значений. В табл. 5.2 указаны способы исключения пропущенных данных.

Для того чтобы задать способы обработки пропущенных значений в пакете *SPSS* нужно вызвать диалоговое окно *Factor Analysis* (см. табл. 5.1) и нажать кнопку 

. В пакете *STATGRAPHICS*, после того как откроется окно с суммарными статистиками факторного анализа, в контекстном меню (правая кнопка мыши) выберите пункт *Analysis Options*. В пакете *STATISTICA* в стартовой панели факторного анализа в раскрывающемся списке *MD deletion*.

Таблица 5.2

Способ исключения пропущенных данных

	SPSS	*STATGRAPHICS*	*STATISTICA*
Пропущенные данные	*Missing Values*	*Missing Value Treatment*	*Missing Data*
Удаление строки	*Listwise*	*Listwise*	*Casewise*
Удаление пары	*Pairwise*	*Pairwise*	*Pairwise MD Deletion*
Подстановка среднего	*Replace with mean*	–	*Mean Substitution*

Если в качестве критерия оптимальности выбирается минимум расхождения между ковариационной матрицей исходных данных и матрицей, которая получается после оценивания нагрузок, то выбирается метод главных компонент. А если в качестве критерия оптимальности выбирается максимальная близость исходных корреляций признаков к тем, которые получены в модели после оценивания факторных нагрузок, то реализовывается метод главных факторов. В табл. 5.3 приведены некоторые методы выделения факторов.

Таблица 5.3

Методы выделения факторов

	SPSS	*STATGRAPHICS*	*STATISTICA*
Метод главных компонент	*Principal components*	*Principal components*	*Principal components*
Максимальное правдоподобие	*Maximum Likelihood*	–	–
Центроидный метод	–	–	*Centroid method*
Метод главных осей	*Principal axis factoring*	–	*Principal axis method*
Классический	–	*Classical*	*Communalities = = multiple R**2*

Также можно определить число факторов, которые будут выделены программой. Задается либо максимальное число факторов (*Maximum no. Of factors*) либо задается ограничение на число выделяемых факторов с помощью минимального собственного значения (*Eigenvalue*). Обычно не рассматриваются значения меньше единицы.

После того как найдено пространство общих факторов, выделенные факторы необходимо интерпретировать. Для содержательного объяснения можно воспользоваться вращением полученной матрицы.

Вращение методом варимакс ставит целью упростить столбцы факторной матрицы, сводя все значения к 1 или 0. Вращение методом квартимакс ставит целью аналогичное упрощение только по отношению к строкам факторной матрицы. Эквимакс занимает промежуточное положение – при вращении факторов по этому методу одновременно делается попытка упростить и столбцы и строки. Рассмотренные методы вращения относятся к ортогональным вращениям, т. е. в результате получаются некоррелированные факторы. Методы прямого облимина и промакс вращения относятся к косоугольным вращениям, в результате которых получаются коррелированные между собой факторы. В табл. 5.4 представлены методы вращения.

Таблица 5.4

Методы вращения факторов

SPSS	*STATGRAPHICS*	*STATISTICA*
Varimax	*Varimax*	*Varimax raw / Varimax normalized*
Quartimax	*Quartimax*	*Quartimax raw / Quartimax normalized*
Equamax	*Equamax*	*Equamax raw / Equamax normalized*
Promax	–	–
Direct Oblimin	–	–

Выходная информация представляет собой набор таблиц и графиков. В таблицах приводятся значения факторных нагрузок до и после вращения, а также значения полученных факторов для каждого испытуемого. Графики включают в себя иллюстрацию процента дисперсии для выделенных факторов, проекцию объектов на плоскость любых двух факторов (по выбору пользователя), графическое (двумерное) изображение факторных нагрузок для всех исходных переменных, отображение факторных нагрузок в пространстве любых трех факторов.

5.3. Примеры использования факторного анализа при решении задач прикладных исследований

Пример 5.1[5]

Исходным материалом для факторного анализа послужили результаты комплексного исследования авторов по оценке технологических и концептуальных качеств руководителей (15 качеств) и сотрудников (16 качеств), проводившегося в одном из акционерных обществ города Томска.

На первом этапе работы, была проанализирована матрица корреляций профессионально значимых качеств руководителей и сотрудников, что позволило психологам сделать вывод: к чему в условиях данного предприятия приводит выбор той или иной стратегии руководства.

Следующий этап – построение факторной модели на основе полученной корреляционной матрицы. В результате, как для руководителей, так и для сотрудников было выделено по три новых обобщенных показателя (или в терминах факторного анализа – три «фактора»), с помощью которых можно охарактеризовать их деятельность.

Для оценки деятельности руководителей:

Фактор 1 – фактор добросовестности.

Фактор 2 – фактор социальной направленности.

Фактор 3 – необходимые качества лидера.

Для оценки деятельности сотрудников:

Фактор 1 – фактор управленческой деятельности, т. е. перспективности сотрудников с точки зрения возможности его использовать на руководящей должности (возможность включения в резерв кадров).

Фактор 2 – фактор добросовестности и профессиональной компетентности.

Фактор 3 – фактор работоспособности и исполнительности.

На следующем этапе анализа полученных факторов психологами были выделены группы, характеризующие кадровый состав предприятия в соответствии с высокими, средними и низкими значениями выделенных факторов (табл. 5.5, табл. 5.6).

Из табл. 5.5 видно, что 12,9 % сотрудников можно отнести в «Группу резерва кадров», то есть это те сотрудники, которые обладают всеми качествами, необходимыми для работы руководителем.

Основная масса (48,4 %) сотрудников попали в группу «Золотая середина». Они имеют средние значения по всем факторам. Сотрудники этой группы «находятся на своем месте», однако они являются малоперспективными с точки зрения дальнейшего профессионального роста. Достаточно большая часть сотрудников (15,8 %) попали в группу с низкими значениями по всем факторам. Это так называемая «Группа риска». Для сотрудников, попавших в эту группу, вопрос об их профпригодности в каждом конкретном случае необходимо решать в индивидуальном порядке.

Анализируя табл. 5.6, можно отметить следующее: что наиболее перспективными в плане дальнейшего профессионального роста, то есть переход на более высокую руководящую должность, представляют сотрудники, попавшие в группы 1 и 4 (39,1 % всех руководителей).

Руководители, имеющие низкие значения по всем факторам «Группа риска» составляют 4,9 %.

Таблица 5.5

Оценки, характеризующие сотрудников предприятия

№	Характеристика группы	Уровень фактора			Процент сотрудников, попавших в данную группу
		Фактор 1 «Способность к управленческой деятельности»	Фактор 2 «Профессиональная компетентность»	Фактор 3 «Работоспособность и исполнительность»	
1	Резерв кадров	высокий	высокий	высокий	12,9
2	«Золотая середина»	средний	средний	средний	48,4
3	«Группа риска»	низкий	низкий	низкий	15,8
4	Профессионалы высокого уровня, не способные к управленческой деятельности	средний	высокий	высокий	4,8
5	«Группа риска»	низкий	средний	низкий	4,8
6	Профессионалы, не способные к управленческой деятельности	низкий	средний	средний	13,3

Таблица 5.6

Оценки, характеризующие руководителей предприятия

№	Характеристика группы	Уровень фактора			Процент сотрудников попавших в данную группу
		Фактор 1 «Профессиональная компетентность»	Фактор 2 «Социальная направленность»	Фактор 3 «Качества лидера»	
1	Резерв управленческих кадров высшего звена	высокий	высокий	высокий	29,3
2	«Золотая середина»	высокий или средний	средний	средний	56
3	«Группа риска»	низкий	низкий	низкий	4,9
4	Резерв	высокий	средний	высокий	9,8

Таким образом, руководителей, имеющих самые низкие значения по всем факторам, оказалось гораздо меньше, чем сотрудников, попавших в аналогичную группу. Возможно, одной из причин является то, что приведенные в табл. 5.5 обобщенные факторы получены на основании анкетирования сотрудников, а сотрудники некоторых подразделений предприятия боялись высказывать свое мнение, несмотря на то, что анкетирование проводилось анонимно.

Полученные данные были использованы при формировании рекомендаций по повышению качества работы с кадрами на данном предприятии.

Пример 5.2 [14]

Оценка учебных дисциплин как критерий качества образовательного процесса.

Под качеством образовательной деятельности авторы понимают способность предлагаемых в учебном процессе услуг удовлетворять запросам обучаемых.

Один из аспектов определения качества образовательной деятельности – оценка учебных дисциплин – исследовался нами через выявление отношения студентов к преподаваемым учебным предметам: выявление наиболее и наименее «привлекательных» учебных дисциплин, их основных характеристик.

Для измерения была выделена целевая группа студентов IV курса факультета автоматики и вычислительной техники Томского политехнического университета. Всего в исследовании приняли участие 105 человек. Уровень надежности и точности, который обеспечивает данная выборка, является приемлемыми, а объем выборки (105 человек) – статистически значимым. Структура выборки имеет вид, представленный на рис. 5.1.

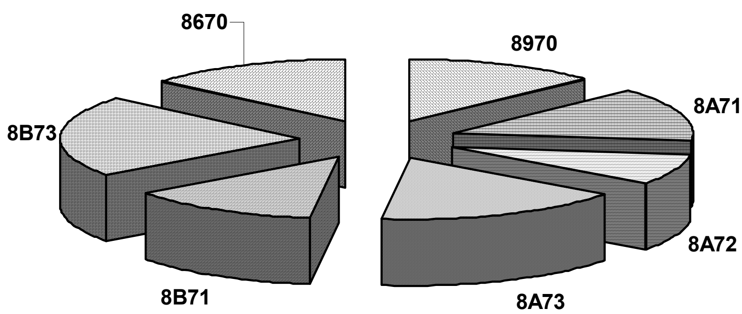


Рис. 5.1. Структура выборки – доля каждой группы в общей выборке

В нашем случае применение факторного анализа позволило перейти от системы 23 признаков к системе 4 новых переменных (факторов) и провести анализ имеющихся объектов (учебных дисциплин) в пространстве этих новых глобальных переменных. Для проведения факторного анализа был использован пакет *STATISTICA*. Анализ осуществлялся методом главных компонент, рассматривались собственные значения больше 1. Вычисленные статистики представлены в табл. 5.7.

Как видно из табл. 5.7, первая компонента объясняет 40 % общей дисперсии, вторая компонента – 16 % общей дисперсии, третья компонента – 10 % общей дисперсии и т. д. Первые 5 главных компонент описывают почти 80 % дисперсии.

Таблица 5.7

Собственные значения. Метод главных компонент

№	Собственные значения	% общей дисперсии	Накопленные собственные значения	Накопительный %
1 компонента	9,24	40,18	9,24	40,18
2 компонента	3,58	15,56	12,82	55,73
3 компонента	2,19	9,51	15,01	65,24
4 компонента	1,77	7,68	16,77	72,93
5 компонента	1,41	6,11	18,18	79,04

Матрицу факторных нагрузок подвергнем вращению методом нормализованного варимакса. Термин нормализованый означает, что факторные нагрузки в процедуре нормализуются, т. е. делятся на корень квадратный из соответствующей общности. Результаты, полученные после вращения представлены в табл. 5.8.

ФАКТОР 1. Данный фактор можно назвать фактором *заинтересованности*: предметы, имеющие высокие значения по данному фактору часто обсуждаются с друзьями, есть желание изменить содержание предмета (внести что-то более интересное, новое). По данным предметам, курсам студентам не хватает аудиторных часов (слишком мало), методических и наглядных пособий, раздаточного материала, сформировались высокий уровень знаний и желание преподавать в дальнейшем эти предметы.

*ФАКТОР 2***.** Это фактор *непонимания, отторжения*: изучение предметов, имеющих высокие нагрузки по данному фактору, дается трудно, наблюдается низкий уровень знаний, желание пропустить занятие.

ФАКТОР 3. Это фактор *удовлетворенности*: предметы, имеющие высокие нагрузки по третьему фактору, не вызывают особой заинтересованности, но нет и отторжения, т. е. предметы воспринимаются нейтрально.

Таблица 5.8

Факторные нагрузки после варимакс нормализованного вращения

	Factor 1	Factor 2	Factor 3	Factor 4
Активность, интерес, побуждение	,157743	,108062	,780835	–,055247
Безусловное восприятие, комфорт	,039844	–,090394	–,508640	–,101450
Невыраженное отношение, неясность, неопределенность, «все равно»	–,396147	,185270	–,625455	,060004
Интерес, поиск смысла для себя, полезности	,421873	–,298931	,040948	,587956
Стабильность, консерватизм, жесткость позиции, ригидность, жесткость структуры мышления	–,509986	–,233327	–,010377	,407919
Невыраженное восприятие	,044755	–,332320	–,065078	–,721312
Трудность изучения	,284554	,694803	,231789	,347498
Низкий уровень знаний	,084200	,867800	,029641	–,056870
Отрицание, отторжение, негативное восприятие	–,227332	,829910	–,091798	,047847
Отсутствие интереса	,012792	,655117	–,624100	,023795
Высокий уровень знаний	,638029	–,129279	,462902	,382701
Желание иметь раздаточный материал	,636630	,298944	,282373	,558778
Недостаточность наглядных пособий	,692101	–,011963	,517719	,246792
Заинтересованность	,975051	–,054275	,099172	,002941
Необходимость изменения	,801079	,278082	–,327694	–,070077
Недостаточность литературы и пособий	,901512	–,208135	,091847	,019947
Недостаточно теории	,762608	,209104	–,106281	,021081
Недостаточно практики	,838565	–,189553	,369814	,171144
Недостаточно теории и практики	,779646	,256641	–,047166	,267291
Дисциплина включает достаточное количество часов	–,903943	,073614	–,259105	–,166794
Желание преподавать	,851740	–,118621	,356938	,015747
Основательность, надежность, прочность	,656777	–,377446	,195621	–,291545
Поиск необычного в новой информации, сложное восприятие	–,056626	,493596	,330403	–,432400

ФАКТОР 4. Это фактор *неудовлетворенности:* наблюдается противоречивое отношение студентов к учебным дисциплинам, имеющим высокие значения по данному фактору, т. е., с одной стороны довольно

высокий уровень знаний по предмету, с другой – трудное понимание предмета, его некачественное преподавание.

В табл. 5.9 приведены коэффициенты корреляции факторов. Как видно из приведенных данных, фактор 1 коррелирует с 3 и 4 факторами, в то время как фактор 2 более независим, что и следовало ожидать, так как взаимосвязь между заинтересованностью предметом и степенью удовлетворенности и неудовлетворенности восприятия предмета очевидна.

Таблица 5.9

Факторная корреляция

	Фактор 1	Фактор 2	Фактор 3	Фактор 4
Фактор 1	1,00	0,03	0,46	0,39
Фактор 2	0,03	1,00	–0,02	0,01
Фактор 3	0,46	–0,02	1,00	0,14
Фактор 4	0,39	0,01	0,14	1,00

В табл. 5.10 представлены коэффициенты нагрузки каждого фактора на дисциплину.

Таблица 5.10

Коэффициенты факторных нагрузок для оцениваемых учебных дисциплин

	Фактор 1	Фактор 2	Фактор 3	Фактор 4
Базы данных и экспертные системы	1,255473	–1,63150	–,47501	,92178
Геометрия и алгебра	–,754561	–,57318	,85742	1,27901
Дифференциальные уравнения	–,994967	,06018	,66005	1,21251
Информатика	1,788954	–1,31443	1,09423	–,67930
Информационные технологии	,384741	–1,69871	–,16473	,63862
Компьютерная графика	1,253351	,09277	1,99255	–,42011
Математический анализ	–,655247	1,23081	2,34136	1,37459
Основы теории случайных функций	–,963377	–,52207	,37268	–,49850
Практикум на ЭВМ	,228331	–1,19233	,73854	–,53818
Теория вероятностей и мат. статистика	–,540180	,75582	,40237	,97417
Физика	–,612473	–,24648	,62824	,67834
Численные методы	–,861736	–,44243	,29121	,48062
Языки программирования и методы трансляций	2,152513	–,49499	–,14666	–1,14103
Безопасность жизнедеятельности	–,398978	,05519	–1,66246	,19753
Иностранный язык	2,632996	2,54238	–1,08126	1,64945

Окончание табл. 5.10

	Фактор 1	Фактор 2	Фактор 3	Фактор 4
История	–,328836	1,40688	,07412	–1,02866
Культурология	–,438174	,29559	–,09375	–1,32866
Политология	–,496346	1,19825	–,29974	–,36024
Правоведение	–,350821	–,39713	–1,39729	,63191
Психология и педагогика	–,041586	–,96330	–1,05443	,44286
Социология	–,371948	–,05379	–,58284	–,32127
Философия	,189427	1,12438	,11364	–1,65914
Экология	–,318101	,43234	–1,66478	,51600
Экономика	–,067606	,39362	,27927	–,35838
Элективные курсы по истории и культурологии	–,945855	–,52465	–1,12551	–,45978
Элективные курсы по философии	–,744996	,36677	–,09724	–2,20415

В таблице отмечены те коэффициенты, величина которых больше или близка к единице. В табл. 5.11 дана интерпретация табл. 5.10.

Таблица 5.11

Характеристика учебных дисциплин,
полученная по результатам факторного анализа

Базы данных и экспертные системы	Заинтересованность, удовлетворенность
Геометрия и алгебра	Больше неудовлетворенности, хотя, в общем, нейтрально
Дифференциальные уравнения	Заинтересованности нет, неудовлетворенность
Информатика	Заинтересованность, удовлетворенность
Информационные технологии	Заинтересованность, неудовлетворенность
Компьютерная графика	Заинтересованность, удовлетворенность
Математический анализ	Заинтересованности нет, отмечается даже отторжение, хотя, скорее всего, предмет воспринимается «как должный»
Основы теории случайных функций	Нейтральное отношение
Практикум на ЭВМ	Вполне удовлетворяет
Теория вероятностей и мат. статистика	Отторжение с неудовлетворенностью
Физика	Предмет воспринимается «как должный»
Численные методы	Предмет воспринимается «как должный»
Языки программирования и методы трансляций	Высокая степень заинтересованности

Окончание табл. 5.11

Безопасность жизнедеятельности	Заинтересованности нет, предмет воспринимается «как должный»
Иностранный язык	Высокая степень заинтересованности, но и высокая степень отторжения с неудовлетворенностью
История	Отторжение
Культурология	Предмет воспринимается «как должный» с низкой степенью отторжения
Политология	Предмет воспринимается «как должный» с низкой степенью отторжения
Правоведение	Предмет воспринимается «как должный» с низкой степенью неудовлетворения
Психология и педагогика	Предмет воспринимается «как должный» с низкой степенью отторжения
Социология	Нейтральное отношение
Философия	Отторжение
Экология	Отторжение с малой долей неудовлетворенности
Экономика	Предмет входит в разряд «трудных», низкая доля удовлетворенности
Элективные курсы по истории и культурологии	Нейтральное отношение

5.4. Кластерный анализ

Кластерный анализ предназначен для разбиения множества объектов на однородные группы (классы) на основании некоторого математического критерия качества классификации (*cluster* (англ.) – гроздь, пучок, скопление, группа элементов, характеризуемых каким-либо общим свойством).

Критерий качества классификации в той или иной мере отражает следующие неформальные требования [18]:

1) внутри групп объекты должны быть тесно связаны между собой;

2) объекты разных групп должны быть далеки друг от друга;

3) при прочих равных условиях распределения объектов по группам должны быть равномерными.

Требования 1 и 2 выражают стандартную концепцию компактности классов разбиения [3]; требование 3 состоит в том, чтобы критерий не навязывал объединения отдельных групп объектов.

Основная цель кластерного анализа – выделить в исходных многомерных данных такие однородные подмножества, чтобы объекты внутри групп были похожи друг на друга, а объекты из разных групп – не похожи. Под «похожестью» понимается близость объектов в многомерном пространстве признаков, и тогда задача сводится к выделению в этом пространстве естественных скоплений объектов, которые и считаются однородными группами [27].

Решением задачи кластерного анализа является разбиение, удовлетворяющее некоторому критерию оптимальности. Этот критерий может представлять собой некоторый функционал, выражающий уровни желательности различных разбиений и группировок. Этот функционал часто называют целевой функцией. Но что означает «два объекта x_i и x_j различны»? Задача была бы решена, если бы объекты i и j попадали в один кластер всякий раз, когда расстояние (отдаленность) между соответствующими точками x_i и x_j было бы достаточно малым и, наоборот, попадали в разные кластеры, если бы расстояние между точками x_i и x_j было достаточно большим.

В табл. 5.12 приведены примеры наиболее часто используемых функций расстояния.

Таблица 5.12

Некоторые функции расстояния, используемые в методах кластерного анализа

Название	Формула	Обозначения
1. Евклидово расстояние	$d_2(x_i,x_j)=\sqrt{\sum_{k=1}^{p}\left(x_{ki}-x_{kj}\right)^2}$	i, j – номера объектов; p – количество переменных, характеризующих объект; k – номер признака; x_{ki} – значение k-го признака у i-го объекта; S – ковариационная матрица
2. $L1$-норма	$d_1(x_i,x_j)=\left[\sum_{k=1}^{p}\left\|x_{ki}-x_{kj}\right\|\right]$	
3. Сюпремум норма	$d_\infty(x_i,x_j)=\sup\limits_{k=1,2,\ldots,p}\left\{\left\|x_{ki}-x_{kj}\right\|\right\}$	
4. Lp-норма	$d_p(x_i,x_j)=\left[\sum_{k=1}^{p}\left\|x_{ki}-x_{kj}\right\|\right]^{1/p}$	
5. Расстояние Махаланобиса	$d\left(x_i,x_j\right)=\sqrt{(x_i-x_j)S^{-1}(x_i-x_j)}$	
5. Расстояние Хэмминга	$d(x_i,x_j)=\sum_{k=1}^{p}\left\|x_{ik}-x_{jk}\right\|$	

Многие процедуры при кластеризации совершаются ступенчато. Это означает, что два наиболее близко расположенных объекта x_i и x_j объединяются и рассматриваются как один кластер. Это приводит к тому, что число объектов уменьшается и становится равным $n - 1$, причем один кластер будет содержать два объекта, а остальные по одному. Наиболее подходящее разбиение выбирает чаще всего сам исследователь, которому предоставляется дендрограмма, отображающая результаты группирования объектов на всех шагах алгоритма (рис. 5.2). Могут одновременно также использоваться и математические критерии качества группирования.

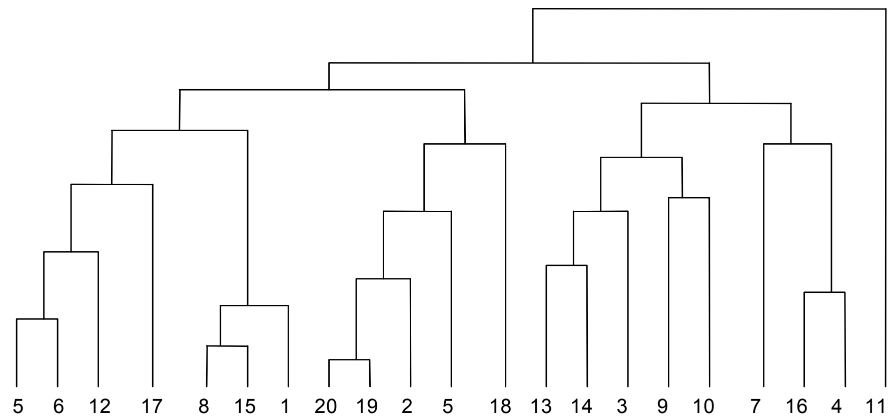


Рис. 5.2. Результаты работы иерархической процедуры группирования объектов, представленные в виде дендрограммы (общее количество объектов – 20)

Как говорилось выше, основные усилия в развитии методов кластерного анализа направлены на построение методов, основанных на минимизации внутригрупповых сумм квадратов (отклонений).

Традиционно различают классификации иерархические и неиерархические (называемые иногда структурными). Соответственно, можно разделить алгоритмы получения этих классификаций.

Принцип работы иерархических алгоритмов состоит в последовательном объединении в кластер сначала самых близких, а затем и все более отдаленных друг от друга элементов. Большинство из этих алгоритмов исходит из матрицы сходства (расстояний), и каждый отдельный элемент рассматривается вначале как отдельный кластер. Общая схема такой иерархической группировки может быть представлена как повторяющееся приложение трех операций к мерам расстояния объект (кластер) – объект (кластер):

1) найти наименьшее расстояние d_{S_1,S_2} между объектом (кластером) S_1 и объектом (кластером) S_2;
2) объединить S_1 и S_2 в один кластер, присвоив общий индекс объединению $S_1 \bigcup S_2$;
3) вычислить расстояние $d_{S,S_1 \cup S_2}$ от кластера $S_1 \bigcup S_2$ до любого другого объекта (кластера) S.

Рассмотрим некоторые методы, основанные на евклидовой мере расстояния.

МЕТОД УОРДА. В качестве целевой функции применяется внутригрупповая сумма квадратов (ВСК) отклонений, которая есть не что иное, как сумма квадратов расстояний между каждой точкой (объектом) и средней по кластеру, содержащему этот объект. Этот метод представляет собой последовательную процедуру: на каждом шаге объединяются такие два кластера, которые приводят к минимальному увеличению целевой функции, т. е. ВСК. Метод Уорда направлен на объединение близко расположенных кластеров.

ЦЕНТРОИДНЫЙ МЕТОД. Расстояние между двумя кластерами I и J в этом методе определяется как евклидово расстояние между центрами (средними) этих кластеров. Кластеризация осуществляется поэтапно: на каждом из $n-1$ шагов объединяют два кластера I и J, имеющие минимальные значения d_{ij}^2. Если n_1 много больше n_2, то центры $I \bigcup J$ и J близки друг к другу и характеристики J при объединении кластеров практически игнорируются [22].

МЕТОД БЛИЖАЙШЕГО СОСЕДА. На первом шаге каждый объект считается отдельным кластером. Расстояние между кластерами равно расстоянию между двумя ближайшими соседями классов. Независимо от общей формы кластера к нему присоединяются ближайшие к границе объекты.

МЕТОД ДАЛЬНЕГО СОСЕДА. Так же как и во всех методах, в этом методе каждый объект на первом шаге считается отдельным кластером, и межкластерное расстояние равно расстоянию между самыми дальними объектами.

МЕДИАННЫЙ МЕТОД. Межкластерное расстояние равно расстоянию между точками с медианными значениями признаков в классах.

Помимо иерархических кластер-процедур, *существуют неиерархические кластер-процедуры*, *которые иногда называют структурными.* Здесь реализуется идея образования кластеров по принципу выделения сгущений – мест наибольшей концентрации точек в рассматриваемом пространстве.

Примером структурной кластер-процедуры является МЕТОД k-средних (вариант Болла и Холла).

Нацеленность алгоритмов кластерного анализа на определенную структуру группировок объектов в пространстве признаков может приводить к неоптимальным или даже неправильным результатам, если гипотеза о типе группировок неверна. В случае отличия реальных распределений от гипотетических указанные алгоритмы часто «навязывают» данным не присущую им структуру и дезориентируют исследователя. Поэтому экспериментатор, учитывающий данный факт, в условиях априорной неопределенности прибегает к применению нескольких алгоритмов кластерного анализа и отдает предпочтение какому-либо выводу на основании общей оценки результатов работы этих алгоритмов.

Алгоритмы кластерного анализа отличаются большим разнообразием. Это могут быть, например, алгоритмы, реализующие полный перебор сочетаний объектов или осуществляющие случайные разбиения множества объектов. В то же время большинство таких алгоритмов состоит из двух этапов. На первом этапе задается начальное (возможно, искусственное или даже произвольное) разбиение множества объектов на классы и определяется некоторый математический критерий качества автоматической классификации. Затем, на втором этапе, объекты переносятся из класса в класс до тех пор, пока значение критерия не престанет улучшаться.

Многообразие алгоритмов кластерного анализа обусловлено также множеством различных критериев, выражающих те или иные аспекты качества автоматического группирования. Простейший критерий качества непосредственно базируется на величине расстояния между кластерами. Однако такой критерий не учитывает «населенность» кластеров – относительную плотность распределения объектов внутри выделяемых группировок. Поэтому другие критерии основываются на вычислении средних расстояний между объектами внутри кластеров. Но наиболее часто применяются критерии в виде отношений показателей «населенности» кластеров к расстоянию между ними. Это, например, может быть отношение суммы межклассовых расстояний к сумме внутриклассовых (между объектами) расстояний или отношение общей дисперсии данных к сумме внутриклассовых дисперсий и дисперсии центров кластеров.

Функционалы качества и конкретные алгоритмы автоматической классификации достаточно полно и подробно рассмотрены в [1, 3, 22]. Эти функционалы и алгоритмы характеризуются различной трудоемкостью и подчас требуют ресурсов высокопроизводительных компьютеров.

Так какой же алгоритм выбрать? И.Д. Мандель и Л.Б. Черный провели экспериментальное сравнение следующих алгоритмов кластеризации:

А1. Метод *k*-средних (вариант Болла и Холла).

А2. Метод *k*-средних (вариант МакКина).

А3. Метод А1, объединенный с алгоритмом Боннера.

А4. Метод *k*-средних (вариант Хартигана).

А5. Метод ближнего соседа.

А5. Метод дальнего соседа.

А7. Медианный метод.

А8. Метод простого среднего.

A9. Метод группового среднего.

A10. Центроидный метод.

A11. Метод Уорда.

A12. Метод A1 при случайном выборе центров класса.

A13. Метод A2 при случайном выборе центров класса.

A14. Метод A1 с предварительным выбором разбиения по «центроиду».

И в результате они приводят следующие выводы и рекомендации [22]:

1. Различные алгоритмы кластер-анализа часто дают существенно несходные разбиения. Близкие результаты наблюдаются только у A9 и A10 и особенно у A1 и A2.
2. Результаты классификации в целом будут лучше, если она проводится в пространстве невысокой размерности и ориентирована на выделение не слишком малого числа классов.
3. Наиболее устойчивы к зашумлению данных алгоритмы A1, A3, A11; наименее – A7 и особенно A5.
4. Для алгоритмов типа *k*-средних очень важен выбор начального разбиения. Случайный выбор (A12, A13) и выбор по центроиду (A14) не могут быть рекомендованы. Эвристики в A1–A4 в целом приемлемы, хотя определение максимально удаленных эталонов – довольно трудоемкая процедура.
5. При малом числе признаков и большом числе классов хорошо работают (восстанавливают истинную структуру) A9, A11, A6; плохо – A5, A2, A4. При большом числе признаков и малом числе классов хорошо работают A11, A2, A3; плохо – A5, A7, A10. Во всех случаях хорошо работают A6 и особенно A11, плохо – A7 и особенно A5.
6. С учетом чувствительности к зашумлению и способности к восстановлению структуры данных наилучшим является алгоритм Уорда A11, наихудшим – метод ближайшего соседа A5 [8].

Некоторые из этих алгоритмов достаточно полно представлены в прикладном пакете *STATGRAPHICS*, в котором реализовано шесть видов иерархических процедур: метод ближайшего соседа, метод дальнего соседа, центроидный метод, медианный метод, метод группового среднего, метод Уорда и одна неиерархическая процедура типа *k*-средних. Такое явное предпочтение иерархическим алгоритмам отдано из-за того, что в отличие от оптимизационных кластерных алгоритмов, представляющих исследователю конечный результат группирования объектов, иерархические процедуры позволяют проследить процесс выделения группировок и иллюстрируют соподчиненность кластеров, образующихся на разных этапах работы.

Преимущества прикладных статистических пакетов *STATGRAPHICS*, *SPSS* и *STATISTICA* для проведения кластерного анализа обусловлены их графическими возможностями. Так, в них предусмотрены двумерные и трехмерные графические отображения полученных кластеров в пространстве исследуемых переменных, а также результаты работы иерархической процедуры группирования объектов (см. рис. 5.2).

5.5. Компьютерная обработка данных при проведении кластерного анализа

Рассмотрим последовательность действий проведения кластерного анализа в статистических пакетах *STATGRAPHICS*, *SPSS* и *STATISTICA*.

Пакет STATGRAPHICS

Активизировав пакет, выберите в главном меню модуль кластерного анализа *Special – Multivariate Methods – Cluster Analysis…* (рис. 5.3).

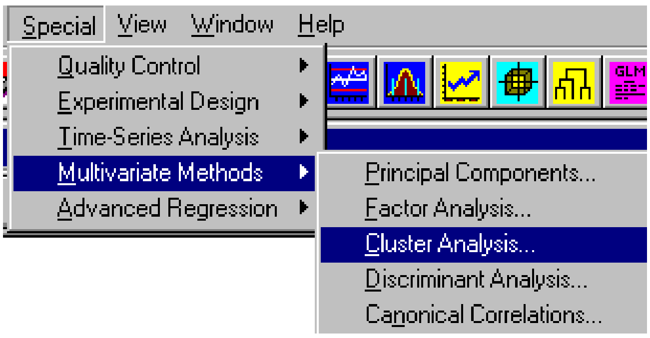


Рис. 5.3. Активизация модуля кластерного анализа

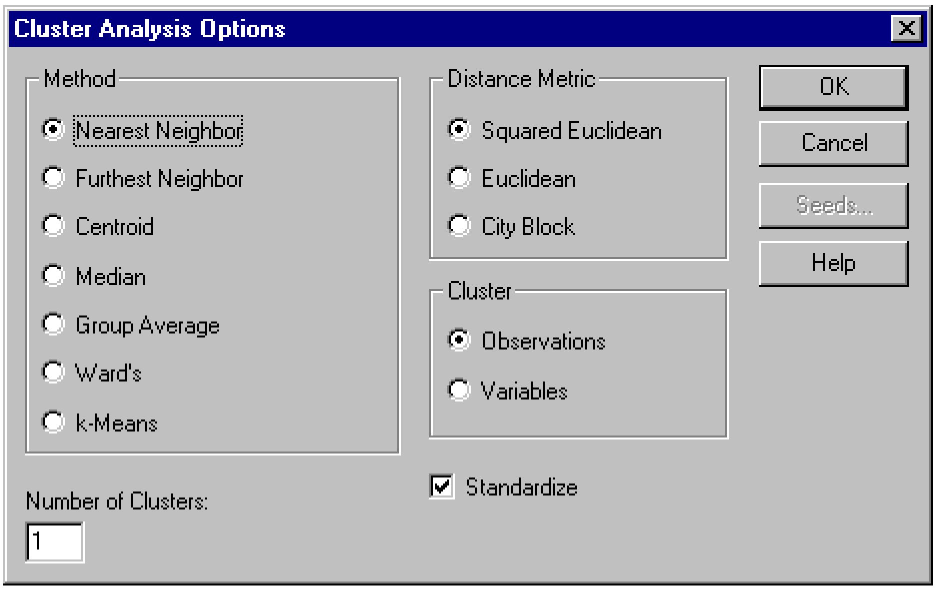


Рис. 5.4. Окно выбора метода кластерного анализа

Откроется диалоговое окно для ввода исследуемых переменных. После нажатия кнопки *OK*. На экране раскроется окно с суммарной статистикой кластерного анализа: количество исследованных объектов, метод выделения кластеров и количество выделенных кластеров. В таблице указаны центроиды выделенных кластеров. Для внесения изменений параметров анализа с помощью контекстного меню (правая кнопка мыши) выберите пункт *Analysis Options*…. Откроется окно ввода параметров проведения кластерного анализа (рис. 5.4).

По умолчанию выделение кластеров реализовывается методом ближайшего соседа, выделяется один кластер и используется в качестве меры расстояния квадрат Евклидова расстояния. Внесите изменения параметров и нажмите кнопку *OK*.

Для вывода результатов развернутого кластерного анализа (например, как объекты распределились по выделенным кластерам) нажмите кнопку табличных опций . Откроется окно выбора, представленное на рис. 5.5.

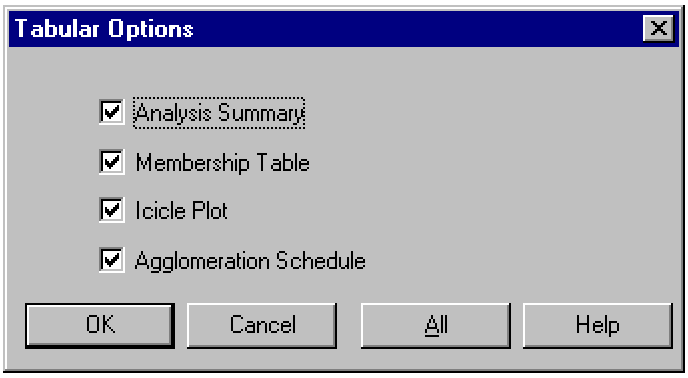


Рис. 5.5. Окно табличных опций кластерного анализа

Для вывода графической информации (например, иерархической дендограммы) нажмите кнопку графических опций . Откроется окно выбора, представленное на рис. 5.6.

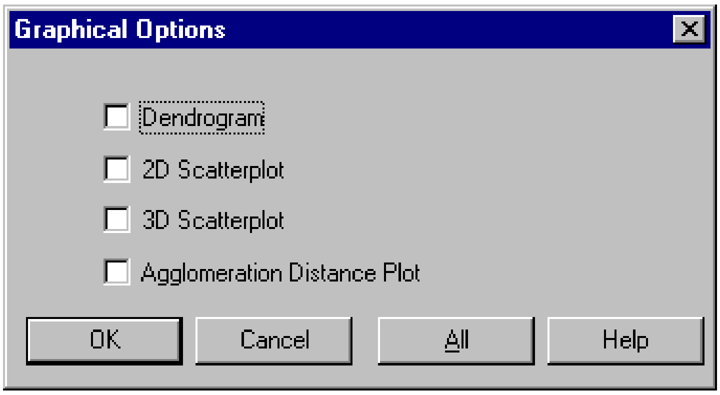


Рис. 5.6. Окно выбора графических опций кластерного анализа

Пакет SPSS

После загрузки пакета выберите рабочий файл с исходными данными. В меню выберите пункт *Analyze – Classify*. Далее исследователю предлагается выбрать один из методов проведения кластерного анализа (рис. 5.7). Рассмотрим каждый.

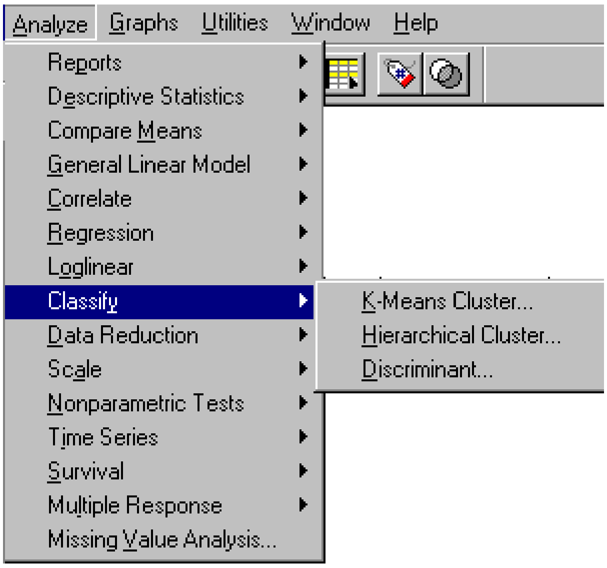


Рис. 5.7. Выбор метода кластеризации

Кластеризация методом k-средних

Откроется диалоговое окно для ввода исходных данных и параметров проведения кластерного анализа. Здесь возможно выбрать метод выделения кластеров либо с помощью итерационной процедуры классификации, либо классификацию с заданием количества выделяемых кластеров (рис. 5.8). Нажав кнопку Options..., можно задать вывод статистики и способ обработки пропущенных значений в исходных данных.

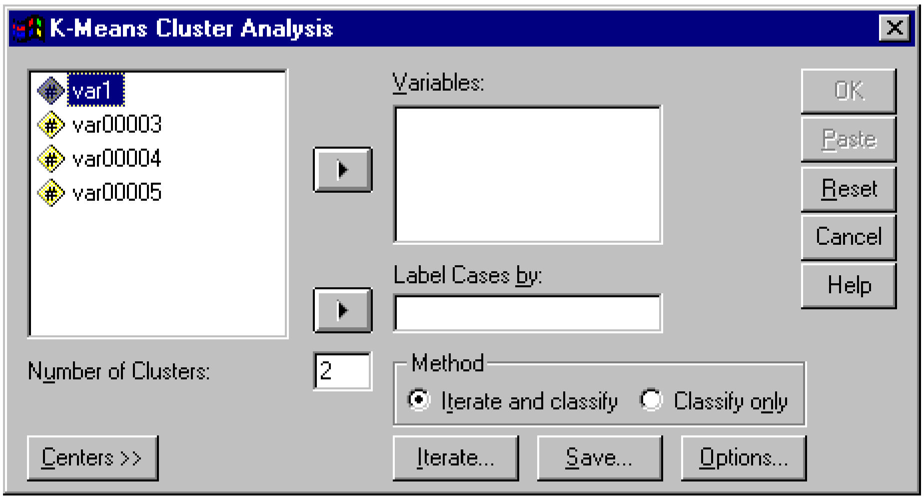


Рис. 5.8. Окно ввода параметров кластеризации методом k-средних

Иерархическая кластеризация

При выборе данного метода кластеризации открывается окно ввода параметров кластерного анализа, представленное на рис. 5.9. В данном окне задаются следующие параметры: проводить кластеризацию объектов наблюдения или признаков, описывающих состояние исследуемых объектов; вывод на экран статистической информации и графиков; выводимой статистической информации Statistics; параметры выводимой графической информации, задаваемые в окне Plots; выбрать метод кластеризации, нажав кнопку Method.

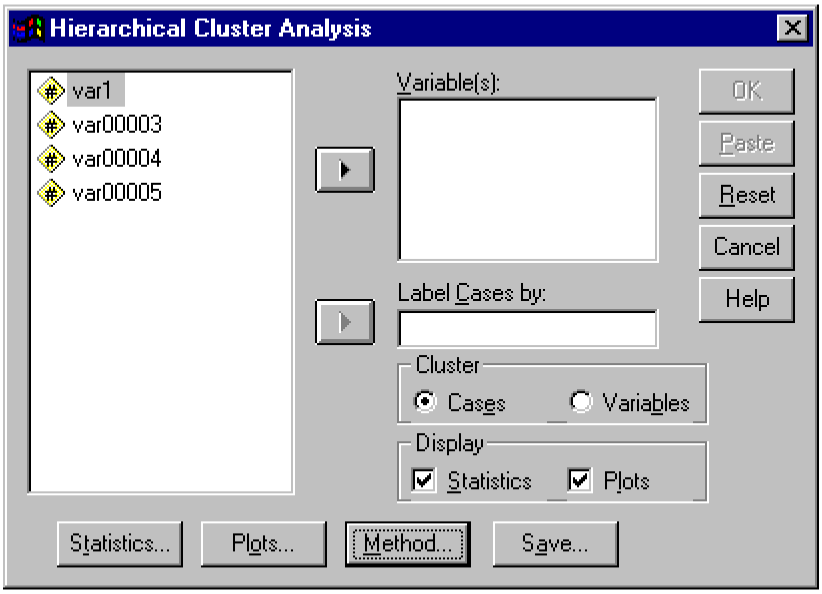


Рис. 5.9. Окно ввода исходных данных иерархической кластеризации

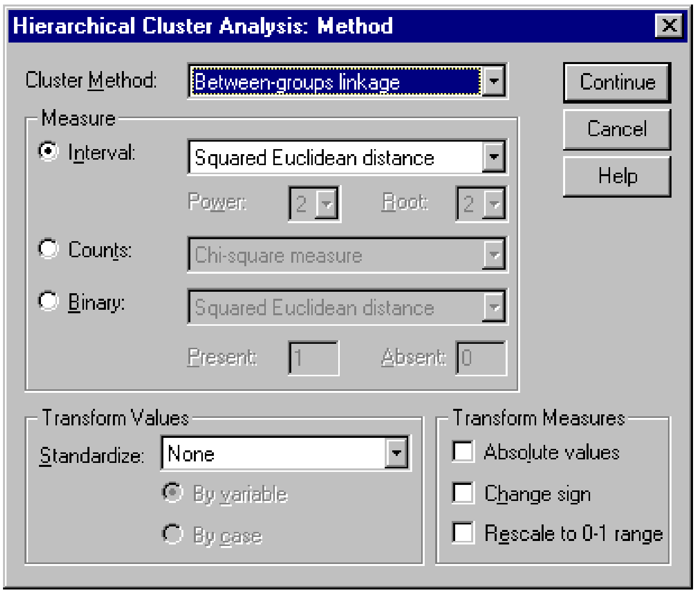


Рис. 5.10. Выбор метода и задание параметров иерархической кластеризации

Ввод параметров проведения иерархической кластеризации (рис. 5.10) требует задания следующих значений: выбор метода кластеризации, выбрать меру расстояния в соответствии со шкалой измерения показателей; стандартизацию данных; и в блоке Transform Measures сохранение знака корреляции.

Пакет STATISTICA

После загрузки модуля кластерного анализа на экране появится окно выбора метода кластеризации: иерархическая кластеризация, кластеризация методом *k*-средних и метод двувходового объединения (рис. 5.11). Двувходовое объединение описано в электронном учебнике, с которым можно ознакомиться на сайте http://www.statsoft/home/textbook.

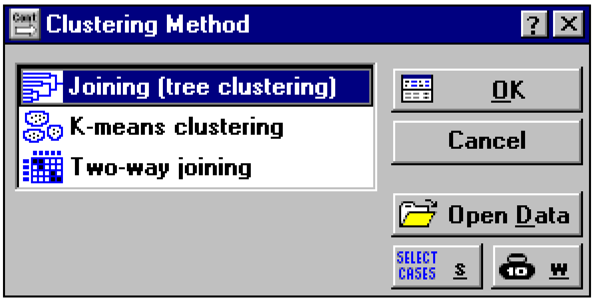


Рис. 5.11. Выбор метода кластеризации в пакете STATISTICA

Иерархическая кластеризация

В окне ввода параметров кластеризации (рис. 5.12) можно изменить следующие параметры:

- исходные данные (*Raw data*). Они могут быть в виде матрицы исследуемых данных и в виде матрицы расстояний (*Distance matrix*);
- Кластеризацию объектов (Cases (*raw*)) или показателей (*Variable (columns)*), описывающих состояние объекта;
- Меры расстояния (*Distance measure*). Здесь возможен выбор следующих мер: евклидово расстояние (*Euclidean distances*), квадрат евклидова расстояния (*Squared Euclidean distances*), расстояние городских кварталов (манхэттенское расстояние) (*City-block (Manhattan) distance*), расстояние Чебышева (*Chebychev distance metric*), степенное расстояние (*Power…*), процент несогласия (*Percent disagreement*);
- Метод кластеризации (*Amalgamation (linkage) rule*). Здесь возможны следующие варианты: одиночная связь (метод ближайшего соседа) (*Single Linkage*), полная связь (метод наиболее удаленных соседей) (*Complete Linkage*), невзвешенное попарное среднее (*Unweighted pair-group average*), взвешенное попарное среднее (*Weighted pair-group average*), невзвешенный центроидный метод (*Unweighted pair-group*

centroid), взвешенный центроидный метод (медиана) (*Weighted pair-group centroid (median)*), метод Уорда (*Ward's method*).

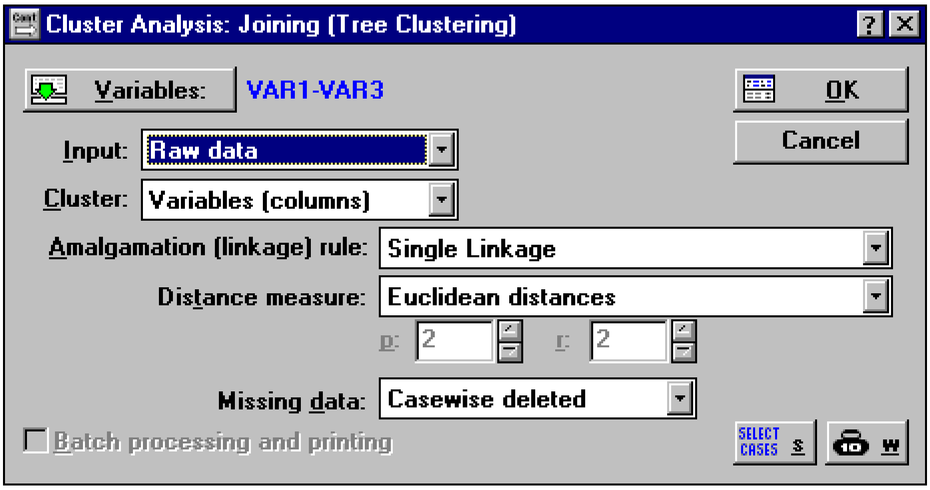


Рис. 5.12. Окно ввода параметров иерархической кластеризации

Кластеризация методом k-средних

На рис. 5.13 представлено окно ввода параметров кластеризации методом *k*-средних. В данном окне выбирается количество итераций или количество выделяемых кластеров, способ обработки пропущенных значений, а также задается центры кластеров.

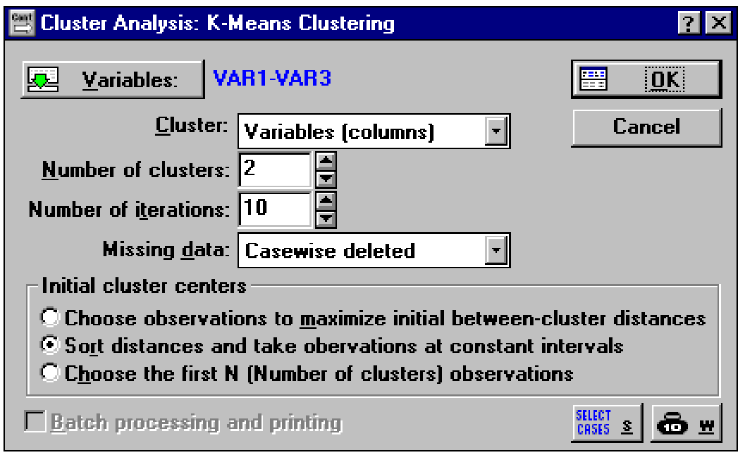


Рис. 5.13. Окно ввода параметров кластеризации методом k-средних

5.6. Примеры использования кластерного анализа при решении задач прикладных исследований

Пример 5.3 [10]

Рассмотрим пример кластерного анализа при исследовании влияния механизмов психологической защиты на эффективность психотерапии избыточного веса (данные Гудкова С.В., СГМУ, Томск 1999). Кластерный анализ проводился с целью выделения однородных групп, т. е. проводилось разбиение множества объектов на заданное число классов (в данном случае – на три класса).

В качестве анализируемых признаков выступали механизмы психологической защиты, измеренные по методике индекса жизненного стиля (ИЖС). Методика ИЖС дает возможность оценить характер трудностей, которые могут возникнуть в процессе психотерапии. Информация, получаемая с помощью методики для диагностики механизмов психологической защиты, дополняет картину личностных особенностей пациента характеристиками стиля его защитного реагирования, анализ которого позволяет использовать в процессе психотерапии разнообразные техники повышения эффективности преодоления тревоги.

В опроснике ИЖС приняты следующие обозначения, соответствующие 8 измеряемым защитным механизмам: А – отрицание, В – подавление, С – регрессия, D – компенсация, Е – проекция, F – замещение, G – интеллектуализация, Н – реактивация.

Кластерный анализ проводился по этим восьми показателям в трех группах пациентов: первый раз на приеме у психотерапевта (первая контрольная точка), группа поддержки (вторичное появление), нормативная группа [10].

Результаты кластерного анализа, приведенные в табл. 5.13–5.15, позволили выделить три основных типа защитной реакции в нормативной группе (табл. 5.13) и в группе женщин с избыточным весом (табл. 5.15).

Таблица 5.13

Механизмы психологической защиты (1-я контрольная точка)

Кластер	Центроиды кластеров								
	A	B	C	D	E	F	G	H	% людей
1	76,46	47,77	57,39	46,77	49,46	41,77	48,85	71,15	28,3
2	59,63	48,88	85,00	79,88	55,81	73,31	46,75	63,88	34,8
3	86,24	49,94	67,47	86,71	68,94	58,94	81,47	85,24	36,9

Таблица 5.14

Механизмы психологической защиты (группа поддержки)

Кластер	Центроиды кластеров								
	A	B	C	D	E	F	G	H	% людей
1	56,89	45,06	77,22	86,72	57,67	82,39	37,11	72,00	56,3
2	77,67	49,17	67,17	38,50	34,83	50,17	58,00	70,83	18,8
3	86,25	57,25	62,38	85,25	78,75	43,38	73,75	93,38	25,0

Таблица 5.15

Механизмы психологической защиты (для нормы)

Кластер	Центроиды кластеров								
	A	B	C	D	E	F	G	H	% людей
1	42,33	44,50	88,33	72,00	57,67	69,67	44,5	58,67	23,1
2	68,69	68,39	65,77	83,39	52,62	54,62	77,92	83,39	50,0
3	74,14	34,71	50,43	35,86	26,29	41,71	34,86	45,14	26,9

Так, в нормативной группе были выделены следующие типы психологической защиты (см. табл. 5.15):

I тип (кластер 1) – с преобладанием защитного механизма *C* (регрессия);
II тип (кластер 1) – с преобладанием защитных механизмов *D* (компенсация) и *H* (реактивация);
III тип (кластер 3) – с преобладанием защитного механизма *A* (отрицание).

В группе женщин с избыточным весом основными типами психологической защиты оказались следующие (см. табл. 5.13):

I тип (кластер 2) – с преобладанием защитного механизма *C* (регрессия);
II тип (кластер 3) – с преобладанием защитных механизмов *A*, *D* и *H*;
III тип (кластер 1) – с преобладанием защитного механизма *A* (отрицание).

Как видно из полученных результатов, I и III типы в нормативной группе и группе женщин с избыточным весом оказались подобными. А тип II отличается только тем, что в группе женщин с избыточным весом к механизмам защиты *D* и *H* добавляется *A* (отрицание). Однако если в нормативной группе преобладающим является II тип психологической защиты (50 % женщин), то в группе с избыточным весом все три типа встречаются практически одинаково часто.

Анализируя данные, приведенные в табл. 5.14, можно сделать вывод о том, как изменились механизмы психологической защиты в группе женщин с избыточным весом после проведения психокоррекционных методик снижения веса.

Количество пациентов с I типом психологической защиты снизилось с 28,26 % до 18,75 %. Самым многочисленным оказался новый тип (кластер 1, табл. 5.14) – с преобладанием механизмов D (компенсация) и F (замещение).

На рис. 5.14 приведено отображение кластеров в виде звездчатых диаграмм, представляющих собой обобщенный графический центроидный образ. Принцип представления данных здесь сводится к построению графов, состоящих из лучей, выходящих из одной точки, концы которых соединяются между собой таким образом, что получаются «звездчатые диаграммы».

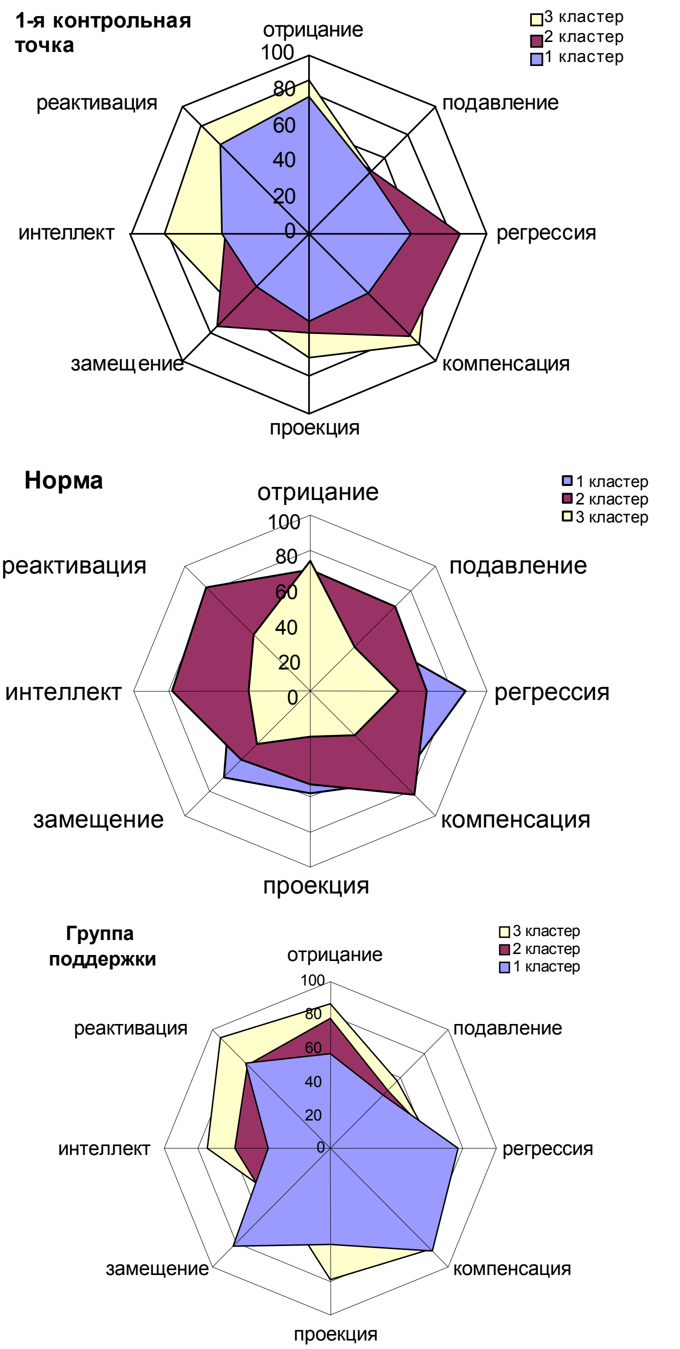


Рис. 5.14. Представление кластеров
в виде звездчатыхдиаграмм для 3-х групп

Анализ полученных результатов позволил авторам реализовать индивидуальный подход при подборе индивидуальных методов психотерапевтических воздействий [10].

Пример 5.4 [19]

Продолжим оценку учебных дисциплин с помощью кластерного анализа (см. пример 5.2).

На рис. 5.15. представлена дендограмма, построенная на основе данных анкетирования студентов IV курса АВТФ ТПУ.

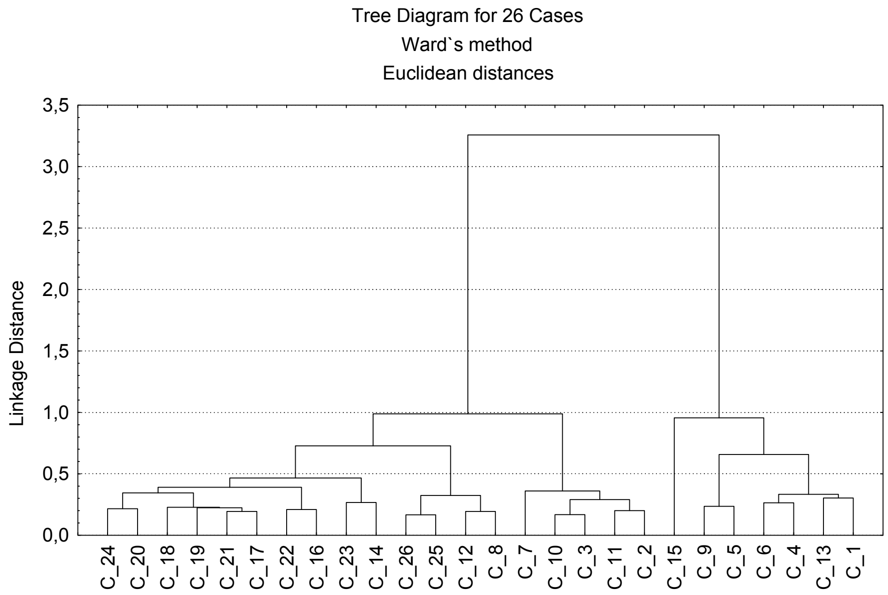


Рис. 5.15. Дендрограмма, построенная
по результатам анкетирования студентов IV курса АВТФ ТПУ

Объекты кластеризации (изучаемые предметы) представлены в табл. 5.15. Признаки объектов кластеризации приведены в табл. 5.17.

Таблица 5.16

Объекты кластеризации

Наименование предмета	Код	Наименование предмета	Код
Базы данных и экспертные системы	C_1	Безопасность жизнедеятельности	C_14
Геометрия и алгебра	C_2	Иностранный язык	C_15
Дифференциальные уравнения	C_3	История	C_16
Информатика	C_4	Культурология	C_17
Информационные технологии	C_5	Политология	C_18
Компьютерная графика	C_6	Правоведение	C_19
Математический анализ	C_7	Психология и педагогика	C_20
Основы теории случайных функций	C_8	Социология	C_21
Практикум на ЭВМ	C_9	Философия	C_22
Теория вероятностей и мат. статистика	C_10	Экология	C_23
Физика	C_11	Экономика	C_24
Численные методы	C_12	Элективные курсы по истории и культурологии	C_25
Языки программирования и методы трансляций	C_13	Элективные курсы по философии	C_26

Таблица 5.17

Признаки объектов кластеризации

RED	Ассоциация с красным – четко выраженные эмоции: активность, интерес, побуждение
YELLOW	Ассоциация с желтым – безусловное восприятие, комфорт
GREEN	Ассоциация с зеленым – интерес, поиск смысла для себя, полезности
MAGENTA	Ассоциация с фиолетовым – поиск необычного в новой информации, сложное восприятие
BLUE	Ассоциация с синим – основательность, надежность, прочность
BROUN	Ассоциация с коричневым – стабильность, консерватизм, жесткость позиции, ригидность, жесткость структуры мышления
GREY	Ассоциация с серым – невыраженное отношение, неясность, неопределенность, «все равно»
BLACK	Ассоциация с черным – отрицание, отторжение, негативное восприятие
WHITE	Ассоциация с белым – невыраженное восприятие
VOPR2	Заинтересованность
VOPR3	Необходимость изменения
VOPR4	Недостаточность литературы и пособий
VOPR6	Отсутствие интереса
VOPR7	Трудность изучения
THEORY%	Недостаточно теории
PRACTIC%	Недостаточно практики
TP%	Недостаточно теории и практики
NORMA%	Дисциплина включает достаточное количество часов
VOPR9	Недостаточность наглядных пособий
VOPR10	Желание иметь раздаточный материал
VOPR11	Высокий уровень знаний
VOPR12	Низкий уровень знаний
VOPR13	Желание преподавать данную дисциплину

Как видно из дендограммы (рис. 5.15), все учебные дисциплины по исследуемым признакам можно разбить на четыре кластера. Для уточнения результата была проведена кластеризация по методу *k*-средних. Число искомых кластеров задавалось равным 4.

Во втором столбце табл. 5.18 размещен список переменных, далее идут суммы квадратов (*SS*) и степени свободы (*df*), затем *F*-критерий Фишера и в восьмом столбце – достигнутый уровень значимости (*p*).

Табл. 5.18 дисперсионного анализа результатов кластеризации на четыре кластера показывает необходимость отклонения нулевой гипотезы о равенстве групповых средних почти по всем 23 признакам, за исключением переменных *RED* (ассоциация с красным – четко выраженные эмоции: активность, интерес, побуждение), *YELLOW* (ассоциация с желтым – безусловное восприятие, комфорт), *MAGENTA* (ассоциация с фиолетовым – поиск необычного в новой информации, сложное восприятие), *BROUN* (ассоциация с коричневым – стабильность, консерватизм, жесткость позиции, ригидность, жесткость структуры мышления), для которых достигнутый уровень значимости оказался более 0,05.

Таблица 5.18

Дисперсионный анализ переменных

№	Оценка учебных дисциплин	Between SS	df	Within SS	df	F	signif. *p*
1	Активность, интерес, побуждение	,005446	3	,038549	22	1,03599	,396077
2	Безусловное восприятие, комфорт	,007304	3	,023612	22	2,26860	,108774
3	Интерес, поиск смысла для себя, полезности	,035614	3	,064794	22	4,03078	,019943
4	Поиск необычного в новой информации, сложное восприятие	,001031	3	,016878	22	,44812	,721113
5	Основательность, надежность, прочность	,033040	3	,028947	22	8,37020	,000673
6	Стабильность, консерватизм, жесткость позиции, ригидность, жесткость структуры мышления	,006025	3	,024059	22	1,83640	,170040
7	Невыраженное отношение, неясность, неопределенность, «все равно»	,030842	3	,048681	22	4,64599	,011579
8	Отрицание, отторжение, негативное восприятие	,029274	3	,028746	22	7,46789	,001261
9	Невыраженное восприятие	,007360	3	,017377	22	3,10612	,047317
10	Заинтересованность	,573883	3	,098435	22	42,75393	,000000
11	Необходимость изменения	,104540	3	,058297	22	13,15033	,000039
12	Недостаточность литературы и пособий	,160605	3	,021078	22	55,87568	,000000
13	Отсутствие интереса	,352482	3	,085194	22	30,34109	,000000
14	Трудность изучения	,058239	3	,034778	22	12,28044	,000063
15	Недостаточно теории	,008238	3	,009344	22	6,46575	,002643
16	Недостаточно практики	,177993	3	,043498	22	30,00765	,000000

Окончание табл. 5.18

№	Оценка учебных дисциплин	Between SS	df	Within SS	df	F	signif. *p*
17	Недостаточно теории и практики	,003338	3	,001009	22	24,25641	,000000
18	Дисциплина включает достаточное количество часов	,294106	3	,072314	22	29,82519	,000000
19	Недостаточность наглядных пособий	,099495	3	,079348	22	9,19534	,000391
20	Желание иметь раздаточный материал	,095418	3	,055622	22	12,58011	,000053
21	Высокий уровень знаний	,139903	3	,144325	22	7,10864	,001636
22	Низкий уровень знаний	,031697	3	,060723	22	3,82791	,023987
23	Желание преподавать	,064740	3	,037028	22	12,82163	,000047

Ниже приведен график (рис. 5.16) средних значений всех переменных по отдельным кластерам.

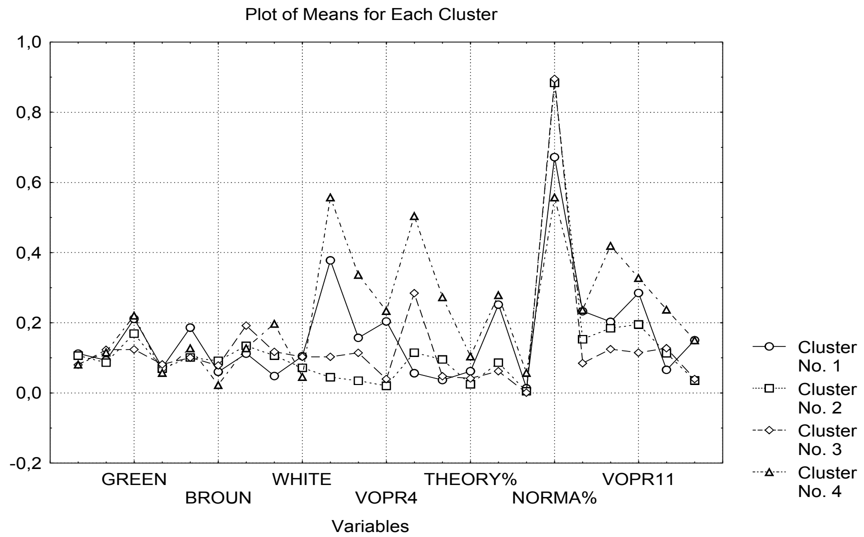


Рис. 5.16. График средних значений переменных каждого кластера

В табл. 5.19 приведены результаты кластеризации – размещение объектов по кластерам и расстояния до центра кластера.

КЛАСТЕР № 1 включает в себя дисциплины, связанные с будущей профессией (компьютерными технологиями). Отношение студентов к этим предметам в целом характеризуется как положительное – спокойная эмоциональная окраска, частое обсуждение с друзьями, интерес, достаточно высокий уровень знаний, желание преподавать один из предметов в будущем. Наряду с этим, отмечается: желание изменить содержание предмета; недостаток методических пособий и литературы, раздаточного материала; недостаточное количество практических занятий, что может свидетельствовать о заинтересованности в получении как можно более глубоких знаний по данным дисциплинам.

Таблица 5.19

Результаты для каждого кластера

№	Объект	Distance
Кластер № 1 содержит 6 объектов		
1	Базы данных и экспертные системы	,034294
2	Информатика	,051216
3	Информационные технологии	,056035
4	Компьютерная графика	,052642
5	Практикум на ЭВМ	,055987
6	Языки программирования и методы трансляций	,043590
Кластер № 2 содержит 7 объектов		
1	Геометрия и алгебра	,039558
2	Дифференциальные уравнения	,025011
3	Математический анализ	,056149
4	Основы теории случайных функций	,058992
5	Теория вероятностей и мат. статистика	,027518
6	Физика	,034542
7	Численные методы	,028838
Кластер № 3 содержит 12 объектов		
1	Безопасность жизнедеятельности	,053897
2	История	,037043
3	Культурология	,024812
4	Политология	,032687
5	Правоведение	,031661
6	Психология и педагогика	,044381
7	Социология	,033310
8	Философия	,048582
9	Экология	,048356
10	Экономика	,045697
11	Элективные курсы по истории и культурологии	,048925
12	Элективные курсы по философии	,043968
Кластер № 4 содержит 1 объект		
1	Иностранный язык	0,00

КЛАСТЕР № 2 включает в себя базовые математические дисциплины, относящиеся к разряду «сложных». Из-за трудного понимания возникает желание пропустить эти предметы, ассоциация предметов с мрачными или серыми цветами, иногда низкий уровень знаний по какому-либо предмету (дифференциальные уравнения, мат. статистика). Здесь же в то же время желание увеличить количество практических и теоретических занятий, понимание необходимости изучения данных предметов.

КЛАСТЕР № 3 включает предметы, имеющие гуманитарную направленность. Отношение к этим предметам можно охарактеризовать как «поверхностное» – предметы «не основные», их можно и не учить, они «довольно легкие». Не выражено желание посещать занятия и, напротив, стремление к частым пропускам. Специфику отношения к данным предметам диктует техническая направленность обучения студентов опрашиваемых специальностей.

КЛАСТЕР № 4 включает всего один предмет – иностранный язык, который действительно занимает особое место среди прочих. Наблюдается активная заинтересованность предметом, наряду с резким недовольством содержания курса и методами его преподавания.

Общие результаты проведенного кластерного и факторного анализа представлены в сводной табл. 5.20.

Таблица 5.20

Совмещенные результаты факторного и кластерного анализов

	Фактор 1	Фактор 2	Фактор 3	Фактор 4
Кластер 1	x		x	
Кластер 2		x		x
Кластер 3		x	x	
Кластер 4	x	x		x

По полученным результатам, авторами был сделан следующий вывод:

Все оцениваемые дисциплины условно разделились на четыре группы, которые можно охарактеризовать четырьмя признаками.

Самая «благополучная» группа – это та, к которой принадлежат дисциплины, связанные с компьютерными технологиями. К этой группе дисциплин практически нет отрицательных замечаний, проявляется высокая заинтересованность.

Особое место заняла одна дисциплина – иностранный язык. Отношение неоднозначное: с одной стороны заинтересованность, высокий уровень знаний, а с другой – желание пропустить занятия, желание изменить содержание. Это говорит о низком качестве преподавания дисциплины.

Кроме того, было выявлено негативное отношение студентов к предметам, имеющим физико-математическую направленность и равнодушное отношение к гуманитарным дисциплинам.

Высокий показатель факторов «лень» и «не интересно» в совокупности с вышеперечисленными выводами позволяет сделать заключение о том, что характер образовательного процесса поверхностный, быстрый. Студент ориентирован на «быстрое» получение знаний по престижным компьютерным технологиям (при этом осознает необходимость изучения иностранного языка). Его не интересуют фундаментальные науки (математика, физика), отсутствие интереса к гуманитарным дисциплинам говорит о нежелании развиваться духовно.

СПИСОК ЛИТЕРАТУРЫ

1. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика, 1974. – 200 с.
2. Александров В.В., Алексеев А.И., Горский Н.Д. Анализ данных на ЭВМ (на примере системы СИТО). – М.: Финансы и статистика, 1990. – 192 с.
3. Аркадьев А.Г., Браверман Э.М. Обучение машины классификации объектов. – М.: Наука, 1971. – 172 с.
4. Артемьева Е.Ю., Мартынов Е.М. Вероятностные методы в психологии. – М.: Изд-во МГУ, 1985. – 206 с.
5. Берестнева О.Г., Дубинина И.А., Марухина О.В. Применение факторного анализа для оценки персонала предприятия // Материалы международной научно-технической конференции «Современные технологии и технологии в медицине и биологии». – Новочеркасск: НАБЛА, 2001. – С. 15–16.
6. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М.: Прогресс, 1976. – 495 с.
7. Гублер Е.В., Генкин А.А. Применение непараметрических критериев статистики в медико-биологических исследованиях. – Л.: Медицина, 1973. – 142 с.
8. Гублер Е.В. Алгоритм оценки расхождения распределений признаков в медицинских автоматизированных системах // Проблемы системотехники и автоматизированные системы управления. – Л.: Медицина, 1978. – 230 с.
9. Гублер Е.В. Вычислительные методы анализа и распознавания патологических процессов. – Л.: Медицина, 1978. – 296 с.
10. Гудков С.В., Бобровский А.В., Гордиенко А.В. и др. Влияние механизмов психологической защиты на эффективность психотерапии избыточного веса // Молодежь и наука: Проблемы и перспективы: Доклады III межвуз. науч. конф. студентов, аспирантов и молодых ученых. – Томск, 1999. – С. 168–171.
11. Гусев А.И., Измайлов С.А., Михалевская М.Б. Измерение в психологии: общий психологический практикум. – М.: Смысл, 1997. – 287 с.
12. Дюк В.А. Компьютерная психодиагностика. – СПб.: Братство, 1994. – 364 с.
13. Иберла К. Факторный анализ. – М.: Статистика, 1980. – 308 с.
14. Иванкина Л.И., Марухина О.В., Берестнева О.Г. Оценка учебных дисциплин как критерий качества образовательного процесса / деп. ВНИИ Высшего образования, № 240–2001, 2001. – 18с.

15. Лакин Г.Д. Биометрия. – М.: Высшая школа, 1980. – 293 с.
16. Логвиненко А.Д. Измерения в психологии: Математические основы. – М.: Изд-во МГУ, 1993. – 476 с.
17. Лоули Д., Максвелл А. Факторный анализ как статистический метод. – М.: Мир, 1967. – 144 с.
18. Мандель И.Д., Черный Л.М. Экспериментальное сравнение алгоритмов кластер-анализа // Автоматика и телемеханика. – 1988. – № 11. – С. 42–48.
19. Моисеева Н.И., Сысуев В.М. Временная среда и биологические ритмы. – Л.: Наука, 1981. – 128 с.
20. Налимов В.В. Теория эксперимента. – М.: Наука, 1971. – 208 с.
21. Окунь Я. Факторный анализ. – М.: Статистика, 1974. – 198 с.
22. Практикум по общей, экспериментальной и прикладной психологии / под ред. А.А. Крылова, С.А. Маричева. – СПб.: Питер, 2000. – 560 с.
23. Плохинский Н.А. Биометрия. 2-е изд. – М.: МГУ, 1970. – 368 с.
24. Сидоренко Е.В. Математические методы в психологии. – СПб.: Изд-во «Социально-психологический центр», 1996. – 346 с.
25. Суходольский Г.В. Основы математической статистики для психологов. – Л.: ЛГУ, 1972. – 428 с.
26. Справочник по прикладной статистике: в 2 т. / под ред. Э. Ллойда, У. Ледермана, Ю.Н. Тюрина. – М.: Финансы и статистика, 1989. – 510 с.
27. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере / под ред. В.Э. Фигурнова. – М.: ИНФРА-М: Финансы и статистика, 1995. – 384 с.
28. Уразаев А.М., Кулаков Ю.А., Медведев М.А. Физиологические закономерности адаптационных процессов, опережающих условия деятельности // Доклады АН СССР. – 1987. – Т. 295. – № 6. – С. 1509–1512.
29. Факторный, дискриминантный и кластерный анализ / под ред. И.С. Енюкова. – М.: Финансы и статистика, 1989.
30. Харман Г. Современный факторный анализ. – М.: Статистика, 1972. – 486 с.
31. Шатилов А.Ю., Рихванов Л.П., Муратова Е.А. и др. Математическая обработка геохимических данных при оценке состояния окружающей среды (на примере мониторинга загрязнения снегового покрова) // «Геоинформатика – 2001». Труды Международной научно-практической конференции. – Томск, 2000. – С. 199–204.

ОГЛАВЛЕНИЕ

Учебное издание

БЕРЕСТНЕВА Ольга Григорьевна
МАРУХИНА Ольга Владимировна
ШЕВЕЛЁВ Геннадий Ефимович

ПРИКЛАДНАЯ МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Выпускающий редактор *Т.С. Савенкова*
Редактор *Е.Л. Тен*
Компьютерная верстка *К.С. Чечельницкая*
Дизайн обложки *А.И. Сидоренко*

Подписано к печати 05.06.2012. Формат 60х84/16. Бумага «Снегурочка».
Печать XEROX. Усл. печ. л. 10,93. Уч.-изд. л. 9,89.
Заказ 740-12. Тираж 100 экз.

Национальный исследовательский Томский политехнический университет
Система менеджмента качества
Издательства Томского политехнического университета сертифицирована
NATIONAL QUALITY ASSURANCE по стандарту BS EN ISO 9001:2008

ИЗДАТЕЛЬСТВО ТПУ. 634050, г. Томск, пр. Ленина, 30
Тел./факс: 8(3822)56-35-35, www.tpu.ru