

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования  
«МОГИЛЕВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени А. А. КУЛЕШОВА»

Н. В. Акулич,  
А. В. Сорока

# СТАТИСТИКА ДЛЯ БИОЛОГОВ

Методические рекомендации



Могилев 2015

Электронный архив библиотеки МГУ имени А.А. Кулешова

*Электронный аналог печатного издания:*

Н. В. Акулич, А. В. Сорока

Статистика для биологов

Могилев : МГУ имени А. А. Кулешова, 2015. – 24 с. : ил.

Издание предназначено для применения статистики в биологических исследованиях. Описываемый подход основан на использовании прикладных программ для персонального компьютера, однако изложенные принципы применения статистики универсальны и могут применяться при использовании любых статистических вычислений.

УДК 311.3:57(075.8)  
ББК 60.652.8

Акулич, Н. В. Статистика для биологов [Электронный ресурс] : методические рекомендации / Н. В. Акулич, А. В. Сорока. – Электрон. данные. – Могилев : МГУ имени А. А. Кулешова, 2015. – Загл. с экрана.

212022, г. Могилев  
ул. Космонавтов, 1  
тел.: 8-0222-28-31-51  
e-mail: alexpzn@mail.ru  
<http://www.msu.mogilev.by>

© Акулич Н. В., Сорока А. В., 2015  
© МГУ имени А. А. Кулешова, 2015  
© МГУ имени А. А. Кулешова,  
электронный аналог, 2015

## ВВЕДЕНИЕ

Для анализа данных в биологических экспериментальных исследованиях при написании курсовых и дипломных работ требуется использование статистических исследований. Строго говоря, такой подход довольно нов и ранее если для каких-либо выводов требовался статистический анализ данных, то это значило, что сам эксперимент был плох (в том смысле, что полученный результат неочевиден, например, зрачок не расширился, лапка не дернулась и т.д.).

Современная наука достигла очень высокого уровня, и поэтому для постижения ее новых тайн необходима (читай обязательно) статистика. Современная биология без опоры на математический аппарат не может обеспечить точный и логически строгий анализ экспериментальных данных. Только с помощью правильно выбранного статистического метода можно описать, объяснить и углубленно исследовать всю совокупность результатов измерений. Но бывает, и наоборот, в курсовых и дипломных работах студенты пренебрегают самым элементарным статистическим анализом, ограничиваясь приростом показателей в процентах и, на этом основании делают вывод, что поскольку применение комплексной добавки приводит к 45% росту всхожести и еще чему-то там, то это чудо добавка на что-то там влияет.

Таким образом, цель настоящего пособия, во-первых предложить студентам простой инструмент для анализа полученных результатов, а, во-вторых, уменьшить количество научных данных с невысокой научной ценностью.

# 1. ОБЩИЕ СВЕДЕНИЯ О СТАТИСТИКЕ

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

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

## 1.1. Определение цели исследования

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

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

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

## 1.2. Планирование исследования

Так сложилось, что исследования планируются довольно редко, чаще они просто проводятся, а затем, для наукообразия проводят статистический анализ, чаще всего «считают по Стюденту».

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

Планирование исследования в общем виде можно разбить на 2 этапа:

1. Определение типа исследования.
2. Определение объемов выборок.

### Типы исследований

**Описание отдельных случаев** – наиболее старый способ исследования. Он состоит в описании редкого наблюдения или феномена. Научные

гипотезы в таком исследовании не выдвигаются и не проверяются. Однако данный способ исследования также важен, так как описание редких случаев или явлений нельзя недооценивать.

**Описание серии случаев** – исследование, включающее обычно описательную статистику группы объектов, отобранных по какому-либо признаку. Описательные исследования используются, например, в экологии для изучения влияния неконтролируемых факторов на биоту.

**Исследование случай-контроль** – ретроспективное исследование, в котором по архивным данным формируют группы по какому-либо признаку, а затем ретроспективно оценивают воздействие предполагаемого фактора. Например, преподаватель предлагает оценить по данным лесхозов, санстанций и т.п. влияние проезжающих по лесной дороге автомобилей на частоту встречаемости одуванчиков. Такие исследования чаще выдвигают научные гипотезы, а не проверяют их. Преимуществом исследования данного типа являются его относительная простота, дешевизна и быстрота выполнения.

**Одномоментное исследование** – описательное исследование, включающее однократно обследуемые группы участников и проводимое с целью оценки распространенности того или иного исхода, течения заболевания, а также эффективности диагностики.

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

**Проспективное исследование** – исследование, в котором выделенная группа участников наблюдается в течение определенного времени. Сначала выделяют группу (например, дети, находящиеся в экологическом лагере и дети, находящиеся дома), а затем проводят наблюдение за ними и осуществляют сбор данных. В этом заключается отличие от ретроспективного исследования, в котором когорты выделяют после сбора данных.

В биологических экспериментальных исследованиях может быть использован особый тип структуры исследования – **активный эксперимент**. Именно исследования такого типа являются классическим научным экспериментом, их результаты являются наиболее доказательными.

При планировании исследования следует стремиться избегать систематических и случайных ошибок.

**Систематическая ошибка** – систематическое (неслучайное, одностороннее) отклонение результатов от истинных значений. К основным видам систематических ошибок относятся следующие.

**Систематическая ошибка, возникающая при отборе.** Эта систематическая ошибка возникает в тех случаях, когда сравниваемые группы участников исследования различаются не только по главным изучаемым признакам, но и по другим факторам, влияющим на результат исследо-

вания, т.е. участники фактически набираются из разных популяций. Эта ошибка возникает на этапе формирования исследуемых групп.

**Систематическая ошибка, возникающая при измерении.** Она возникает, когда объекты в сравниваемых группах обследованы в разной степени (разные методы исследования, частота обследований) или используются нестандартизованные схемы получения данных и субъективные оценки.

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

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



Первым шагом, предваряющим статистический анализ данных, является анализ типов данных. Это необходимо делать для того, чтобы определить правомочность использования в дальнейшем того или иного способа представления данных и статистического метода. Тип получаемых данных необходимо также учитывать на этапе планирования исследования при определении необходимых объемов выборок.

Принято выделять в качестве основных типов данных количественные и качественные.

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

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

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

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

**Дискретные данные** – количественные данные, которые не могут иметь дробную часть. Пример: количество конечностей у особи.

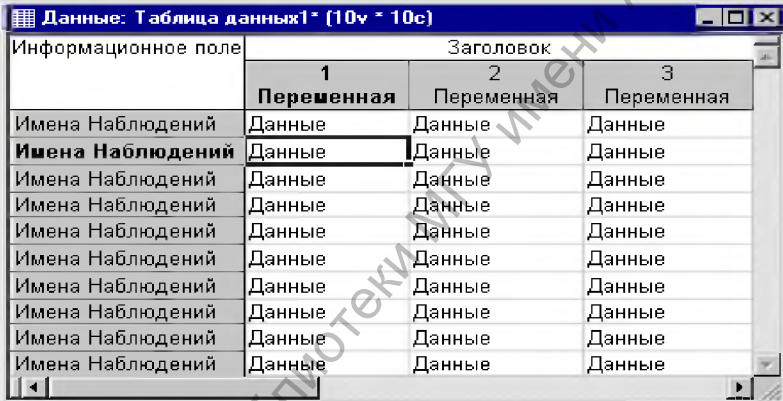
**Качественные данные** – отражают условные коды неизмеряемых категорий, или условную степень выраженности какого-либо признака (например, условные единицы тропической сети). Их основное отличие от дискретных количественных данных заключается в отсутствии пропорциональной шкалы для измерения выраженности признака.

**Бинарные (дихотомические) данные** – особо выделяемый вид качественных данных. Признак такого типа имеет лишь два возможных значения (например, пол).

## 2. ПОДГОТОВКА ДАННЫХ К СТАТИСТИЧЕСКОМУ АНАЛИЗУ

Подготовка данных может осуществляться в любой программе, поддерживающей электронные таблицы, а также в любых базах данных, поддерживающих возможность их экспорта.

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



| Информационное поле | Заголовок    |              |              |
|---------------------|--------------|--------------|--------------|
|                     | 1 Переменная | 2 Переменная | 3 Переменная |
| Имена Наблюдений    | Данные       | Данные       | Данные       |
| Имена Наблюдений    | Данные       | Данные       | Данные       |
| Имена Наблюдений    | Данные       | Данные       | Данные       |
| Имена Наблюдений    | Данные       | Данные       | Данные       |
| Имена Наблюдений    | Данные       | Данные       | Данные       |
| Имена Наблюдений    | Данные       | Данные       | Данные       |
| Имена Наблюдений    | Данные       | Данные       | Данные       |
| Имена Наблюдений    | Данные       | Данные       | Данные       |
| Имена Наблюдений    | Данные       | Данные       | Данные       |

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

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

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



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

## 2.1. Основные принципы и методы статистического анализа

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

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

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

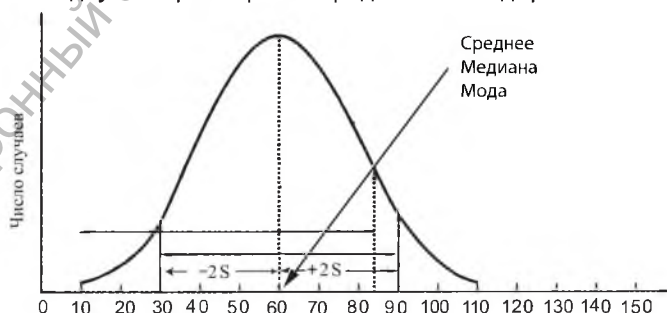
- статистическая оценка параметров распределения;
- проверка статистических гипотез;
- статистическое моделирование.

### Понятия описательной статистики

**Генеральная совокупность** – все множество изучаемых объектов.

**Выборка** – множество объектов, выбранных из генеральной совокупности для изучения, а **объем выборки** – количество.

Нормальное распределение выборки более предпочтительно по следующим причинам. В большинстве случаев оно является хорошим приближением функций. Распределение многих статистик является нормальным или может быть получено из нормальных с помощью некоторых преобразований. Нормальное распределение представляет собой одну из эмпирически проверенных истин относительно общей природы действительности и его положение может рассматриваться как один из фундаментальных законов природы. Форма нормального распределения (характерная определяется только двумя параметрами: средним и стандартным отклонением.



Характерное свойство нормального распределения состоит в том, что 68% всех его наблюдений лежат в диапазоне  $\pm 1$  стандартное отклонение от среднего, а диапазон  $\pm 2$  стандартных отклонения содержит 95% значений.

**Среднее значение** (арифметическое) – сумма, деленная на количество.

Используется также среднее геометрическое и квадратичное.

**Медиана** – значение в середине упорядоченной выборки.

**Мода** – значение, встречающееся в выборке чаще всех. Если таковых нет, тогда для непрерывных величин принимается середина интервала с максимальной частотой.

Мода и медиана часто применяются для описания дискретных величин, где среднее арифметическое может не иметь физического смысла.

**Стандартное отклонение** (среднее квадратическое отклонение) обозначается:  $\sigma$ .

Отражает разброс: если величина больше, значит показатели более разнообразные.

**Дисперсия** – это стандартное отклонение в квадрате ( $\sigma^2$ ).

**Коэффициент вариации** – дополнительный показатель = стандартное отклонение/среднее арифметическое. Применяют: если надо сравнить изменчивость величин измеряемых в разных единицах.

**Статистическая ошибка** – ошибка, связанная с перенесением результатов, полученных при выборке, на всю генеральную совокупность. Она показывает среднее расхождение между исследуемой величиной в выборке. На основе ошибки находится показатель точности. Хорошо, если меньше 3%, удовлетворительно – 3-5%. Чтобы было лучше брать больше выборку.

**Доверительные интервалы** – это диапазон, в котором с заданной вероятностью находятся значения исследуемого показателя для генеральной совокупности.  $P = 0,95$  и  $P = 0,99$ .

### 3. ПРОВЕДЕНИЕ СТАТИСТИЧЕСКОГО АНАЛИЗА

**Пример 1.** (выполнен в пакете анализа EXCEL, далее будет обозначено как Ex).

Определена масса 24 кошек (кг):

|     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 5,3 | 4,5 | 5,2 | 5,6 | 6,0 | 3,8 | 4,7 | 7,3 | 5,7 | 6,3 | 4,1 | 4,9 |
| 5,8 | 6,4 | 4,3 | 5,0 | 3,2 | 5,8 | 6,7 | 4,3 | 5,1 | 5,4 | 5,9 | 5,2 |

Построить таблицу частот и гистограмму.

1. В ячейке A1 ввести заголовок “Кошки”. Чтобы заголовок состоял из нескольких строк, настроить формат ячейки следующим образом. Выбрать ячейку A1. Из меню *Формат* выбрать элемент *Ячейки*. На вкладке *Выравнивание* установить *По горизонтали – По центру*, *По вертикали – По центру*; установить флажок *Переносить по словам*.

2. В ячейках A2-A25 ввести данные (значения масса кошек).

3. В ячейке B1 ввести заголовок “Масса”; в ячейках B2-B7 ввести границы интервалов гистограммы (3, 4, 5, 6, 7, 8).

4. Проверить, установлен ли пакет анализа данных. Для этого выбрать элемент меню *Сервис – Настройки* и убедиться, что установлен флажок *Пакет анализа* (при необходимости установить его).

5. Получить таблицу частот и гистограмму. Для этого выбрать элемент меню *Сервис – Анализ данных*; из полученного списка выбрать инструмент *Гистограмма*. Установить следующие параметры: *Входной интервал – A2:A25*, *Интервал карманов – B1:B7*. Установить флажки *Метки* (так как в ячейках A1 и B1 указаны не данные, а заголовки), *Интегральный процент* (для получения накопленных относительных частот) и *Вывод графика* (для построения гистограммы). Чтобы результаты были выведены на отдельный лист, установить переключатель *Новый рабочий лист*. Для получения результатов нажать кнопку *ОК*.

#### **Пример 2.** (Ex)

По данным задания 1 найти основные статистические показатели для массы кошек (среднее, стандартное отклонение, дисперсию, коэффициент вариации), а также статистическую ошибку и 95-процентный доверительный интервал для среднего, используя функции EXCEL.

1. В ячейках C1,C2,C3 ввести обозначения «Среднее», «Станд. отклон.», «Дисперсия». В ячейках D1,D2,D3 найти среднее, стандартное отклонение и дисперсию, используя функции СРЗНАЧ, СТАНДОТКЛОН и ДИСП соответственно.

2. В ячейке C4 ввести обозначение «Коеф. вариации». В ячейке D4 найти коэффициент вариации, используя следующую формулу:  $=D2/D1$ .

3. В ячейке C5 ввести обозначение «Стат. ошибка». В ячейке D5 найти статистическую ошибку для среднего, используя следующую формулу:  $=D2/КОРЕНЬ(25)$ .

4. Для определения доверительного интервала, в ячейке F6 найти квантиль  $t_{\alpha; n-1}$ . Для этого использовать функцию СТЬЮДРАСПОБР со следующими параметрами: *Вероятность* – 0,05, *Степени свободы* – 24.

5. В ячейке С6 ввести обозначение «Нижняя граница», в ячейке С7 – «Верхняя граница». В ячейке D6 найти нижнюю границу доверительного интервала, используя формулу: =D1-F6\*D5. Аналогично в ячейке D7 найти верхнюю границу доверительного интервала: =D1+F6\*D5.

### Пример 3. (Ex)

Найти статистические показатели, указанные в задании 2, используя инструмент анализа *Описательная статистика*.

1. Выбрать элемент меню *Сервис – Анализ данных*; из полученного списка выбрать инструмент *Описательная статистика*. Установить следующие параметры: *Входной интервал* – A1:A26, *Группирование* – *По столбцам*. Установить флажки *Метки в первой строке* (так как в ячейке A1 указан заголовок), *Итоговая статистика* (для получения всех статистических показателей) и *Уровень надежности* (для получения доверительного интервала). В поле *Уровень надежности* указать доверительную вероятность 95. Установить переключатель *Новый рабочий лист*. Для получения результатов нажать кнопку *ОК*.

2. Определить границы доверительного интервала. Пусть среднее находится в ячейке B2, а уровень надежности – в ячейке B16. Тогда для определения верхней границы доверительного интервала следует в любой свободной ячейке ввести формулу =B2+B16, а для определения нижней границы – =B2-B16.

### Пример 4. (Ex)

Определить, является ли статистически значимым различие между двумя группами кошек по дисперсии (изменчивости) веса.

1. Перейти на новый лист. Скопировать на него столбец с данными о массе кошек.

2. В ячейке B1 ввести заголовок «Масса кошек». В столбец B (ячейки B2–B22) ввести вес 21 кошки группы 2:

3.

|     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 3,7 | 4,2 | 4,7 | 5,3 | 4,9 | 4,8 | 5,1 | 6,4 | 6,2 | 4,7 | 4,1 |
| 5,5 | 3,8 | 4,1 | 4,5 | 5,6 | 5,3 | 4,8 | 4,3 | 3,2 | 3,6 |     |

4. Вычислить  $F$ -критерий, используемый для сравнения дисперсий:

$$F = \frac{\sigma_1^2}{\sigma_2^2},$$

где  $\sigma_1, \sigma_2$  – дисперсии веса в выборках (в числителе указывается большая из дисперсий).

5. Найти квантиль распределения Фишера:  $F_{\alpha; n_1-1; n_2-1}$ , где  $n_1-1$  и  $n_2-1$  – число степеней свободы. Эта величина находится с помощью функции ФРАСПОБР со следующими аргументами: Вероятность – 0,05, Степени\_свободы\_1 – 24, Степени\_свободы\_2 – 20.

Если  $F > F_{\alpha; n_1-1; n_2-1}$ , то различие пород по значениям дисперсии статистически значимо.

6. Оценить значимость различия между дисперсиями другим способом: найти расчетный уровень значимости ( $P$ ) с помощью функции ФРАСП со следующими аргументами:  $X - F$  ( $F$  – критерий, найденный выше), Степени\_свободы\_1 –  $n_1-1$ , Степени\_свободы\_2 –  $n_2-1$ . Если  $P < a$ , то нулевая гипотеза (о незначимости различия) отклоняется.

Ответ:  $F = 1,26$ ,  $F_{\alpha; n_1-1; n_2-1} = 2,0825$ ,  $P = 0,3$ , различие незначимо.

### Пример 5. Проверка соответствия теоретических предположений и фактических данных. Критерий $\chi^2$ . (Ex)

Рассчитано, что если бы фактические данные полностью соответствовали этому предположению, значения исследуемой величины были бы следующими.

При скрещивании некоторых пород животных получено 22 животных, из них 7 – с темной окраской, 9 – со смешанной, 6 – со светлой. Предполагается, что соотношение животных с темной, смешанной и светлой окраской должно составлять 1:2:1. Проверить, соответствуют ли фактические данные этому предположению.

Для оценки значимости различий между фактическими и теоретическими величинами находится следующая статистика:

$$\chi^2 = \sum_{j=1}^m \frac{(X_{Фj} - X_{Тj})^2}{X_{Тj}}$$

Статистика  $\chi^2$  сравнивается с квантилем распределения  $\chi^2: \chi^2_{\alpha, k}$ , где  $\alpha$  – заданный уровень значимости,  $k$  – число степеней свободы. Если  $\chi^2 < \chi^2_{\alpha, k}$ , то можно считать, что различия между фактическими данными и теоретическими величинами статистически незначимы.

Другой способ проверки – по значению статистики  $\chi^2$  и числу степеней свободы  $k$  из таблиц распределения  $\chi^2$  найти расчетный уровень значимости  $P$ . Если  $P > a$ , то различия между фактическими данными и теоретическими величинами статистически незначимы.

Перейти на свободный рабочий лист. В ячейке A1 ввести заголовок «Фактические данные», в ячейке A2 – заголовок «Теоретические величины». В ячейках B1:D1 ввести фактическое количество животных с темной, смешанной и светлой окраской (7, 9 и 6).

В ячейках B2:D2 рассчитать теоретическое количество животных с темной, смешанной и светлой окраской. Если бы соотношение 1:2:1 полностью соответствовало фактическим данным, то количество животных с

темной окраской составляло бы четвертую часть от всех животных, со смешанной окраской – половину, со светлой – четвертую часть. Таким образом, теоретические величины можно найти следующим образом:  $X_{T1}=22/4=5,5$ ;  $X_{T2}=22/2=11$ ;  $X_{T3}=22/4=5,5$ .

Оценить значимость различия между фактическими и теоретическими величинами. Для этого в ячейке F1 ввести заголовок «Расчетный уровень значимости». В ячейке G1 найти расчетный уровень значимости ( $P$ ), используя функцию ХИ2ТЕСТ со следующими параметрами: Фактический интервал – B1:D1, Ожидаемый интервал – B2:D2. По найденному значению  $P$  (ячейка G1) и заданному значению  $\alpha$  (в данном примере  $\alpha = 0,05$ ) сделать вывод о соответствии теоретического предположения фактическим данным.

Для изучения других способов использования критерия  $\chi^2$  в ячейках B3:D3 рассчитать значения  $\frac{(X_{Фj} - X_{Tj})^2}{X_{Tj}}$ . Например, в ячейке B3 для

этого потребуется ввести формулу:  $=(B1-B2)^2/B2$ . В ячейке E3 найти значение статистики  $\chi^2$ , т.е. сумму ячеек B3:D3.

В ячейке E4 найти квантиль  $\chi^2_{\alpha,k}$ , используя функцию ХИ2ОБР со следующими параметрами: Вероятность – 0,05, Степени свободы – 2. Сравнив значения статистики  $\chi^2$  (ячейка E3) и квантиля (ячейка E4), сделать вывод о соответствии теоретического предположения фактическим данным.

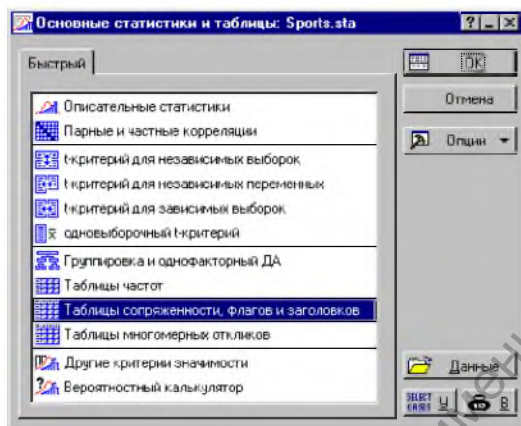
В ячейке E5 найти расчетный уровень значимости ( $P$ ), используя функцию ХИ2РАСП со следующими параметрами:  $X$  – E3 (т.е. ссылка на ячейку со значением статистики  $\chi^2$ ), Степени свободы – 2. По найденному значению  $P$  (ячейка E5) и заданному значению  $\alpha$  (в данном примере  $\alpha = 0,05$ ) сделать вывод о соответствии теоретического предположения фактическим данным.

Ответ:  $\chi^2 = 0,82$ ,  $\chi^2_{\alpha,k} = 5,99$ ,  $P = 0,66$ , фактические данные не противоречат предполагаемому соотношению.

**Пример 5. Параметрические корреляции** (выполнено в пакете Statistica, далее обозначено как Sta). Этот пример основывается на данных из файла *Adstudy.sta*. Этот файл содержит 25 переменных и 50 наблюдений. Эти (фиктивные) данные были собраны путем социологического опроса в рекламном исследовании, где женщины и мужчины оценивали качество двух реклам. Пол респондента кодировался с помощью переменной 1 (Пол – Gender: 1=мужчина – male, 2=женщина – female). Каждому респонденту случайным образом предлагался на рассмотрение один из двух рекламных роликов (Предпочтение – Advert: 1=Coke®, 2=Pepsi®). Затем они оценивали привлекательность соответствующего продукта по 23 различным шкалам (с *Мера1 – Measure1* до *Мера23 – Measure23*). В каждой из этих шкал респонденты могли дать ответы между 0 и 9.

**Запуск модуля Основные статистики и таблицы.** Запустите программу STATISTICA и откройте файл *Adstudy.sta* с помощью команды

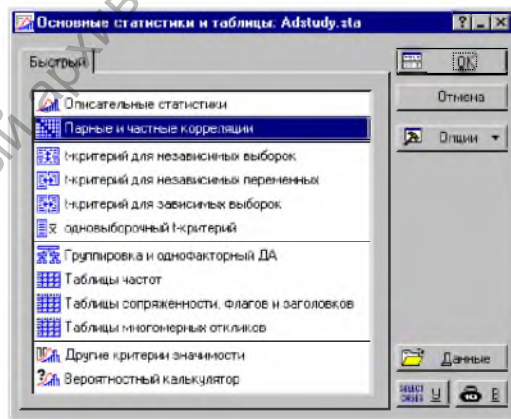
Открыть в меню *Файл*, наиболее вероятно, что файл находится в папке / Examples/Datasets.



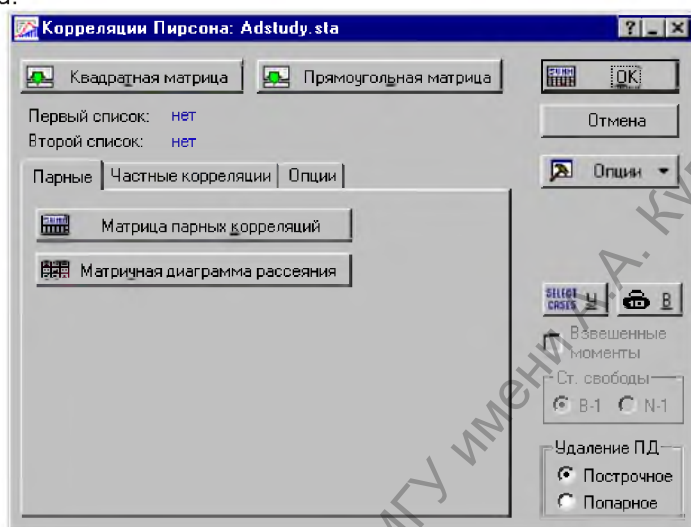
Также вы можете открыть файл данных из стартовой панели любого модуля. Например, выберите команду *Основные статистики и таблицы* в меню *Анализ* для отображения стартовой панели модуля *Основные статистики и таблицы*.

Нажмите кнопку *Данные* для отображения диалогового окна *Выберите Таблицу данных или документ*; в этом диалоге нажмите кнопку *Файлы* и выберите файл *Adstudy.sta*.

**Корреляции.** Сначала проверим, не коррелируют ли между собой оценки в различных шкалах (другими словами, не измеряют ли некоторые шкалы одни и те же свойства). В стартовой панели *Основные статистики и таблицы*, выберите процедуру *Парные и частные корреляции* и щелкните *ОК* (или можете просто дважды щелкнуть на процедуре *Парные и частные корреляции*).



После выбора этой процедуры откроется диалоговое окно *Корреляции Пирсона*.



Вы можете выбрать переменные как из одного списка (т.е. матрица будет квадратной), так и из двух списков (прямоугольная матрица). Для нашего примера, щелкните на кнопке *Квадратная матрица*, и появится стандартное диалоговое окно *Выбор переменных*. Выберите все переменные и затем нажмите кнопку *ОК* для отображения таблицы результатов с корреляциями.

Данные: Корреляции (Adstudy)\*

Корреляции (Adstudy.sta)  
Отмеченные корреляции значимы на уровне  $p < ,05000$   
N=50 (Построчное удаление ПД)

| Переменная | GENDER | ADVERT | MEASURE01 | MEASURE02 | MEASURE03 | MEASURE04 |
|------------|--------|--------|-----------|-----------|-----------|-----------|
| GENDER     | 1,00   | -0,17  | -0,19     | -0,04     | -0,08     | 0,02      |
| ADVERT     | -0,17  | 1,00   | -0,03     | 0,13      | -0,03     | 0,11      |
| MEASURE01  | -0,19  | -0,03  | 1,00      | 0,01      | -0,11     | 0,19      |
| MEASURE02  | -0,04  | 0,13   | 0,01      | 1,00      | -0,06     | 0,01      |
| MEASURE03  | -0,08  | -0,03  | -0,11     | -0,06     | 1,00      | -0,09     |
| MEASURE04  | 0,02   | 0,11   | 0,19      | 0,01      | -0,09     | 1,00      |
| MEASURE05  | 0,26   | -0,28  | 0,04      | 0,08      | -0,21     | 0,10      |
| MEASURE06  | 0,06   | -0,15  | -0,01     | 0,15      | 0,14      | -0,16     |
| MEASURE07  | -0,37  | 0,05   | -0,12     | 0,05      | 0,04      | 0,01      |
| MEASURE08  | -0,04  | -0,02  | -0,02     | -0,08     | -0,19     | 0,01      |
| MEASURE09  | -0,15  | 0,38   | -0,12     | -0,21     | 0,21      | 0,10      |
| MEASURE10  | 0,04   | 0,03   | -0,15     | 0,05      | -0,09     | -0,24     |



**Выделение значимых корреляций (Sta).** Вы можете указать уровень значимости (*альфа*) (.05, по умолчанию) для выделения значимых коэффициентов корреляции в таблице результатов. Чтобы заменить уровень (*альфа*) во вкладке *Опции* диалогового окна *Корреляции Пирсона* измените значение поля *Уровень значим. для выделения*, например, на .001. Снова нажмите кнопку *ОК*. Появится новая таблица результатов с корреляциями, в которой все соответствующие корреляции будут выделены. Теперь стало очень легко обнаружить эти высокие корреляции (например, корреляция между *Мерой5 – Measure5* и *Мерой9 – Measure9* равна *-.47*). Такая высокая корреляция показывает, что две шкалы оценок могут измерять одну и ту же характеристику зрительного восприятия рекламы (хотя одна мера этой характеристики возрастающая, а другая – убывающая).

Две опции из диалогового окна *Корреляции Пирсона* позволяют получить таблицу данных с коэффициентами корреляции, а также более подробными статистиками (например, *p*-значение, число пар *N*, *r*<sup>2</sup> *t*-значения, и т.д.). Когда вы выберете установку *Отображать p-уровень и N*, во вкладке *Опции*, вместе с коэффициентами корреляции будут также выведены *p*-значения и число пар.

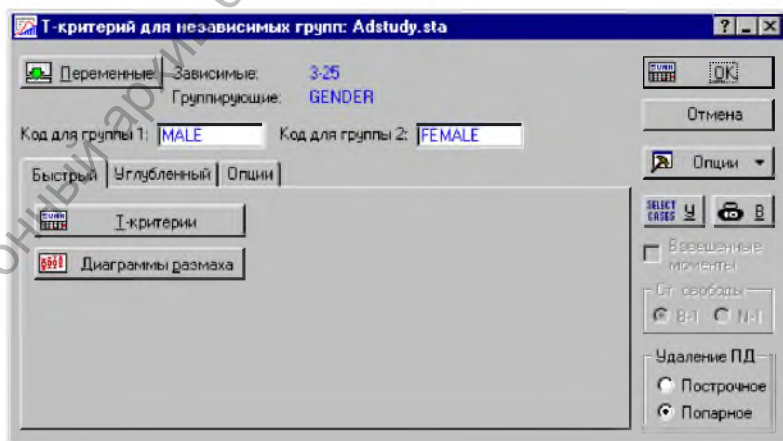
Выбор опции *Подробная таблица результатов* во вкладке *Опции* диалогового окна *Корреляции и частные корреляции* возможен только при выборе 20 или меньше переменных для анализа, так как для каждой корреляции автоматически будет выводиться большое количество информации. После выбора этой опции будет построена таблица результатов, содержащая соответствующие описательные статистики, коэффициенты корреляции, *p*-значения и число пар *N*, а также наклон и отрезок регрессии для каждой переменной.

| Переменная | MEASURE04    | MEASURE05    | MEASURE06    | MEASURE07    | MEASURE08    |
|------------|--------------|--------------|--------------|--------------|--------------|
|            | $r = .506$   | $r = ---$    | $r = .112$   | $r = .721$   | $r = .175$   |
| MEASURE05  | $r = -.1623$ | $r = .2276$  | $r = 1,0000$ | $r = .1199$  | $r = -.3305$ |
| MEASURE07  | $r = .260$   | $r = .112$   | $r = ---$    | $r = .407$   | $r = .019$   |
| MEASURE08  | $r = .0087$  | $r = -.0518$ | $r = .1199$  | $r = 1,0000$ | $r = .0531$  |
| MEASURE09  | $r = .952$   | $r = .721$   | $r = .407$   | $r = ---$    | $r = .714$   |
| MEASURE10  | $r = .0090$  | $r = -.1948$ | $r = -.3305$ | $r = .0531$  | $r = 1,0000$ |
| MEASURE11  | $r = .951$   | $r = .175$   | $r = .019$   | $r = .714$   | $r = ---$    |
|            | $r = .0953$  | $r = -.4672$ | $r = -.2701$ | $r = .0288$  | $r = .0030$  |
|            | $r = .510$   | $r = .001$   | $r = .058$   | $r = .843$   | $r = .984$   |
|            | $r = -.2374$ | $r = -.0663$ | $r = .1295$  | $r = -.0804$ | $r = .2266$  |
|            | $r = .097$   | $r = .647$   | $r = .370$   | $r = .579$   | $r = .114$   |
|            | $r = .0945$  | $r = -.0051$ | $r = -.1497$ | $r = -.0737$ | $r = .0885$  |

Эту опцию следует использовать только для отдельных корреляций (но не для подробного анализа), потому что в этом формате для каждого коэффициента корреляции будут заняты 22 ячейки таблицы результатов; таким образом, для матрицы корреляций 20x20 получится таблица результатов с 8,800 ячейками. Вы можете увидеть выше, что корреляция между *Мерой5 – Measure5* и *Мерой9 – Measure9* действительно, значима ( $p=.0006$ ), это означает, что ошибка, связанная с принятием результата составляет 6 из 10,000. То есть, если делать случайные выборки такого же размера из совокупности, в которой эти две переменные не коррелируют, то только шесть раз из каждых 10,000, получим выборку, в которой корреляция будет  $-0.47$  или меньше.

**Пример 7. Разность между средними (t-критерий) (Sta).** На следующем шаге анализа будет исследована возможность различия между моделью ответов мужчин и женщин. Подробнее, мужчины могут в некоторых шкалах давать более высокие или низкие оценки по сравнению с женщинами. *t*-критерий для независимых выборок может использоваться для идентификации такой возможной разницы. Выборки мужчин и женщин будут сравнены относительно среднего их оценок по каждой шкале. Вернитесь к стартовой панели *Основные статистики и таблицы* и щелкните на процедуре *t-критерий для независимых выборок*, чтобы открыть диалоговое окно *T-критерий для независимых групп*.

Щелкните по кнопке *Переменные*, чтобы открыть стандартное диалоговое окно *Выбор переменных*. Здесь вы можете выбрать и независимые (группирующие), и зависимые переменные. Для нашего примера выберите (выделите) переменную *Пол – Gender* как независимую переменную и переменные от 3 до 25 (содержащие ответы) в качестве зависимых переменных.



Поскольку вы выбрали группирующую переменную, STATISTICA автоматически предложит коды, используемые в этой переменной для идентификации групп, которые будут сравниваться (в нашем случае коды Мужчины – Male и Женщина – Female). Вы можете дважды щелкнуть по каждому из этих полей ввода для вызова окна Значения переменной, где вы можете посмотреть и выбрать коды для каждой группы. Для быстрого выбора значения из списка (и выхода из диалогового окна), щелкните дважды на желаемом значении.

| Данные: Т-критерии: Группир.: GENDER: Gender of the subjects (May 15, 1996). (Adstudy)* |          |          |          |         |          |         |         |           |           |
|---|----------|----------|----------|---------|----------|---------|---------|-----------|-----------|
| Т-критерии, Группир.: GENDER: Gender of the subjects (May 15, 1996). (Adstudy sta)      |          |          |          |         |          |         |         |           |           |
| Группа 1: MALE  |          |          |          |         |          |         |         |           |           |
| Группа 2: FEMALE  |          |          |          |         |          |         |         |           |           |
| Переменная  | Среднее  | Среднее  | t-знач.  | ст. св. | p        | N набл. | N набл. | Ст. откл. | Ст. откл. |
|   | MALE     | FEMALE   |          |         |          |         |         |           |           |
| MEASURE01   | 6,285714 | 5,409091 | 1,30945  | 48      | 0,196615 | 28      | 22      | 2,088011  | 2,648613  |
| MEASURE02   | 4,642857 | 4,409091 | 0,28152  | 48      | 0,779520 | 28      | 22      | 2,971647  | 2,839502  |
| MEASURE03   | 4,321429 | 3,909091 | 0,52707  | 48      | 0,600572 | 28      | 22      | 2,931989  | 2,486326  |
| MEASURE04   | 5,464286 | 5,590909 | -0,16547 | 48      | 0,869267 | 28      | 22      | 2,987407  | 2,239453  |
| MEASURE05   | 3,357143 | 4,727273 | -1,87198 | 48      | 0,067309 | 28      | 22      | 2,831232  | 2,186143  |
| MEASURE06   | 4,714286 | 5,000000 | -0,32910 | 48      | 0,743511 | 28      | 22      | 3,125251  | 2,943920  |
| MEASURE07   | 5,464286 | 3,636364 | 2,73550  | 48      | 0,008703 | 28      | 22      | 1,835497  | 2,870962  |
| MEASURE08   | 3,821429 | 3,590909 | 0,28554  | 48      | 0,776461 | 28      | 22      | 2,708745  | 2,986622  |
| MEASURE09   | 4,571429 | 3,636364 | 1,07920  | 48      | 0,285892 | 28      | 22      | 3,155578  | 2,887501  |
| MEASURE10   | 3,821429 | 4,090909 | -0,30691 | 48      | 0,760242 | 28      | 22      | 3,019119  | 3,160908  |
| MEASURE11   | 5,642857 | 4,272727 | 1,67689  | 48      | 0,100067 | 28      | 22      | 2,895901  | 2,831487  |
| MEASURE12   | 4,607143 | 4,000000 | 0,72504  | 48      | 0,471948 | 28      | 22      | 2,935596  | 2,943920  |
| MEASURE13   | 4,928571 | 3,818182 | 1,31895  | 48      | 0,193445 | 28      | 22      | 2,930410  | 2,986260  |
| MEASURE14   | 3,821429 | 4,227273 | -0,46423 | 48      | 0,644584 | 28      | 22      | 2,982089  | 3,176278  |
| MEASURE15   | 3,750000 | 3,883333 | -0,14331 | 48      | 0,886648 | 28      | 22      | 2,875503  | 2,980130  |

**Просмотр результатов t-критерия (Sta).** Самым быстрым способом изучения таблицы является просмотр пятого столбца (содержащего *p*-уровни) и определение того, какие из *p*-значений меньше установленного уровня значимости .05. Для большинства зависимых переменных средние по двум группам (Мужчины – Males и Женщины – Females) очень близки. Единственная переменная, для которой *t*-критерий соответствует установленному уровню значимости .05 – это Measure7, для нее *p*-уровень равен .0087. Как показывают столбцы, содержащие средние значения (см. две первые колонки) для мужчин, эта переменная принимает в среднем существенно большие значения – в выбранной шкале измерений для мужчин она равна 5.46, а для женщин – 3.63.

## 4. НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА

Этот вид статистического анализа используется в том случае, когда распределение ваших данных не имеют нормального распределения. Другим фактором является объем или размер выборки, доступной для анализа. До тех пор пока выборка достаточно большая (например, 100 или больше наблюдений), можно считать, что распределение часто бывает нормальным, но, если выборка очень мала, то необходимо его оценить, и при необходимости использовать непараметрические методы.

Обычно, когда имеются две выборки (например, мужчины и женщины), которые вы хотите сравнить относительно среднего значения некоторой изучаемой переменной, вы используете *t*-критерий для независимых выборок. Непараметрическими альтернативами этому критерию являются: *Критерий серий Вальда-Вольфовица*, *U-критерий Манна-Уитни* и *Двухвыборочный критерий Колмогорова-Смирнова*.

Каждая непараметрическая процедура в модуле имеет свои достоинства и свои недостатки. Например, двухвыборочный критерий Колмогорова-Смирнова чувствителен не только к различию в положении двух распределений, например, к различиям средних, но также чувствителен и к форме распределения. Критерий Вилкоксона парных сравнений предполагает, что можно ранжировать различия между сравниваемыми наблюдениями. Если это не так, лучше использовать Критерий знаков. В общем если результат исследования является важным (например, оказывает ли людям помощь определенная очень дорогостоящая и болезненная терапия?), то всегда целесообразно применить различные непараметрические тесты. Возможно, результаты проверки (разными тестами) будут различны.

С другой стороны, непараметрические тесты имеют меньшую статистическую мощность (менее чувствительны), чем их параметрические конкуренты, и если важно обнаружить даже слабые отклонения (например, является ли данная пищевая добавка опасной для людей?) следует особенно внимательно выбирать статистику критерия.

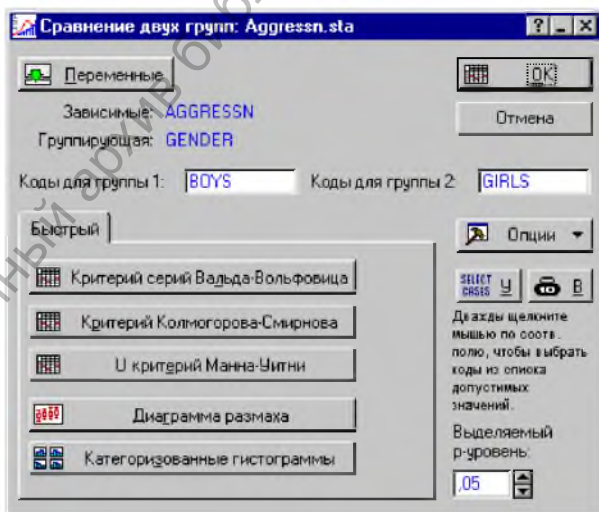
**Пример 7. Критерий серий Вальда-Вольфовица, Манна-Уитни U-критерий, двухвыборочный критерий Колмогорова-Смирнова (Sta).**

Пример основан на исследовании агрессивности четырехлетних мальчиков и девочек (Siegel, 1956, стр. 138). Данные содержатся в файле *Aggressn.sta*. Чтобы открыть этот файл выберите в меню *Файл* команду *Открыть*; наиболее вероятно, что этот файл данных находится в директории */Examples/Datasets*.

Двенадцать мальчиков и двенадцать девочек наблюдались в течение 15-минутной игры; агрессивность каждого ребенка оценивалась в баллах (в терминах частоты и степени проявления агрессивности) и суммировалась в один индекс агрессивности, который вычислялся для каждого ребенка.

| Aggression score of 12 |          |
|------------------------|----------|
| 1                      | 2        |
| GENDER                 | AGGRESSN |
| 1 BOYS                 | 86       |
| 2 BOYS                 | 69       |
| 3 BOYS                 | 72       |
| 4 BOYS                 | 65       |
| 5 BOYS                 | 113      |
| 6 BOYS                 | 65       |
| 7 BOYS                 | 118      |
| 8 BOYS                 | 45       |
| 9 BOYS                 | 141      |
| 10 BOYS                | 104      |
| 11 BOYS                | 41       |
| 12 BOYS                | 50       |
| 13 GIRLS               | 55       |
| 14 GIRLS               | 40       |

Задание анализа. Выберите команду Непараметрическая статистика в меню Анализ для отображения стартовой панели модуля Непараметрическая статистика. Далее выберите Сравнение двух независимых групп и нажмите кнопку ОК для отображения диалогового окна Сравнение двух групп. Нажмите кнопку Переменные для отображения диалогового окна Выбор переменных. Выберите переменную Aggressn как зависимую, а переменную Gender как независимую. Коды для однозначного отнесения каждого наблюдения к определенному полу будут автоматически выбраны программой.



**Просмотр результатов.** Теперь нажмите кнопку *Критерий серий Вальда-Вольфовица*, чтобы выполнить анализ.

Данные: Критерий серий Вальда-Вольфовица (Aggressn)\*

Критерий серий Вальда-Вольфовица (Aggressn.sta)  
По перем. GENDER  
Отмеченные критерии значимы на уровне  $p < ,05000$

| Перем.   | N набл. BOYS | N набл. GIRLS | Среднее BOYS | Среднее GIRLS | Z        | p-уров.  | Z скорр. | p-уров.  | Число серий | Число совп. |
|----------|--------------|---------------|--------------|---------------|----------|----------|----------|----------|-------------|-------------|
| AGGRESSN | 12           | 12            | 80,75000     | 26,66667      | -3,75681 | 0,000172 | 3,548100 | 0,000388 | 4           | 0           |

Как видно из таблицы результатов, разница агрессивности мальчиков и девочек в этом исследовании высоко значима, независимо от того, какой критерий используется. Вернитесь во вкладку *Быстрый* диалогового окна *Сравнение двух групп* и нажмите кнопки *Двухвыборочный критерий Колмогорова-Смирнова* и *U критерий Манна-Уитни*.

Данные: Критерий Колмогорова-Смирнова (Aggressn)\*

Критерий Колмогорова-Смирнова (Aggressn.sta)  
По перем. GENDER  
Отмеченные критерии значимы на уровне  $p < ,05000$

| Перем.   | Макс. отр Разн. | Макс. по Разн. | p-уров.    | Среднее BOYS | Среднее GIRLS | Ст. откл BOYS | Ст. откл GIRLS | N набл. BOYS | N набл. GIRLS |
|----------|-----------------|----------------|------------|--------------|---------------|---------------|----------------|--------------|---------------|
| AGGRESSN | 0,00            | 0,833333       | $p < ,001$ | 80,75000     | 26,66667      | 31,82373      | 16,99911       | 12           | 12            |

### Пример 8. Непараметрическая корреляция Спирмена (Sta)

Используется, если представлена в нечисловой форме.

Определяется значение вариант для каждой из величин присваиваются ранги мин – 1 ранг. Если варианты равны то одинаковый усредненный ранг.

Вычисляются разности рангов.

Вычисляется ранговый коэф Спирмена.

Для проверки ранговой корреляции вычисляется *T*-критерий Стьюдента сравнивается с квантилем.

Используя коэффициент ранговой корреляции Спирмена, выяснить, имеется ли статистически значимая линейная связь между анализируемыми величинами.

1. Подготовить таблицу исходных данных из 11 строк (по объему выборок).

2. Присвоить переменным имена *T* и *ZABOL*.

3. В столбце переменной *T* ввести значения температуры, в столбце *ZABOL* – значения заболеваемости.

4. Из меню *Statistics* выбрать команду *Nonparametrics*, затем – команду *Correlations* (в скобках после этой команды указаны вычисляемые величины, в том числе *Spearman* – коэффициент ранговой корреляции Спирмена).

5. В появившемся окне *Nonparametric correlations* нажать кнопку *Variables*. Указать переменные для анализа: в одном из списков выбрать переменную *T*, в другом – *ZABOL*. Нажать *OK*.

6. В окне *Nonparametric correlations* для получения результатов нажать кнопку *Spearman R* (или *Spearman rank R*).

Примечание. В окне *Nonparametric correlations* имеются также кнопки для вычисления других коэффициентов, отражающих связь между величинами: коэффициента “гамма” (*Gamma*) и коэффициента корреляции Кендалла (*Kendall Tau*).

На экран выводится окно результатов. Коэффициент ранговой корреляции Спирмена обозначается как *Spearman R*. Указывается также расчетный уровень значимости (*p-level*): если он меньше заданного уровня значимости (обычно – 0,05), то связь между исследуемыми величинами статистически значима. Если при этом коэффициент ранговой корреляции положителен, то между величинами имеется прямая связь (например, чем выше температура летом, тем больше заболеваемость). Если коэффициент ранговой корреляции отрицателен, то связь между величинами обратная.

Другие величины, выводимые в окне результатов: *Valid N* – объем выборки; *T(N-2)* – статистика, используемая для определения расчетного уровня значимости.

### **Пример 9. Однофакторный дисперсионный анализ (Ex)**

Исследуется влияние кормовой добавки на привесы животных трех пород (П1, П2, П3). Наблюдалось пять животных породы П1, шесть – породы П2, пять – породы П3. Получены следующие величины привеса (в граммах):

- порода П1: 640, 637, 648, 624, 639;
- порода П2: 638, 625, 662, 645, 640, 637;
- порода П3: 625, 617, 631, 629, 619.

Используя метод дисперсионного анализа, выяснить, различается ли действие кормовой добавки на разные породы животных.

1. Перейти на свободный рабочий лист. В ячейках А1, В1, С1 ввести заголовки “П1”, “П2”, “П3”. В ячейках А2:А6, В2:В7 и С2:С6 ввести значения привесов для пород П1, П2, П3.

2. Из меню *Сервис – Анализ данных* выбрать инструмент *Однофакторный дисперсионный анализ*. Указать *Входной интервал* – А1:С7. Установить переключатель *Группирование* – по столбцам. Установить флажок *Метки*. Указать заданный уровень значимости *Альфа* – 0,05. В области *Параметры вывода* выбрать переключатель *Выходной интервал* и указать ячейку, с которой будет начинаться вывод результатов (например, ячейку А10). Для получения результатов нажать кнопку *OK*. Смысл полученных результатов следующий:

- Счет – объемы выборок;
- Сумма – суммы по выборкам;

Среднее – средние значения по выборкам;  
Дисперсия – дисперсии по выборкам;  
SS – величины  $Q_1$  и  $Q_2$ , рассчитанные по формулам:

$$Q_1 = \sum_{k=1}^m n_k \cdot (\bar{X}_k - \bar{X})^2, \quad Q_2 = \sum_{k=1}^m \sum_{i=1}^{n_k} n_k \cdot (x_{jk} - \bar{X}_k)^2,$$

где  $df$  – числа степеней свободы  $m-1$  и  $n-m$ ;

MS – величины  $Q_1/(m-1)$  и  $Q_2/(n-m)$ ;

$F$  – статистика, рассчитанная по формуле:

$$F = \frac{Q_1/(m-1)}{Q_2/(n-m)},$$

$P$ -значение – расчетный уровень значимости, найденный из таблиц распределения Фишера по значению статистики  $F$  и числам степеней свободы  $m-1$  и  $n-m$ ;

$F$  критическое – квантиль распределения Фишера  $F_{\alpha; m-1; n-m}$ .

Если величина  $F$  больше, чем  $F$  критическое, и  $P$ -значение меньше заданного уровня значимости  $\alpha$ , то можно считать, что различие между средними значениями выборок значимо. Это означает, что кормовая добавка по-разному влияет на вес животных разных пород.

Используя метод дисперсионного анализа, выяснить, различается ли действие кормовой добавки на разные породы животных.

1. Присвоить переменным имена *PORODA* и *PRIVES*. Если это требуется, с помощью команды *Insert – Add cases* добавить необходимое количество строк (всего потребуется 16 строк).

2. В столбце переменной *PORODA* ввести признаки породы животных, например,  $p_1$ ,  $p_2$  и  $p_3$ . В столбце *PRIVES* ввести значения привеса. Фрагмент исходных данных показан на рис. 4.

3. Из меню *Statistics* выбрать команду *ANOVA*, затем – команду *One-way ANOVA* (однофакторный дисперсионный анализ).

4. Нажать кнопку *Variables*. В списке *Categorical Predictor (factor)* выбрать переменную *PORODA*, в списке *Dependent variable list* – *PRIVES*. Нажать *OK*. Происходит возврат в окно *General ANOVA/MANOVA*. В нем также нажать *OK*.

5. Для получения результатов нажать *All effects*. Оценка значимости различия между величинами привеса для разных пород выполняется по величине  $p$  (расчетный уровень значимости) в строке переменной *PORODA*.

### Пример 10. Многофакторный дисперсионный анализ (Ex)

Исследуется влияние способа обработки почвы и вида используемых удобрений на урожай некоторого растения. Анализируются два способа обработки почвы (глубокая и мелкая) и три вида удобрений (А, В, С). Исследования выполнялись на 24 участках (по 4 участка для каждой ком-



бинации способа обработки почвы и вида удобрений). На участках получен следующий урожай (в центнерах):

- глубокая обработка, удобрение А: 18, 24, 9, 15;
- глубокая обработка, удобрение В: 25, 17, 14, 11;
- глубокая обработка, удобрение С: 30, 26, 37, 34;
- мелкая обработка, удобрение А: 14, 17, 25, 22;
- мелкая обработка, удобрение В: 18, 14, 19, 15;
- мелкая обработка, удобрение С: 24, 22, 17, 19.

Используя метод дисперсионного анализа, выяснить, какие факторы оказывают значимое влияние на урожай.

1. Перейти на свободный рабочий лист. Ввести исходные данные, как показано на рис.

|   | А        | В           | С           | Д           |
|---|----------|-------------|-------------|-------------|
| 1 |          | Удобрение А | Удобрение В | Удобрение С |
| 2 | Глубокая | 18          | 25          | 30          |
| 3 |          | 24          | 17          | 26          |
| 4 |          | 9           | 14          | 37          |
| 5 |          | 15          | 11          | 34          |
| 6 | Мелкая   | 14          | 18          | 24          |
| 7 |          | 17          | 14          | 22          |
| 8 |          | 25          | 19          | 17          |
| 9 |          | 22          | 15          | 19          |

2. Из меню *Сервис – Анализ данных* выбрать инструмент *Двухфакторный дисперсионный анализ с повторениями*. Указать следующие параметры: *Входной интервал – A1:D9*, *Число строк для выборки – 4*, *Альфа – 0,05*. В области *Параметры вывода* выбрать переключатель *Выходной интервал* и указать ячейку, с которой будет начинаться вывод результатов (например, ячейку A12). Для получения результатов нажать кнопку *ОК*.

Значимость факторов определяется *P*-значение (или в столбцах *F* и *F* критическое). Строка *Выборка* соответствует фактору, указанному в исходных данных в строках (в данном примере это способ обработки почвы), строка *Столбцы* – фактору, указанному в столбцах (в данном примере – вид удобрений), строка *Взаимодействие* – комбинации факторов. Если *P*-значение меньше заданного уровня значимости  $\alpha$ , то соответствующий фактор оказывает значимое влияние на исследуемую величину (урожай).

В данном примере значимое влияние на урожай оказывает вид удобрений, а также совместное действие видов удобрений и способа обработки почвы.

Используя метод дисперсионного анализа, выяснить, какие факторы оказывают значимое влияние на урожай (Sta).

1. Присвоить переменным имена *OBRAB*, *UD* и *UROZAI*. Если это требуется, с помощью команды *Insert – Add cases* добавить необходимое количество строк (всего потребуется 24 строки).

2. В столбце переменной *OBRAB* ввести признаки способа обработки, например, *g* (для глубокой обработки) и *m* (для мелкой). В столбце переменной *UD* ввести обозначения удобрений, например, *A*, *B*, *C*. В столбце *UROZAI* ввести значения урожая.

3. Из меню *Statistics* выбрать команду *ANOVA*, затем – команду *Factorial ANOVA* (многофакторный дисперсионный анализ).

4. Нажать кнопку *Variables*. В списке *Categorical Predictor (factor)* выбрать переменные *OBRAB* и *UD*, в списке *Dependent variable list* – *UROZAI*. Нажать *OK*. Происходит возврат в окно *General ANOVA/MANOVA*. В нем также нажать *OK*.

5. Для получения результатов нажать кнопку *All effects*. Оценка значимости различия между величинами привеса для разных пород выполняется по величине  $p$  (расчетный уровень значимости). Вид результатов показан на рис. 5.

В данном примере значимое влияние на урожай оказывает вид удобрений, а также совместное действие видов удобрений и способа обработки почвы (в строках *UD* и *OBRAB\*UD* расчетный уровень значимости  $p < 0,05$ ). Влияние способа обработки незначимо.

Найти средние значения по каждой выборке (для определения комбинаций факторов, которым соответствует максимальный урожай). Для этого из меню *Statistics* выбрать команду *Basic Statistics and Tables*, затем – команду *Breakdown & One-way ANOVA*. На вкладке *Individual Tables* нажать кнопку *Variables*; в окне *Grouping variables* выбрать переменные *OBRAB* и *UD*, а в окне *Dependent variables* – переменную *UROZAI*. Нажать *OK*. Для получения результатов в окне *Statistics by Groups* нажать кнопку *OK*, затем – кнопку *Summary*.

## СОДЕРЖАНИЕ

|  |    |
|--|----|
| ВВЕДЕНИЕ.....  | 1  |
| 1. ОБЩИЕ СВЕДЕНИЯ О СТАТИСТИКЕ.....                          | 2  |
| 1.1. Определение цели исследования.....                      | 2  |
| 1.2. Планирование исследования.....                          | 2  |
| 2. ПОДГОТОВКА ДАННЫХ К СТАТИСТИЧЕСКОМУ АНАЛИЗУ.....          | 6  |
| 2.1. Основные принципы и методы статистического анализа..... | 7  |
| 3. ПРОВЕДЕНИЕ СТАТИСТИЧЕСКОГО АНАЛИЗА.....                   | 9  |
| 4. НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА.....                         | 18 |