Министерство здравоохранения Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Уральский государственный медицинский университет»

В.Я. Крохалев, С.А. Скопинов, В.А. Телешев

СТАТИСТИКА

Учебное пособие

Екатеринбург
Издательство УГМУ
2018

УДК 31:61(075.8)
ББК 51.1,02
С78

Печатается по решению Ученого совета медико-профилактического факультета ФГБОУ ВО УГМУ Минздрава России (протокол №1 от 27 сентября 2017 года)

Ответственный редактор
д-р биол. наук, проф. Ф.А. Бляхман

Рецензенты:
д-р физ.-мат. наук, проф. Шашкин С.А.,
д-р мед. наук, проф. Н.В. Ножкина

Крохалев, В.Я.
С78 ***Статистика*** [Текст] : уч. пособие / В.Я. Крохалев, С.А. Скопинов, В.А. Телешев; ФГБОУ ВО УГМУ Минздрава России. — Екатеринбург : Изд-во УГМУ, 2018. — 114 с.

ISBN 978-5-89895-860-2

Учебное пособие включает основные положения теории вероятностей и статистики и практические работы на основе современных компьютерных технологий применительно к решению задач практической медицины. Практические навыки статистических расчетов студенты получают на основе освоения и использования на конкретных примерах специализированного компьютерного пакета STATISTICA.

Пособие рассчитано на студентов медицинских вузов, изучающих курс статистики.

УДК 31:61(075.8)
ББК 51.1,02

ISBN 978-5-89895-860-2

ОГЛАВЛЕНИЕ

ОСНОВНЫЕ ПОЛОЖЕНИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Математическая статистика как наука начинается с работ знаменитого немецкого математика Карла Фридриха Гаусса, который на основе теории вероятностей исследовал и обосновал метод наименьших квадратов, созданный им в 1795 г. и примененный для обработки астрономических данных (с целью уточнения орбиты малой планеты Церера). Его именем часто называют одно из наиболее популярных распределений вероятностей — нормальное, а в теории случайных процессов основной объект изучения — гауссовы процессы.

В конце XIX — начале XX в. крупный вклад в математическую статистику внесли английские исследователи, прежде всего К. Пирсон и Р. Фишер. В частности, Пирсон разработал критерий хи-квадрат (χ^2) проверки статистических гипотез, а Фишер — дисперсионный анализ, теорию планирования эксперимента, метод максимального правдоподобия оценки параметров. В 30-е годы XX в. Д. Нейман и К. Пирсон развили общую теорию проверки статистических гипотез, а советские математики А.Н. Колмогоров и Н.В. Смирнов заложили основы непараметрической статистики. В сороковые годы XX в. А. Вальд построил теорию последовательного статистического анализа.

Математическая статистика — это раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических целей.

Основные понятия математической статистики

Генеральная совокупность — совокупность всех возможных результатов наблюдений некоторого признака всех изучаемых объектов (иногда — сами объекты).

Например, генеральной совокупностью может быть: рост всех людей, частота употребления определенной части речи во всех произведениях изучаемого автора, средний балл аттестата всех выпускников и т.п.

Выборка (выборочная совокупность) — совокупность результатов наблюдений, выбранных случайно из генеральной совокупности.

Выборкой может быть: рост 20 случайно выбранных студентов, количество глаголов в выбранных произвольно 50 однородных отрывках текста длиной 500 словоупотреблений, средний балл аттестата 100 выпускников, выбранных случайно из школ города и т.п.

Выборка называется **репрезентативной**, если она правильно отражает свойство генеральной совокупности. Репрезентативность выборки достигается случайностью отбора, когда все объекты генеральной совокупности имеют одинаковую вероятность быть отобранными.

Математическое ожидание — ожидаемое среднее значение случайной величины.

Относительная частота — вероятность регистрации определенного значения случайной величины в заданном интервале ее возможных значений.

Дисперсия случайной величины — математическое ожидание квадрата отклонения случайной величины от ее среднего значения. Она характеризует степень разброса значений случайной величины относительно ее математического ожидания, т.е. ширину диапазона значений.

Дисперсия случайной величины имеет размерность квадрата случайной величины. Для наглядности характеристики рассеивания пользуются величиной стандартного (среднеквадратического) отклонения, размерность которого совпадает с размерностью самой случайной величины.

Стандартное (среднеквадратическое) отклонение — величина, равная квадратному корню из дисперсии.

Статистическая гипотеза — предположение о законе распределения изучаемых случайных величин или событий. Это понятие встречается в задаче анализа данных при статистической проверке гипотез. В теории статистической проверки гипотез рассматривается, как экспериментальные данные могут быть использованы
для выбора одной из альтернативных гипотез либо для того, чтобы подтвердить или опровергнуть теорию (гипотезу). Решение принимается с помощью специальных статистических критериев (критерии Стьюдента, Фишера, Пирсона и др.).

ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В МЕДИЦИНЕ

Один из основателей теории статистики бельгийский статистик Адольф Кетле (1796—1874) приводит пример использования статистических наблюдений в медицине: «Два профессора сделали любопытное наблюдение относительно скорости пульса. Сравнив мои наблюдения с их данными, они заметили, что между ростом и числом пульса существует зависимость. Возраст может влиять на пульс только при изменении роста, который играет в этом случае роль регулирующего элемента. Число ударов пульса находится, таким образом, в обратном отношении с квадратным корнем роста. Приняв за рост среднего человека 1,684 м, они полагают число ударов пульса равным 70. Имея эти данные, можно вычислить число ударов пульса у человека какого бы то ни было роста».

Активным сторонником использования статистики был основоположник военно-полевой хирургии Н.И. Пирогов. Еще в 1849 году, говоря об успехах отечественной хирургии, он указывал: «Приложение статистики для определения диагностической важности симптомов и достоинства операций можно рассматривать как важное приобретение новейшей хирургии».

В 60-е годы XX века, после очевидных успехов прикладной статистики в технике и точных науках, вновь начал расти интерес к использованию статистики в медицине. В.В. Алпатов в статье «О роли математики в медицине» писал: «Чрезвычайно важна математическая оценка терапевтических воздействий на человека. Новые лечебные мероприятия имеют право заменить собою мероприятия, уже вошедшие в практику, лишь после обоснованных статистических испытаний сравнительного характера». Огромное применение может получить статистическая теория в постановке клинических и неклинических испытаний новых терапевтических и хирургических мероприятий.

Основными задачами медицинской статистики являются разработка специальных методов исследования массовых процессов и явлений в медицине и здравоохранении; выявление наиболее существенных закономерностей и тенденций в здоровье населения в целом и в различных его группах (возрастных, половых, профессиональных и др.) во взаимосвязи с конкретными условиями и образом жизни: изучение и оценка состояния и динамики развития сети деятельности учреждений здравоохранения и медицинских кадров.

При работе врача, и в том числе санитарного врача, часто возникает необходимость делать выводы на основе обработки данных наблюдений. При этом статистика базируется на теории вероятностей и предназначена для того, чтобы делать разумные заключения в тех случаях, когда число наблюдений не одно и не два, а хотя бы более чем несколько. Например, три пациента принимали один и тот же лекарственный препарат в течение одного и того же времени, и при этом одному стало лучше, второму — хуже, у третьего ничего не изменилось, — никакого вывода сделать нельзя. Но если было пролечено 15 пациентов и 10-ти стало лучше, трем — хуже, у двоих состояние без изменений, то это уже задача, вывод по которой можно сделать, опираясь на методы статистики.

ОСНОВНЫЕ ПОЛОЖЕНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Теория вероятностей как наука возникла в середине XVII века. В ее создании принимали участие выдающиеся ученые своего времени Б. Паскаль, П. Ферма, Х. Гюйгенс и Я. Бернулли. Интерес к математическому описанию случайных событий появился в связи с подсчетом вероятностей различных исходов в азартных играх. В XVIII в. и начале XIX в. теория вероятностей находит применение в физических и технических науках — изучении теории ошибок наблюдений в астрономии и геодезии, теории стрельбы и др. С этим периодом связано участие в развитии теории вероятностей таких великих математиков как С. Лаплас, К. Гаусс и С. Пуассон. Во второй половине XIX в. дальнейшее развитие теории вероятностей связано в основном с российской математической школой в лице П.Л. Чебышева, А.М. Ляпунова и А.А. Маркова.

Современная логическая схема построения основ теории вероятностей разработана в 1933 советским математиком А. Н. Колмогоровым.

Теория вероятностей базируется на следующих основных понятиях:

Случайное событие — это явление, которое при одних и тех же условиях может или произойти, или не произойти.

Испытание — это возникновение этих неопределенных условий и наблюдение за результатом. Любое испытание приводит к результату или исходу, который заранее невозможно в точности предсказать.

Случайная величина — величина, которая в результате опыта принимает какое-либо одно из множества ее возможных значений, при этом до опыта невозможно сказать, какое именно.

Случайные величины подразделяются на два класса:

— **дискретные**, принимающие некоторые фиксированные значения (например, пульс, принимающий значения только из области целых чисел);

— **непрерывные**, принимающие любые возможные значения (например, температура тела, принимающая любые значения, ограниченные только точностью используемого термометра).

Примеры случайных величин:

— количество очков, выпавших на игральной кости (X), — дискретная случайная величина;

— оценка, полученная студентом на экзамене (Y), — дискретная случайная величина;

— рост наугад выбранного студента (Z) — непрерывная случайная величина;

— время действия лекарства (T) — непрерывная случайная величина.

Случайные события могут быть:

1. **Достоверными** или **невозможными**.

Достоверным называется событие, которое в данных условиях всегда происходит, невозможным — если оно никогда не может быть результатом данного испытания. Например, при бросании монеты событие А — «выпадение какой-либо стороны монеты» — будет достоверным, а В — «одновременное выпадение “решки” и “орла”» — невозможным.

2. **Зависимыми** или **независимыми**.

Если появление одного события влечет за собой появление другого, то говорят, что второе событие зависит от первого.

3. **Равновероятными** или **не равновероятными**.

Например, в случае бросания игральной кости события выпадения каждой цифры равновероятны.

Определение вероятности

Вероятность — это отношение числа благоприятных исходов событий *m* к общему числу всех равновозможных исходов *n*.

Обычно вероятность обозначают буквой ***P*** (от англ. *probability* — вероятность). Вероятность понимается как количественная мера объективной возможности появления случайного события *A*.

Вероятность обладает следующими свойствами:

1) $0<P<1$, так как количество благоприятных исходов не может быть больше их общего числа;

2) вероятность достоверного события = 1;

3) вероятность невозможного события = 0.

ПРАКТИЧЕСКИЕ ЗАНЯТИЯ

Занятие № 1. Основы теории вероятностей

Решение простейших задач на нахождение вероятностей. Нахождение вероятностей независимых, несовместных событий и использование формул сложения и умножения вероятностей.

— **Событие** — возникновение комплекса условий (G), результатом которого является тот или иной исход.

— Обозначение события — *A, B, C…*

— Какие бывают события:

— **достоверное событие** (Ω) — событие, которое всегда происходит при выполнении данного комплекса условий *G;*

— **невозможное событие** ($\varnothing$) — событие, которое не может произойти при выполнении комплекса условий *G* (никогда!);

— **Случайное событие** — такое, которое при реализации комплекса условий *G* может произойти или не произойти.

Пример случайного события:

— выпадение орла при подбрасывании «правильной» монетки.

Вероятность — это числовая характеристика возможности наступления события.

Разные определения (подходы и способы нахождения вероятностей):

— статистическое;

— классическое;

— геометрическое.

Рассмотрим способы нахождения вероятностей более подробно.

Статистический подход (опыт)

При проведении N испытаний событие A появляется m раз — за вероятность события *A* принимают $P(A) = {}^{m}\!/_{N}$.

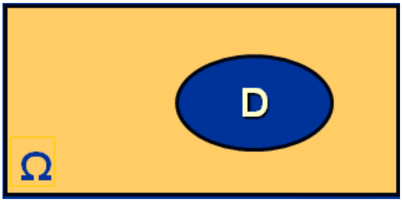
Пример. Вероятность рождения мальчика, найденная в одной из областей страны, равна 0,515.

Классическое определение (априори)

Исходы равно возможны. Пусть в этом случае P — вероятность события A, m — число событий, благоприятствующих наступлению A, n — общее число событий; тогда искомую вероятность можно найти как $P(A) = {}^{m}\!/_{n}$.

Геометрический подход

Исходы опытов равно возможны. Рассмотрим прямоугольник Ω, на нем область D. Пусть в Ω случайным образом вбрасывается точка X, тогда, если событие A — попадание точки в D, то вероятность можно определить как: $P(A) = {}^{S_D}/_{S_\Omega}$.



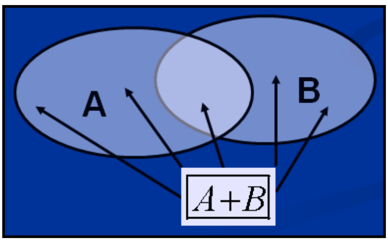
При этом предполагается, что точка всегда попадает в область Ω. Иными словами, смысл этого подхода состоит в том, что вероятность попадания в мишень пропорциональна ее площади.

Действия над событиями. Диаграммы Эйлера.

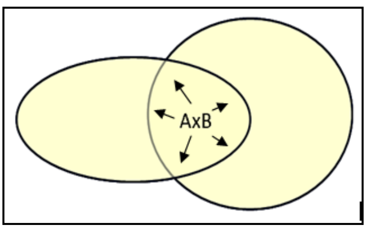
<u>На диаграммах:</u>

- ***достоверное событие (Ω) прямоугольник;***
- ***элементарное событие — точка;***
- ***случайное событие — область (D) внутри прямоугольника.***

1. Сумма событий: $(A + B)$ — событие, соответствующее наступлению A либо B, либо A и B вместе.



2. Произведение событий: (AB) — событие, соответствующее наступлению A и B одновременно.



Правило сложения вероятностей

1. Для ***несовместных*** (области событий не пересекаются $A \cdot B = 0$) событий: $P(A + B) = P(A) + P(B)$.

2. В остальных случаях: $P(A + B) = P(A) + P(B) - P(A \cdot B)$.

Пример. В группе 27 студентов. Из них 17 изучали (успешно!) английский, 6 — французский, 2 — оба языка. Найти вероятность того, что случайно выбранный студент изучал хотя бы один язык.

Решение:

$$P(A) = {}^{17}/_{27}, P(B) = {}^{6}/_{27}, P(A \cdot B) = {}^{2}/_{27}.$$

$$P(A + B) = P(A) + P(B) - P(A \cdot B) = {}^{7}/_{9}.$$

Событие, противоположное событию A (обозначаемое как $\bar{A}$), определяется соотношением $A + \bar{A} = \Omega$.

При этом указанные события предполагаются несовместными, для таких событий отсутствует общая область (пересечение) и выполняется условие $A \cdot \bar{A} = 0$.

Тогда $P(\Omega) = 1 = P(\bar{A} + A) = P(\bar{A}) + P(A)$. Отсюда вытекает, что $P(\bar{A}) = 1 - P(A)$.

Условная вероятность

Иногда наступление одного события изменяет вероятность наступления других событий.

Если произошло событие B, то новая вероятность события *A* называется ***условной вероятностью*** и обозначается как $P_B(A)$. При этом *B* оказывается достоверным событием и играет роль пространства элементарных событий (Ω). Условная вероятность определяется формулой

$P_B(A) = \frac{P(A \cdot B)}{P(B)}$.

Из определения условной вероятности следует:

$$P(AB) = P(B) \cdot P_B(A) = P(A) * P_A(B).$$

Если *A* и *B* не зависят друг от друга, то:

$$P_A(B) = P(B) \text{ или } P_B(A) = P(A).$$

Для независимых событий выполняется «правило умножения вероятностей»:

$$P(AB) = P(A) \cdot P(B).$$

Задачи

1. В ящике 10 шаров: 5 красных, 2 синих и 3 белых. Найти вероятность того, что при единственном опыте из корзины достанут цветной шар.
2. В ящике 10 шаров: 5 красных, 2 синих и 3 белых. Найти вероятность того, что два случайно взятые шара оба будут:

а) красными; б) синими; в) белыми.

3. В ящике 10 шаров: 5 красных, 2 синих и 3 белых. Найти вероятность того, что из двух случайно взятых шаров: а) один будет красный, а второй — синий; б) один будет красный, а второй — белый; в) один будет белый, а второй — синий.

4. В ящике 10 шаров: 5 красных, 2 синих и 3 белых. Найти вероятность того, что два случайно взятых шара будут иметь один и тот же цвет (Указание: использовать выводы задачи №2).

5. В ящике 10 шаров: 5 красных, 2 синих и 3 белых. Найти вероятность того, что два случайно взятых шара будут иметь разные цвета (Указание: использовать выводы задачи №3).

6. В первом ящике 5 красных и 10 синих шаров, во втором — 4 красных и 2 синих. Из произвольного ящика достают один шар. Какова вероятность того, что это будет синий шар?

7. В первом ящике 4 красных и 6 белых, во втором — 9 синих шаров. Из произвольного ящика достают один шар. Какова вероятность того, что это будет синий шар?

8. Сигнальная лампочка прибора перегорает при включении в сеть с вероятностью $P = 0{,}1$. Найти вероятность того, что она перегорит при втором включении.

9. Имеется два одинаковых орудия. Вероятность попадания снаряда в цель для каждого равна 0,9. Производится выстрел из обоих орудий одновременно. Найти вероятности событий: A (цель поражена); B (цель не поражена).

10. Для повышения надежности блок прибора дублируется другим таким же блоком. При выходе из строя первого блока происходит мгновенное переключение на второй. Надежность каждого блока $P = 0{,}9$. Найти надежность системы.

11. Два блока прибора соединены последовательно, причем вероятность отказа каждого из них при включении прибора всегда одинакова и равна 0,1. Определить вероятность того, что прибор не сработает.

12. На складе в клинике имеется 15 электрокардиографов. У пяти из них имеются неисправности. Какова вероятность того, что из трех наугад взятых приборов хотя бы один неисправен?

13. Опыт состоит в одновременном (или последовательном) бросании двух монет. Найти вероятность события A — хотя бы на одной монете выпадет «орел».

14. Опыт состоит в одновременном (или последовательном) бросании двух монет. Найти вероятность события C — монетки выпадут разными сторонами.

15. На обследование прибыла группа в 12 человек, среди которых 4 больны инфекцией. Одновременно обследование проходят 3 человека. Какова вероятность того, что в группе из 3-х человек хотя бы один окажется инфицированным?

16. Определить надежность схемы, состоящей из двух параллельно соединенных элементов, если известно, что вероятность срабатывания каждого элемента
$p = 0{,}8$.

17. Определить вероятность отказа (несрабатывания) схемы, состоящей из двух элементов, соединенных последовательно, если известно, что вероятность отказа каждого элемента $q = 0{,}1$.

Вопросы для обсуждения:
— понятие вероятности события;
— вероятность одиночного события;
— итоговая вероятность в цепочке событий;
— несовместные события;
— независимые события;
— сложение вероятностей;
— умножение вероятностей.
Задание: решение задач № 1—17.

Занятие № 2. Формула Бернулли

Пусть производится n независимых испытаний. Ставится задача определения вероятности того, что ровно в m испытаниях наступит событие A, если вероятность наступления этого события в каждом испытании равна p.

Определим вероятность того, что в первых m испытаниях событие A наступит, а в остальных $n - m$ испытаниях — не наступит. Вероятность такого события может быть получена на основании формулы вероятности произведения независимых событий:

$P = p^m q^{n-m}$, где $q = 1 - p$.

Так как рассматривалась только одна из возможных комбинаций, когда событие A произошло только в первых m испытаниях, то для определения искомой вероятности нужно перебрать все возможные комбинации. Их число будет равно числу сочетаний из n элементов по m, т.е. C_n^m.

Таким образом, вероятность того, что событие A наступит ровно в m испытаниях определяется по формуле:

$$P_n(m) = C_n^m p^m q^{n-m}, \text{ где } C_n^m = \frac{n!}{(n-m)!m!}.$$

Задачи на применение формулы Бернулли

1. Вероятность удачного выполнения сложного химического опыта $p = {}^{1}/_{3}$. Найти: а) вероятность того, что из пяти испытаний удачными будут два; б) из пяти испытаний удачными будут более двух; в) из пяти испытаний удачными будут менее двух.

2. Если в среднем левши составляют 1%, то какова вероятность, что среди 200 человек ровно 2 левши? Указание: решить задачу двумя способами (по формуле Бернулли и по формуле Пуассона) и сравнить результаты.

3. Вероятность рождения мальчика равна 0,51. Найти вероятность того, что среди 50 новорожденных окажется 50 мальчиков.

4. Вероятность брака при изготовлении изделия равна 0,02. Найти вероятность того, что среди 200 произведенных изделий будет не более одного бракованного.

5. Студенты сдают входной контроль по математике. Необходимо ответить на 10 вопросов посредством выбора одного верного ответа из четырех. Студент не подготовлен и ставит ответы случайным образом (наугад). Найти вероятности того, что верными будут: а) ноль ответов; б) один ответ; в) два ответа; г) три ответа; д) четыре ответа; е) пять ответов; ж) шесть ответов; з) более шести ответов. Указание: ответы привести с точностью до тысячных.

6. Какова вероятность того, что в результате бросания игральной кости шесть раз подряд выпадут единицы? (Ответ: (1/6) в шестой степени или 0,000021…).

7. Какова вероятность того, что в результате бросания игральной кости шесть раз подряд выпадут только четные числа? (Ответ: (1/2) в шестой степени или 0,015625).

8. По многолетним наблюдениям, вызов врача в некоторый дом оценивается вероятностью 0,4. Найти вероятность того, что из пяти вызовов врача два будут в данный дом. (Ответ: 0,3456).

9. Найти распределение случайной величины, соответствующей выпадению одной из двух сторон подброшенной монеты («орла»). Всего опытов 5. Проверить выполнение условия нормировки. (Ответ: ${}^{1}/_{32}, {}^{5}/_{32}, {}^{10}/_{32}, {}^{10}/_{32}, {}^{5}/_{32}, {}^{1}/_{32}$).

Вопросы для обсуждения:

— применение метода графов для расчета надежности приборов;

— применения комбинаторики и формулы Бернулли;

Задание: решение задач №1—9.

Занятие № 3. Дискретная случайная величина (ДСВ)

Случайная величина называется **дискретной**, если она принимает отдельные, изолированные возможные значения с определенными вероятностями. Д.с.в. обычно задают в виде таблицы, которую называют законом распределения. В качестве примера приведем закон распределения ДСВ количество правильных ответов из 10, по условию задачи № 5 из предыдущего занятия.

Таблица 1

X	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
P	*0,056*	*0,188*	*0,282*	*0,250*	*0,146*	*0,058*	*0,016*	*0,003*	*0,0004*

Числовые характеристики дискретной случайной величины (д.с.в.)

Определение: **математическим ожиданием *M (X)*** дискретной случайной величины *X* называется сумма произведений ее значений на соответствующие вероятности:

$$M(X) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_n p_n = \mu.$$

Из определения вытекают, например, следующие свойства математического ожидания:

1) $M(C) = C$;

2) $M(k \cdot X) = k \cdot M(X)$.

Для характеристики разброса значений случайной величины относительно математического ожидания вводят понятие **дисперсии**:

$$D(X) = \sigma^2 = M[(X-\mu)^2] = M(X^2) - \mu^2$$

и **стандартного отклонения** $\sigma = \sqrt{D}$.

Для иллюстрации найдем основные числовые характеристики дискретной случайной величины рассматриваемого примера. Для этого добавим в таблицу еще одну строку сверху, в которой приведены квадраты значений случайной величины, и строку снизу, в которой приведены частоты встречаемости результата при условии повтора указанной ситуации 1000 раз. Получим таблицу.

X^2	0	1	4	9	16	25	36	49	64
X	0	1	2	3	4	5	6	7	8
P	0,056	0,188	0,282	0,250	0,146	0,058	0,016	0,003	0,0004
m	56	188	282	250	146	58	16	3	0,4

По данным таблицы и с учетом формулы для математического ожидания найдены математическое ожидание случайной величины (1), математическое ожидание квадрата значений случайной величины (2), квадрат математического ожидания (3). Последний нужен для нахождения характеристик разброса значений случайной величины, а именно: дисперсии (4) и стандартного отклонения (5).

1. $M(X) = 2{,}49 = \mu$.
2. $M(X^2) = 8.10$.
3. $[M(X)]^2 = \mu^2 = 6{,}2$.
4. $D(X) = M(X^2) - \mu^2 = 1{,}9$.
5. $\sigma = \sqrt{D(X)} = 1{,}38$.

Для графической иллюстрации закона распределения д.с.в. применяют так называемый полигон. Полигоном частот называют ломаную линию, получаемую посредством соединения точек: $(x_1, m_1), (x_2, m_2), \dots, (x_n, m_n)$ — отрезками прямой. Полигоном относительных частот называют ломаную линию, отрезки которой соединяют точки $(x_1, p_1), (x_2, p_2), \dots, (x_n, p_n)$.

Задачи

1. Вероятность удачного выполнения сложного химического опыта p=1/3. Найти закон распределения дискретной случайной величины X — число удачных опытов из пяти.

2.

x_i	0	1	2	3	4	5
p_i	32/243	80/243	80/243	40/243	10/243	1/243
m_i	32	80	80	40	10	1

Контроль: $\sum_0^5 p_i = 1$; $\sum_0^5 m_i = 243$.

Построить многоугольник (полигон) распределения.

3. На складе в клинике имеется 15 электрокардиографов. У пяти из них имеются неисправности. Случайным образом берут три прибора. Найти закон распределения дискретной случайной величины X — числа неисправных приборов (из трех).

x_i	0	1	2	3
p_i	$24/91$	$45/91$	$20/91$	$2/91$
m_i	24	45	20	2

Контроль: $\sum_0^3 p_i = 1$; $\sum_0^3 m_i = 91$.

Построить многоугольник (полигон) распределения.

4. Опыт состоит в одновременном (или последовательном) бросании двух одинаковых монет. Найти закон распределения случайной величины X — числа выпавших орлов.

x_i	0	1	2
p_i	$1/4$	$1/2$	$1/4$
m_i	1	2	1

Контроль: $\sum_0^2 p_i =$ 1. $\sum_0^2 m_i = 4$.

5. Монета бросается 4 раза. Построить многоугольник распределения случайной величины X — числа выпадения герба.

x_i	0	1	2	3	4
p_i	1/16	4/16	6/16	4/16	1/16
m_i	1	4	6	4	1

Контроль: $\sum_0^4 p_i = 1$.

6. В урне 8 шаров, из которых 5 белых, остальные — черные. Из нее вынимаю наудачу 3 шара. Найти закон распределения числа белых шаров в выборке.

X	0	1	2	3
P	$1/56$	$15/56$	$30/56$	$10/56$
m_i	1	15	30	10

Контроль: $\sum_0^3 p_i = 1$.

Вопросы для обсуждения:
— математическое ожидание ДСВ;
— дисперсия, стандартное отклонение ДСВ;
— построение полигона ДСВ.
Задание: решение задач №1—5.

Занятие № 4.
Непрерывная случайная величина (НСВ)

Функция распределения и плотность вероятности непрерывной случайной величины

Множество значений непрерывной случайной величины несчетно и обычно представляет собой некоторый промежуток — конечный или бесконечный.

Непрерывную случайную величину описывают **функцией распределения**, которая определяется как вероятность того, что случайная величина X в результате эксперимента примет значение, меньшее x:

$$F(x) = P(X < x).$$

Из определения функции распределения следует, что она принимает только положительные значения от нуля до единицы и является неубывающей функцией.

Если функция распределения непрерывной случайной величины дифференцируема, то более наглядное представление о случайной величине дает **плотность вероятности случайной величины** $f(x)$, которая связана с функцией распределения $F(x)$ формулами:

$$F(x) = \int_{-\infty}^{x} f(t)dt \text{ и } f(x) = \frac{dF}{dx}, \text{ при этом } \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Свойства плотности распределения вероятностей:
— $f(x)$ неотрицательная, т.е. $f(x) \geq 0$;
— вероятность попадания в интервал $[a;b]$ равна:

$$P\{a \leq x \leq b\} = \int_{a}^{b} f(x)dx = F(b) - F(a).$$

Кроме математического ожидания, дисперсии и стандартного отклонения, для описания случайной величины используются следующие параметры:

Медиана — значение случайной величины, которое делит всю совокупность на две равные части: половина из них меньше медианы, а другая половина — больше.

Мода — наиболее вероятное значение случайной величины (центр сгущения). Распределение с одной модой называется унимодальным, а с несколькими модами — мультимодальным. Мультимодальность распределения свидетельствует о существенной неоднородности множества значений исследуемой величины. Для симметричного унимодального распределения мода совпадает с математическим ожиданием, с медианой и с абсциссой центра симметрии графика функции распределения.

Коэффициентом вариации случайной величины X называется число $V(X)$, равное

$$V(X) = \frac{\sigma(X)}{M(X)}.$$

Коэффициент вариации, как и дисперсия и стандартное отклонение, является мерой рассеяния распределения, служит для измерения стандартного отклонения в долях математического ожидания.

Пример. *Плотность распределения случайной величины* X *задана функцией*

$f(x) = \frac{a}{1+x^2}$. Найти значение параметра a.

Согласно свойству 4 функции плотности распределения имеем:

$$\int_{-\infty}^{+\infty} \frac{a}{1+x^2}\, dx = 1, \text{ т.е. } a\cdot\operatorname{arctg} x \,\Big|_{-\infty}^{+\infty} = 1,\ a\cdot\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = 1,$$

следовательно, $a\cdot\pi=1$ и $a=\frac{1}{\pi}$.

Задачи

1. Случайная величина X задана функцией распределения:

$$F(x) = \begin{cases} 0, \text{при } x \le -1, \\ a(x+1)^2, \text{при } -1 < x \le 2, \\ 1, \text{при } x > 2. \end{cases}$$

Найти значение a и построить графики $F(x)$ и $f(x)$.

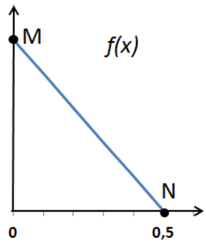
2. Является ли плотностью распределения случайной величины каждая из следующих функций:

a) $f(x) = \frac{x}{\pi(1+x^2)}$ при $x \in (-\infty; +\infty)$;

b) $f(x) = \begin{cases} \frac{1}{2}, \text{при } x \in (-1; 1], \\ 0, \text{при } x \overline{\in} (-1; 1]; \end{cases}$

c) $f(x) = \begin{cases} 0, \text{при } x < 0 \text{ и } x > 2 \\ ax^2, \text{при } 0 \le x \le 2. \end{cases}$

3. Кривая плотности вероятности распределения непрерывной случайной величины X имеет вид, указанный на рисунке.

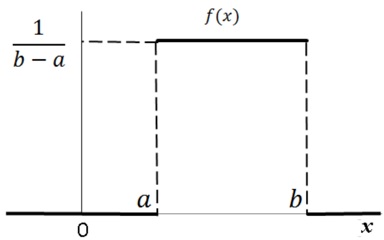


Найти выражение для *f(x)*, функцию распределения *F(x)* и вероятность события $\left\{X \in \left(\frac{1}{4}; 1\right)\right\}$.

Равномерный закон распределения

Непрерывная случайная величина X имеет равномерное распределение на отрезке *[a,b]*, если ее плотность вероятности $f(x)$ постоянна на этом отрезке, а вне него равна нулю:

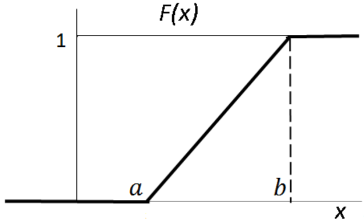
$$f(x) = \begin{cases} \dfrac{1}{b-a}, \text{при } x \in [a,b], \\ 0 \text{ при } x \notin [a,b]. \end{cases}$$



$$F(x) = \int_a^x \frac{dt}{b-a} = \frac{t}{b-a}\bigg|_a^x = \frac{x-a}{b-a}.$$

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{cases} \mathbf{0}, \text{при } \boldsymbol{x} \le \boldsymbol{a}, \\ \dfrac{\boldsymbol{x}-\boldsymbol{a}}{\boldsymbol{b}-\boldsymbol{a}}, \text{при } \boldsymbol{a} < \boldsymbol{x} \le \boldsymbol{b}, \\ \mathbf{1}, \text{при } \boldsymbol{b} < \boldsymbol{x}. \end{cases}$$

График F (x):



Показательный закон распределения

Непрерывная случайная величина *X* имеет *показательный (или экспоненциальный)* закон распределения, если ее плотность вероятности имеет вид:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, \text{при x } \geq 0, \\ 0, \text{при } x < 0, \end{cases}$$

где λ >0 — параметр распределения.

График функции плотности распределения приведен на рисунке:

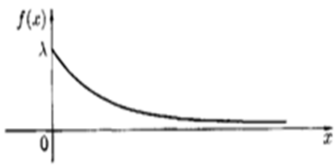
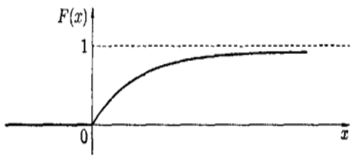


График функции распределения *F (x)* имеет вид:



Вопросы для обсуждения:
— математическое ожидание НСВ;
— дисперсия, стандартное отклонение, анизотропия и эксцесс НСВ;
— функция распределения и функция плотности вероятности НСВ.

Задание: решение задач №1—3.

Занятие № 5. Нормальный закон распределения

Нормальное распределение (распределение Гаусса) имеет следующие характеристики:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$P\{a < X < b\} = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$\Phi_0(x) = \frac{1}{2\pi} \cdot \int_0^x e^{-\frac{t^2}{2}} dt,$$

$$P\{a < X < b\} = \Phi_0\left(\frac{b-\mu}{\sigma}\right) - \Phi_0\left(\frac{a-\mu}{\sigma}\right),$$

где μ — математическое ожидание, σ — стандартное отклонение, Φ_0 — функция Лапласа.

Распределение соответствует нормальному закону в случае, когда на результат измерения влияют многочисленные случайные факторы.

Плотность вероятности $f(x)$ в этом случае имеет вид симметричного колокольчика.

Проблема формулируется следующим образом: можно ли считать, что случайная величина имеет **нормальное распределение**?

Сформулируем основную и альтернативную гипотезы.

Гипотеза: случайная величина имеет нормальное распределение, значения параметров распределения заранее не известны. Альтернативная гипотеза: распределение случайной величины отличается от нормального.

Параметрами распределения обычно являются математическое ожидание и дисперсия. Иногда вместо дисперсии рассматривают корень из нее, то есть стандартное отклонение.

Критерии соответствия распределения нормальному закону:

1. Критерий χ^2 Пирсона.

Критерий согласия Пирсона позволяет осуществлять проверку эмпирического и теоретического (либо другого эмпирического) распределений одного признака. Данный критерий применяется, в основном, в двух случаях:

— для сопоставления эмпирического распределения признака с теоретическим распределением (нормальным, показательным, равномерным либо каким-то иным законом);

— для сопоставления двух эмпирических распределений одного и того же признака.

Идея метода — определение степени расхождения соответствующих частот m_i (эмпирической) и m_i' (теоретической) попадания значений случайной величины в интервалы. Чем больше это расхождение, тем больше значение

$$\chi^2 = \sum_i \frac{(m_i - m_i')^2}{m_i'}.$$

Объемы выборок должны быть не меньше 50 и необходимо равенство сумм частот:

$$\sum_i m_i = \sum_i m_i'.$$

Нулевая гипотеза H_0 ={два распределения практически не различаются между собой}; альтернативная гипотеза — H_1 ={расхождение между распределениями существенно}.

Гипотеза H_0 принимается, если $\chi^2_{эмп} < \chi^2_{крит}$ (где $\chi^2_{эмп}$ — вычисленное значение критерия Пирсона, а $\chi^2_{крит}$ — минимальное значение критерия соответствующее гипотезе H_0), и отвергается в противном случае.

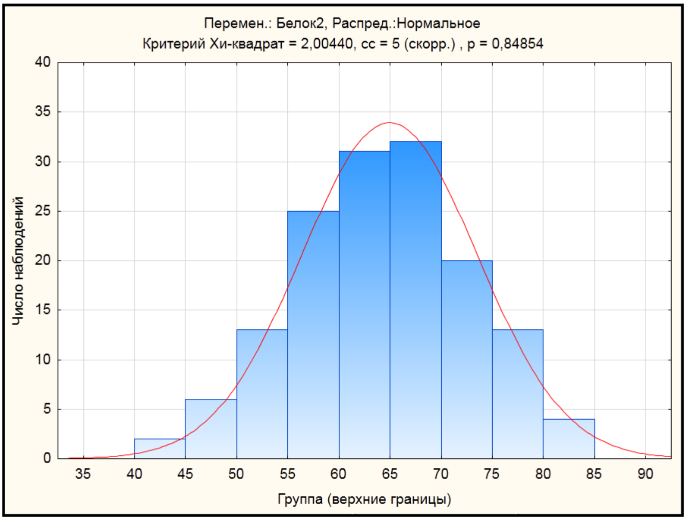
Критерий χ^2 используется для выборок объема 50 и более.

Для выборок меньшего объема мы будем использовать критерий **Шапиро-Уилка** (в стандарте ГОСТ РИСО 5479-2002 применяется при объемах выборки от 8 до 50).

Гистограмма — это способ представления статистических данных в графическом виде — в виде столбчатой диаграммы. Она отображает распределение отдельных наблюдений. Иногда ее называют частотным распределением, так как гистограмма показывает частоту появления измеренных значений параметров объекта.

Высота каждого столбца указывает на число наблюдений в выбранном диапазоне, а количество столбцов — на число выбранных диапазонов.

Важное преимущество гистограммы заключается в том, что она позволяет наглядно представить тенденции изменения измеряемых параметров объекта и зрительно оценить закон их распределения. Кроме того, гистограмма дает возможность быстро определить центр, разброс и форму распределения случайной величины.



Пример гистограммы для анализа крови (общий белок) для 146 пациентов после операции (см. таблицу № 7 приложения).

Задачи

1. Нормально распределенная случайная величина *X* задана плотностью вероятностей:

$$f(x) = \frac{1}{2\pi} e^{-x^2/2}.$$

Найти:

а) вероятность попадания случайной величины в интервал (1,3);

б) симметричный относительно математического ожидания интервал, в который с вероятностью 0,8926 попадет случайная величина *X* в результате опыта;

в) построить нормальную кривую $\boldsymbol{f(x)}$.

2. Математическое ожидание нормально распределенной случайной величины равно μ = 3 и стандартное отклонение σ = 2. Записать плотность вероятности *X*.

3. Записать плотность вероятности нормально распределенной случайной величины *X* если $M(X) = 3$, а $D(X) = 16$.

4. Нормально распределенная случайная величина *X* задана плотностью распределения:

$f(x) = \frac{1}{5\sqrt{2\pi}}\ e^{-(x-1)^2/50}$.

Найти математическое ожидание и дисперсию.

5. Задана функция распределения нормального закона:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

Найти плотность распределения $f(x)$.

Вопросы для обсуждения:

- математическое ожидание нормального распределения;
- дисперсия, стандартное отклонение нормального распределения;
- построение графиков функций нормального распределения.

Задание: решение задач №1—6.

Занятие № 6. Знакомство с программой STATISTICA

Программа STATISTICA была создана в 1984 году в США. В 1993 году разработана версия для Windows. STATISTICA — программный пакет для статистического анализа, разработанный компанией StatSoft, реализующий функции анализа данных, управления данными, добычи данных, визуализации данных с привлечением статистических методов. Торговая марка — STATISTICA.

Сейчас имеет высочайшие рейтинги. Руководство компании находится в Оклахоме.

Мы будем использовать программу STATISTICA в учебных целях для решения следующих задач:

— Получение описательных статистик, построение различных графиков, удобных для восприятия и анализа имеющейся информации.

— Использование методов статистики для решения задач о различии переменных.

— Использование методов статистики для решения задач о взаимосвязи (корреляции) переменных.

— Научимся решать вопрос о принадлежности выборок (различного объема) генеральной совокупности, подчиняющейся нормальному закону распределения, и использовать критерии t-Стьюдента, F- Фишера в задачах о различии, критерий (коэффициент корреляции) Пирсона в задачах о корреляции.

— В случае если гипотеза о нормальности не может быть принята (в частности, при работе с малыми выборками, когда проверить гипотезу о нормальности не всегда представляется возможным), мы научимся использовать так называемые непараметрические критерии. Например, критерий знаков, критерий Вилкоксона, Манна-Уитни и другие для решения задач о различии; критерий (коэффициент корреляции) Спирмена и другие для решения задач о корреляции.

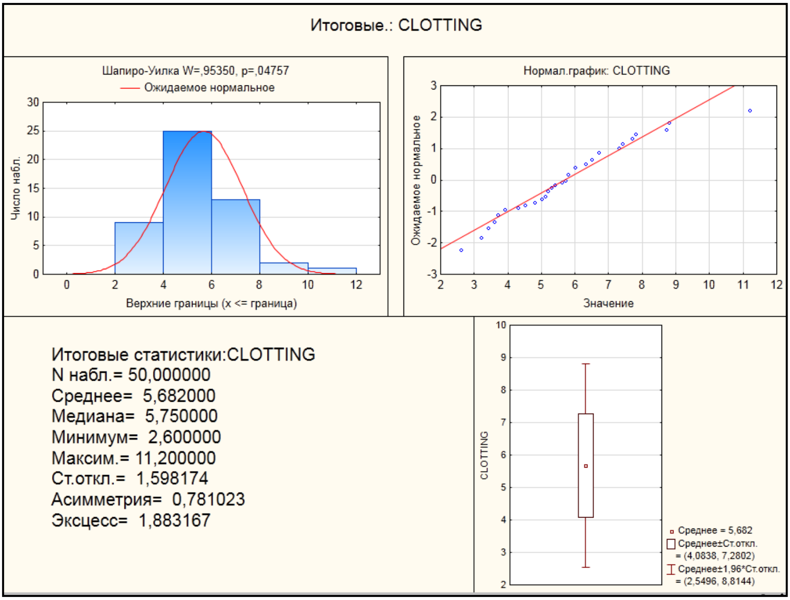
— Мы научимся использовать статистические критерии (точный критерий Фишера и хи-квадрат) для так называемых таблиц сопряженности 2x2.

Ход выполнения практического занятия

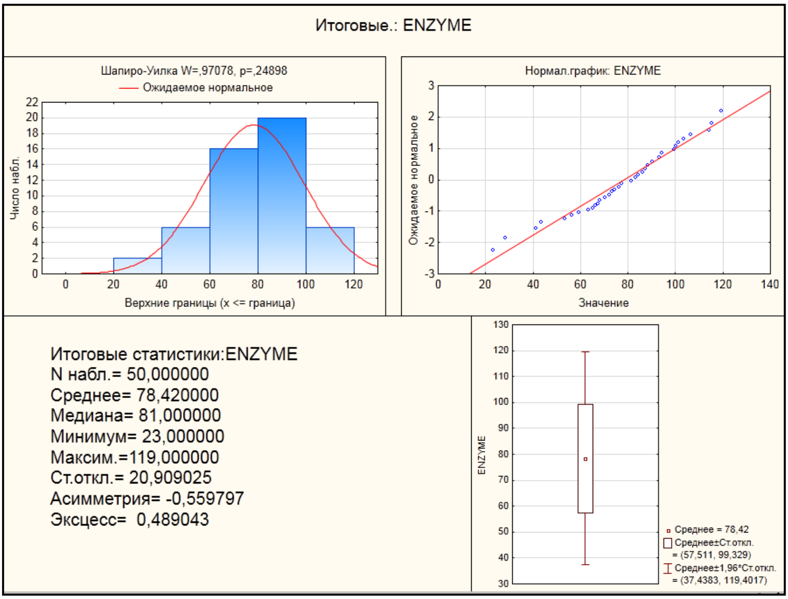
1. Открыть программу. Закрыть все лишние окна. Создать таблицу данных: 5 переменных и 54 наблюдения формат данных числовой с двумя знаками после запятой (число десятичных разрядов 2).

2. Открыть папку примеров. Открыть «Datasets». Открыть в ней таблицу «Surgical units». Скопировать с заголовками таблицу и вставить данные в созданную нами таблицу. Теперь удалим из нашей таблицы переменные (колонки) 2 и 5. Удалим также наблюдения с 51 по 54. Копируем таблицу с заголовками и вставляем ее в отчет Word. Преобразуем таблицу, представив ее в «две колонки». Получим: (Таблица №1 приложения).

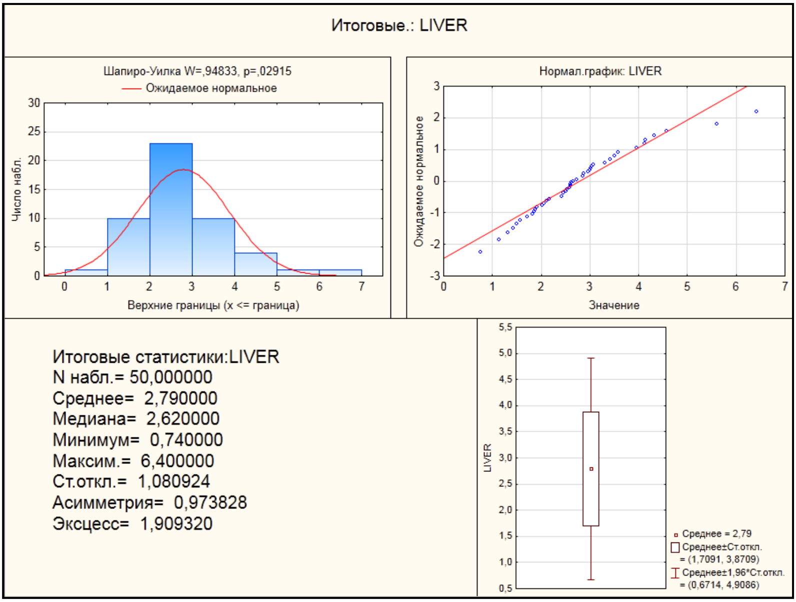
3. <u>Выполнить анализ данных на предмет нормальности</u>: Анализ. Основные статистики и таблицы, ОК. Добавляем на вкладках: <u>Дополнительно</u> → Медиана, Асимметрия, Эксцесс. <u>Нормальность</u> → оставляем только критерий Шапиро-Уилка. <u>Переменные</u> → Все. На вкладке Быстрый выбираем График №1. Получим три графика и скопируем их в отчет с помощью клавиш Правка → Захват экрана → Захват прямоугольника.



Гипотеза о соответствии данных нормальному закону (H_0) отвергается.



H_0 принимается.



H_0 отвергается.

Использование встроенной опции позволяет проанализировать данные как количественно (описательные статистики, критерий нормальности Шапиро-Уилка), так и качественно (вид гистограммы, нормальный вероятностный график, график, показывающий среднее и разброс значений).

Мы не будем сразу анализировать всю информацию, которую дает график №1. Отметим только, что **качественно** вид распределения характеризуется гистограммой и нормальным вероятностным графиком, на котором близость распределения к нормальному закону определяется тем, насколько тесно точки группируются вблизи прямой. Красная прямая — это преобразованный с помощью логарифмирования и замены переменных «колокольчик» нормального распределения, а точки — это преобразованные таким же путем эмпирические наблюдения. Если точки плохо стыкуются с прямой, то, скорее всего, данные не подчиняются нормальному закону. Доводом в пользу гипотезы о нормальном законе распределения является совпадение или близость значений среднего и медианы. Если параметры формы — асимметрия и эксцесс отличаются от нуля не более чем на 0,5 в ту или другую сторону, это также свидетельствует в пользу гипотезы о нормальности распределения. Для нормального распределения — симметричного колокольчика указанные параметры формы равны нулю. **Количественно**, то есть более определенно, выводы о нормальности дают критерии. Мы будем использовать в дальнейшем критерий Шапиро-Уилка и критерий хи-квадрат. Последний критерий используется для выборок большего объема.

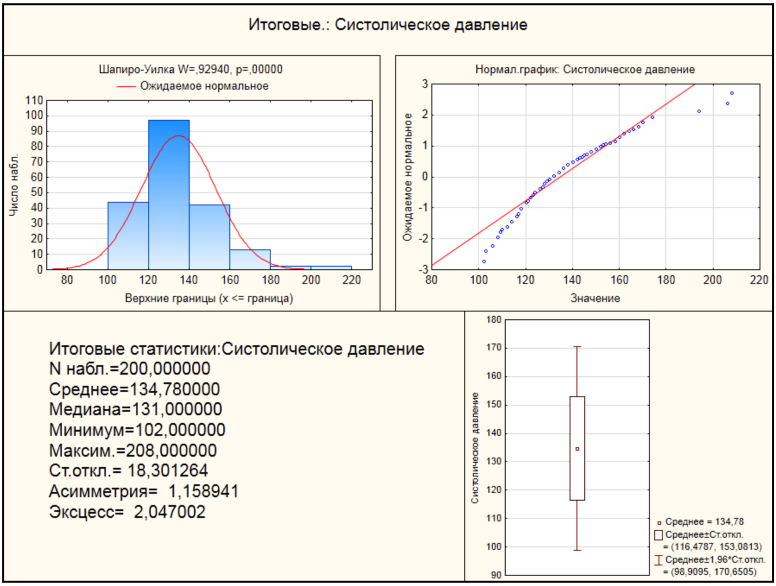
Укажем причины внимания к проверке нормальности:

1. Факт нормальности сам по себе несет информацию об однородности выборок (здоровые пациенты близкого возраста, пола, рода занятий и др.), многие параметры и анализы подчиняются нормальному закону.
2. Все критерии, которые мы будем использовать, делятся на две группы. Одни применяются для нормально распределенных переменных, другие, называемые непараметрическими, — для всех остальных.

Теперь про полученные значения критерия Шапиро-Уилка (W). Если $W_{\text{эмп.}} > W_{\alpha,n}^{\text{кр.}}$ (где α — уровень значимости, n — объем выборки), нулевая гипотеза (H_0 — переменная подчиняется нормальному закону) принимается. Параметр p показывает вероятность нулевой гипотезы. При принятом уровне значимости 0,05 мы можем отвергнуть нулевую гипотезу, если $p < 0{,}05$. Применение критерия позволяет принять гипотезу о нормальности только для переменной ENZYME.

Пример выполнения практического задания

1. Открыть программу. Закрыть все лишние окна. Создать таблицу данных: 10 переменных и 200 наблюдений.
2. Открыть папку примеров. Открыть «Datasets». Открыть в ней таблицу «HeartDiseases» (Таблица №2 приложения). Скопировать первые 200 наблюдений из колонки «Систолическое давление». Вставить наблюдения в созданную нами таблицу. Изменить название переменной с Пер1 на «Систолическое давление».
3. <u>Выполнить качественный анализ данных на предмет нормальности</u>: Анализ. Описательные статистики. Опции: добавляем медиану, асимметрию и эксцесс. Строим график 1:

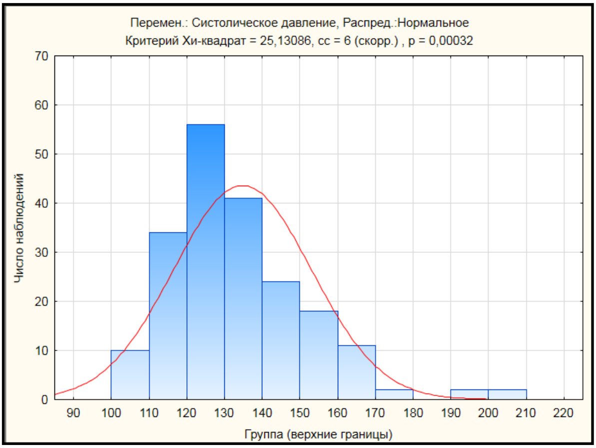


4. Прокомментировать вид гистограммы, ее соответствие подгоночному колокольчику. Обсудить, насколько близко точки ложатся на прямую линию. Проанализировать итоговые статистики: различие среднего и медианы, отличие от нуля (как должно быть у нормального закона) асимметрии и эксцесса. Обсудить график в правом нижнем углу — среднее, среднее±σ, среднее±1,96σ (доверительный интервал, в который попадают значения случайной величины, распределенной по нормальному закону, с вероятностью ровно 95%).

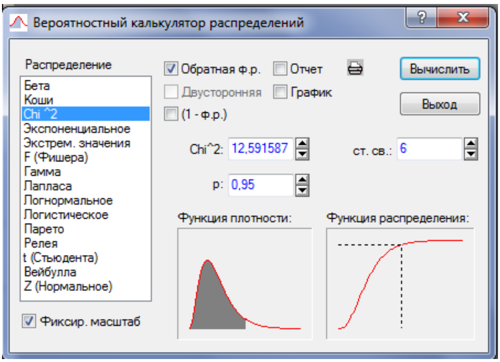
5. Открыть вероятностный калькулятор (Анализ → Вероятностный калькулятор) и сравнить распределение Chi^2 с симметричным распределением t-Стьюдента.

6. Создать файл Word с заголовком Занятие №1, который потом поместить в папку со своей фамилией, а ее — в папку, соответствующую номеру группы. Копировать построенный график 1, используя: Правка. Захват прямоугольника.

7. Проанализировать соответствие данных нормальному закону с помощью критерия χ^2 (хи-квадрат). Анализ. Подгонка распределений (посмотреть опции и комментировать). Гистограмма.



8. Анализируем график визуально. Значение эмпирического критерия χ^2 равно 25,13. Далее открываем вероятностный калькулятор: Анализ. Вероятностный калькулятор: выбираем распределение Chi^2, ставим степени свободы 6 и задаем $p = 0{,}95$. Вычислить. Получаем значение $\chi^2_{\text{крит.}}$=12,59.



Поскольку $\chi^2_{\text{эмп.}} > \chi^2_{\text{крит.}}$, делаем вывод: нулевая гипотеза о нормальном законе распределения отвергается. Можно также просто воспользоваться найденной вероятностью нулевой гипотезы $p = 0{,}00032$.

Вопросы для обсуждения:

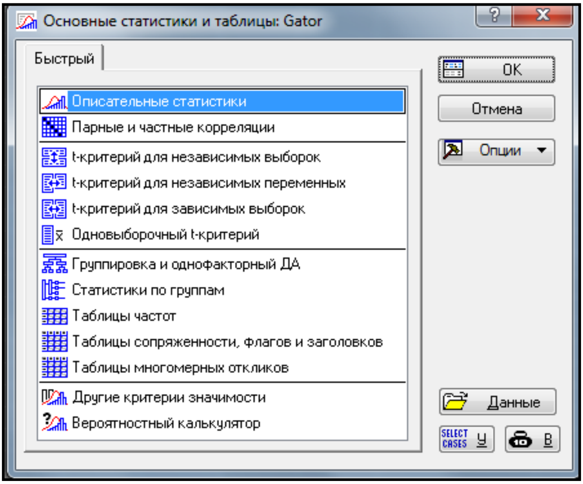
— основные окна пакета;

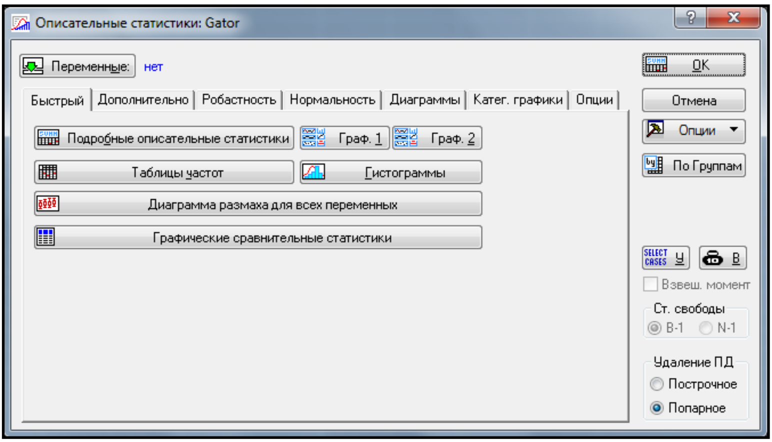
— вероятностный калькулятор;

— ввод и вывод данных в пакете STATISTICA.

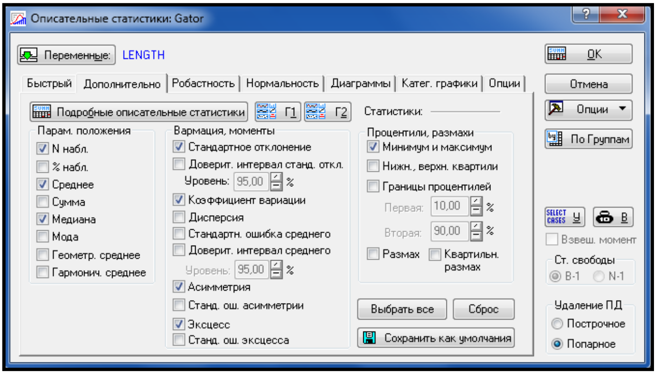
Занятие № 7. Основные описательные статистики. Построение и анализ гистограмм. Проверка гипотезы о нормальности с помощью критериев согласия

Открыть папку примеров программы STATISTICA «Аллигаторы» (приложение, табл. №3).





В открывшемся окне выбрать «Подробные описательные статистики, но сначала нажать кнопку «Дополнительно» и выбрать нужные статистики.

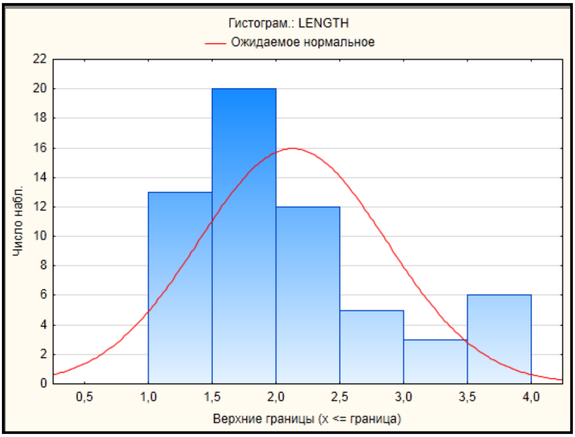


Нажать кнопку «Быстрый», «Подробные описательные статистики», выбираем переменную «Длина» и ОК! Получим:

Переменная	Описательные статистики (Gator)								
	N набл.	Среднее	Медиана	Минимум	Максим.	Ст.откл.	Коэф.Вар.	Асимметрия	Эксцесс
LENGTH	59	2,130339	1,850000	1,240000	3,890000	0,737407	34,61453	0,953906	-0,138218

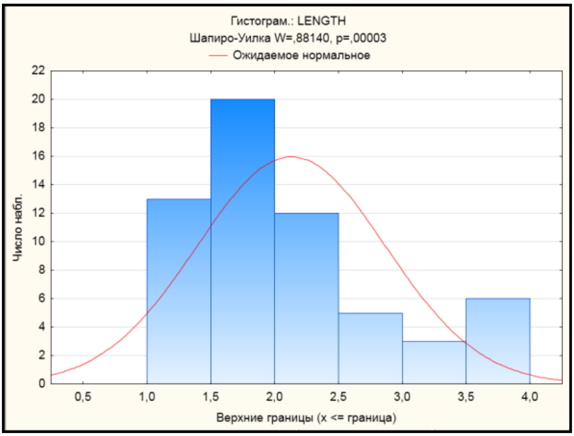
Значения среднего и медианы заметно отличаются друг от друга (2,13 и 1,85), что указывает на асимметрию распределения. Минимальное и максимальное значения параметра (длина) составляют 1,24 и 3,89 соответственно. Стандартное отклонение $\sigma = 0{,}737407$. Коэффициент вариации — 34,61453. Далее приведены значения параметров формы: асимметрия (скошенность) равна 0,953906, что указывает на существенное отклонение распределения для нашей переменной от симметричного (нормального) закона. И последнее — эксцесс (или показатель островершинности) отрицателен и равен 0,138218.

Для визуальной иллюстрации построить гистограмму распределения. Для этого нажать кнопку «Гистограмма»:



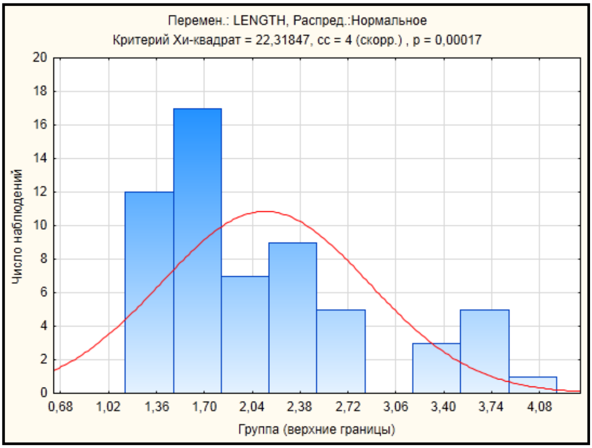
Видно, что гистограмма действительно несимметрична (правый край более пологий) и весьма заметно отличается от красной кривой для нормального распределения.

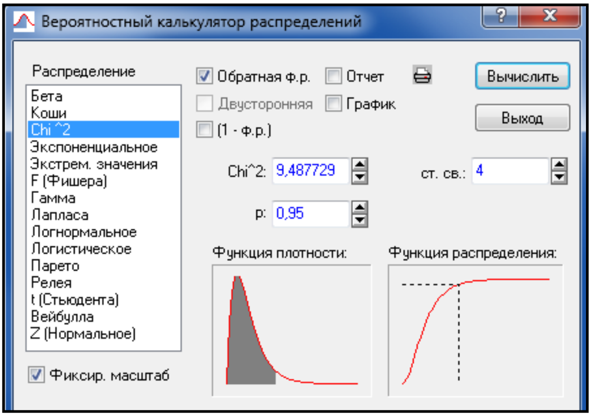
Применим для проверки выборки на нормальность тест Шапиро-Уилка (задано 10 интервалов).



Результат теста Шапиро-Уилка — вероятность того, что выборка взята из генеральной совокупности, которая распределена по нормальному закону p=0,00003. Следовательно, гипотеза о нормальности для нашей выборки отвергается.

Применим для проверки нормальности тест (критерий) хи-квадрат. Число интервалов задано равным 10.

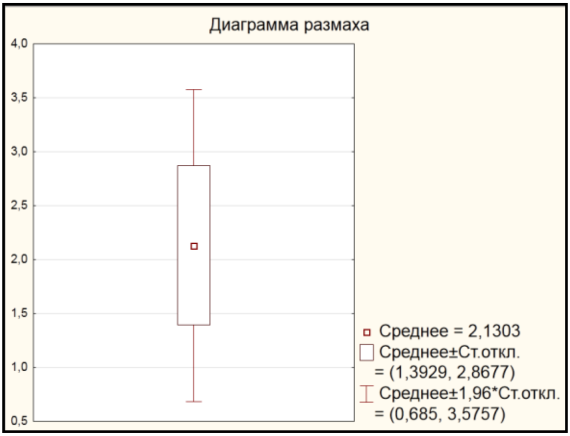




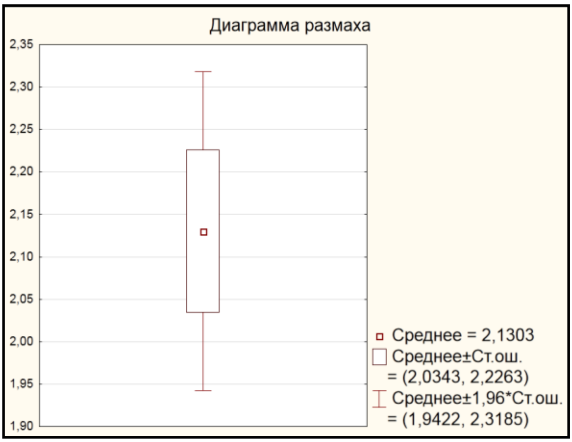
Поскольку эмпирическое значение критерия хи-квадрат (22,3) больше критического (9,5), нулевая гипотеза о нормальности отвергается. Найдена вероятность нулевой гипотезы 0,00017.

Таким образом, оба критерия показывают, что для анализируемой переменной гипотеза о нормальном законе отвергается.

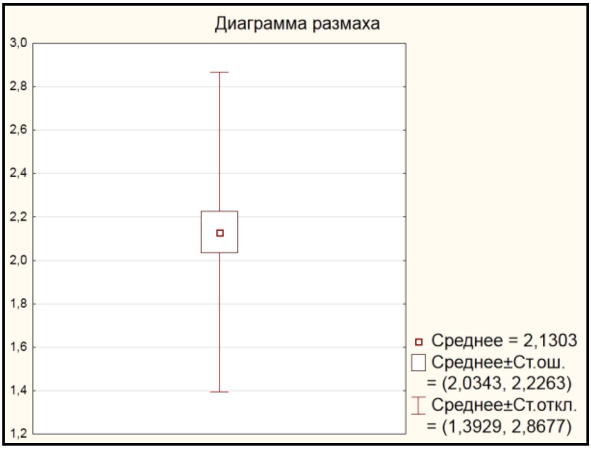
Построим диаграммы размаха в четырех вариантах, предлагаемых программой STATISTICA. Используем кнопку «Опции», где необходимо отметить последовательно все 4 варианта.



На первом графике точка в середине показывает среднее, а ящик — среднее $\pm\sigma$. Усы показывают диапазон, в который попадает 95% значений переменной.



На втором графике показано среднее, среднее ± ошибка среднего и 95% — доверительный интервал для среднего.



На третьем графике точка в середине показывает среднее, ящик — среднее $\pm$ ошибка среднего. Усы показывают диапазон, в который попадает 95% значений переменной (среднее $\pm\sigma$).

На четвертом графике точкой показана медиана, ящик показывает 50% квартиль, а усы — минимальное и максимальное значение переменной. Данный график отражает несимметричность распределения, замеченную ранее по виду гистограммы.

Вопросы для обсуждения:

— описательные статистики;

— гистограмма;

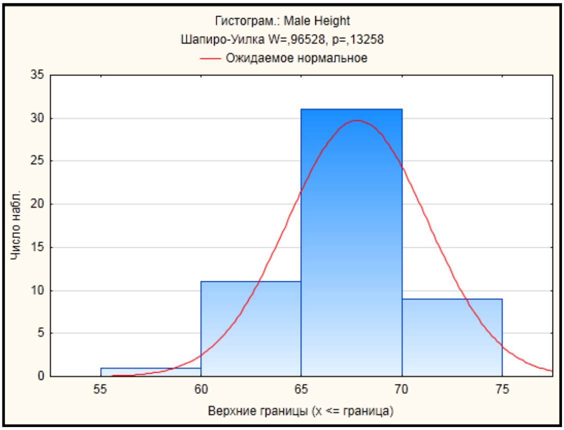
— качественная оценка нормальности в пакете STATISTICA.

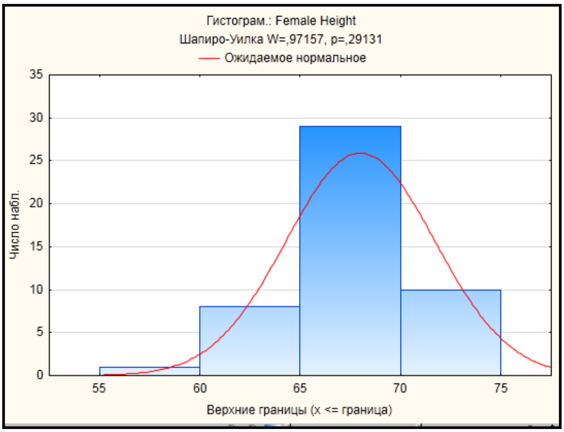
Занятие № 8. Критерии Шапиро-Уилка и χ^2 для проверки гипотезы нормальности (самостоятельная работа)

Разбор задач из папки примеров программы STATISTICA: «Characteristic Height» (таблица №4 приложения), «Heart Diseases» (таблица №2 приложения).

Папка «**Characteristics Height**» — сравнение независимых переменных.

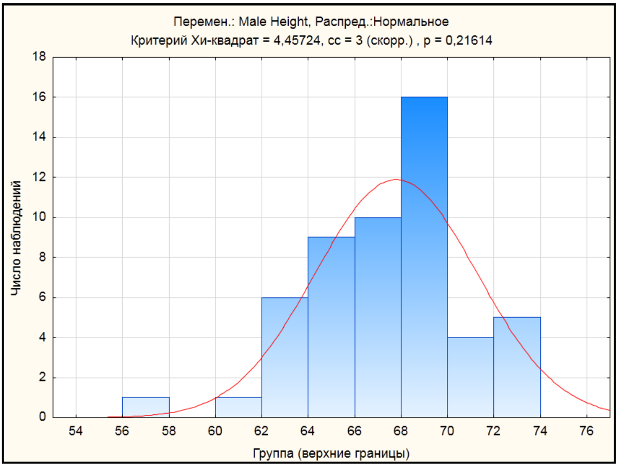
1. Проверка нормальности с помощью теста Шапиро-Уилка.

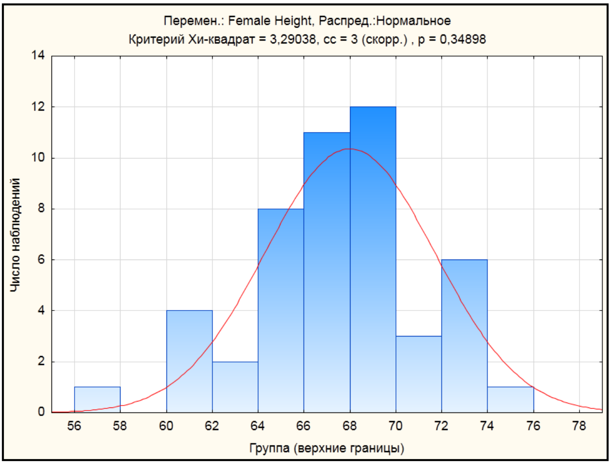


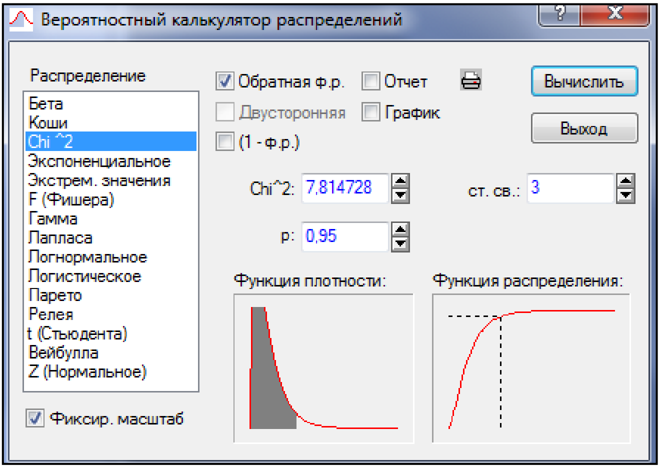


Результат теста — вероятность того, что выборка взята из генеральной совокупности, которая распределена по нормальному закону p=0,13 и 0,29 соответственно, следовательно, гипотеза о нормальности для обеих выборок принимается.

2. Проверка нормальности с помощью критерия хи-квадрат (Пирсона).







В обоих случаях эмпирические значения критерия хи-квадрат (4,46 и 3,29) меньше критического значения (7,81), вероятность нулевой гипотезы в обоих случаях больше, чем 0,05. Вывод: обе переменные подчиняются нормальному закону распределения.

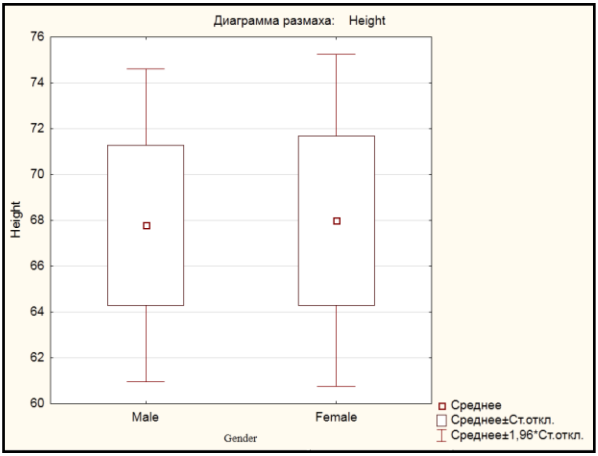
Результаты обоих тестов — Шапиро-Уилка и хи-квадрат — дают близкие значения вероятности нулевой гипотезы. Оба теста, подкрепляя друг друга, дают одинаковый вывод о возможности принятия нулевой гипотезы.

1. Сравним переменные, используя критерии для независимых переменных.

	Т-критерий независимых выборок (CharacteristicsHeight) Замечание: Переменные рассм. как независимые выборки										
Группа 1 и Группа 2	Среднее Группа 1	Среднее Группа 2	t-знач.	сс	p	N набл. Группа 1	N набл. Группа 2	Ст.откл. Группа 1	Ст.откл. Группа 2	F-отн. дисперс.	p дисперс.
Male Height vs. Female Height	67,78846	68,00000	-0,294428	98	0,769054	52	48	3,488561	3,695886	1,122392	0,684868

По t-критерию Стьюдента и по F-критерию Фишера переменные не различаются на уровне значимости 0,05.

2. Построим диаграмму размаха.



Вид диаграммы позволяет визуально убедиться в справедливости вывода об отсутствии различий роста в группах Male и Female.

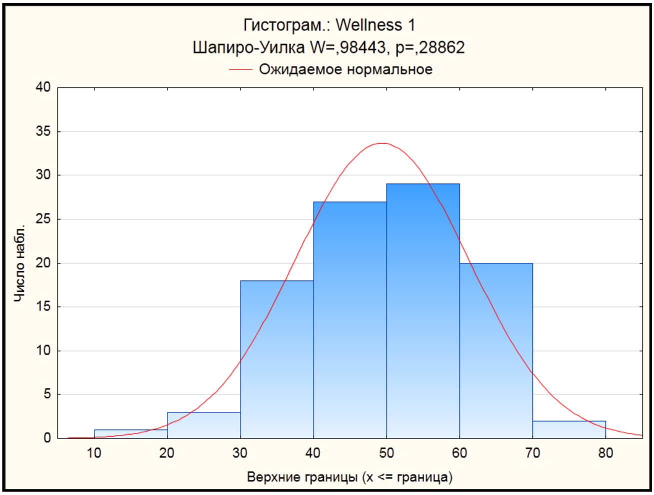
Вопросы для обсуждения:
- критерии нормальности;
- критерии различия;
- количественная оценка нормальности в пакете STATISTICA.

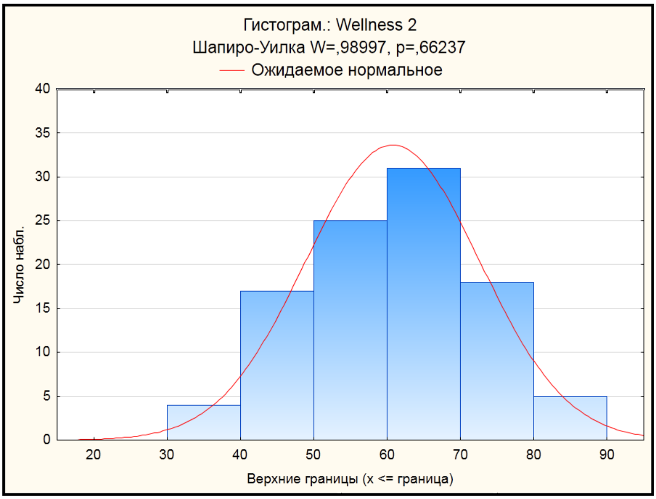
Занятие № 9. Критерий Стьюдента для зависимых выборок

Сравнение зависимых (парных) выборок — Папка «Characteristics»

В таблице №5 приведены показатели уровня здоровья «Wellness1» и «Wellness2», которые в сокращенном виде будем называть W1 и W2. Согласно легенде показатели здоровья измерены у лиц мужского и женского пола возраста от 20 до 79 лет до (W1) и после (W2) занятий в тренажерном зале.

1. Проверка нормальности.





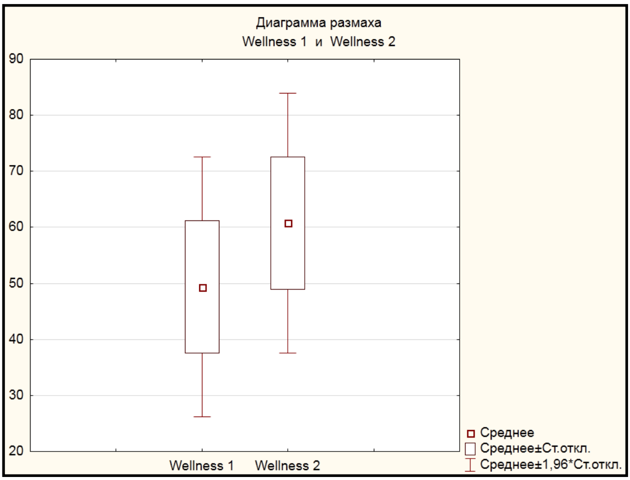
В обоих случаях вероятности нулевой гипотезы (0,29 и 0,66) превышают значение 0,05. Вывод: для обеих переменных принимается гипотеза о нормальности.

2. Сравнение зависимых выборок с помощью t-критерия Стьюдента.

	Т-критерий для зависимых выборок (Characteristics) Отмечены разности, значимые на уровне p < ,05000									
Переменная	**Среднее**	Стд.откл	N	разн.	Стд.откл разн.	t	сс	p	Доверит. -95,000%	Доверит. +95,000%
Wellness 1	49,37074	11,84380								
Wellness 2	60,75626	11,85713	100	-11,3855	17,18203	-6,62641	99	0,000000	-14,7948	-7,97623

Вывод: установлено статистически значимое различие между переменными Wellness1 и Wellness2. Программа STATISTICA выделяет числа красным, что соответствует о том, что модуль эмпирического значения t-критерия выше критического значения. При этом приведена вероятность нулевой гипотезы об отсутствии различия — шесть нулей после запятой.

3. Диаграммы размаха.



Вывод: показатель «здоровье 2» выше чем «здоровье 1». Результат статистически значим на уровне 0,05 (на самом деле — на уровне 0,000000).

Вопросы для обсуждения:

— сравнение различия зависимых выборок по величине математического ожидания;

— расчет критического значения t-критерия с использованием вероятностного калькулятора.

Занятие № 10. Критерии Стьюдента и Фишера для независимых переменных

Программа STATISTICA позволяет работать с данными, представленными в зависимости от желания пользователя: a) в виде двух столбцов (сравнение независимых переменных); b) в виде основного столбца и соседнего, в котором приведена группирующая переменная (сравнение независимых выборок). В остальном алгоритм и результаты одинаковы.

Сравнение роста лиц мужского (52 чел.) и женского (48 чел.) пола.

Данные взяты из таблицы № 4 приложения «Characteristic Height».

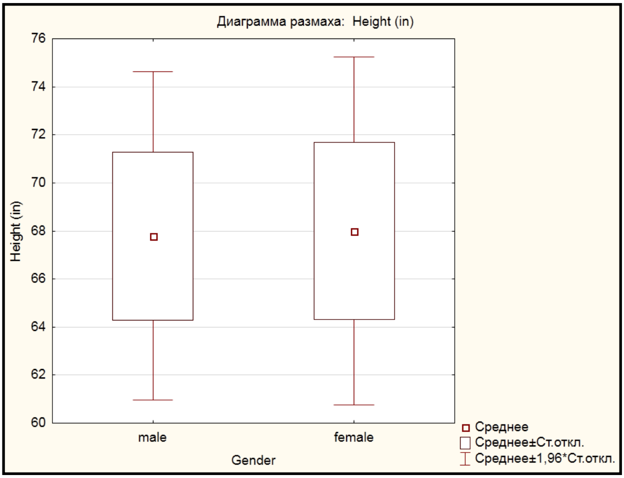
В обоих случаях принимается гипотеза нормальности (критерий Шапиро-Уилка дает вероятность нулевой гипотезы $p = 0{,}13$ и 0,29 для первой и второй группы соответственно).

Анализ. Основные статистики и таблицы, t- критерий для независимых выборок:

	Т-критерии; Группир.: Gender (Characteristics) Группа 1:male Группа 2:female										
Переменная	**Среднее male**	Среднее female	t-знач.	сс	p	N набл. male	N набл. female	Ст.откл. male	Ст.откл. female	F-отн. дисперс.	p дисперс.
Height (in)	67,78846	68,00000	-0,294428	98	0,769054	52	48	3,488561	3,695886	1,122392	0,684868

Нулевая гипотеза (H_0) — различия нет. Альтернативная гипотеза (H_1) — различие есть. Числа в таблице не выделены красным — установлено отсутствие статистически значимого различия между переменными. Вероятность нулевой гипотезы по критерию Стьюдента $p = 0{,}769054$, а по критерию Фишера $p = 0{,}684868$.

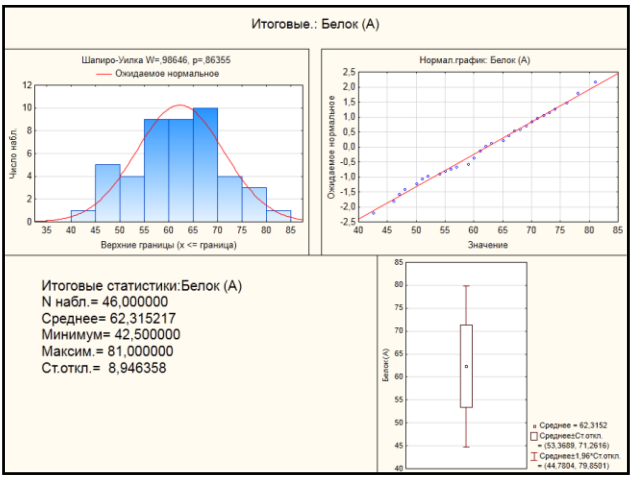
Диаграммы. Быстрый → Дополнительно. Диаграммы размаха (задаем: среднее ±σ и среднее ±1,96σ).

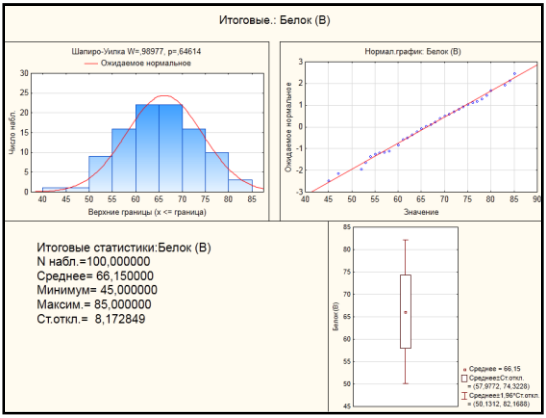


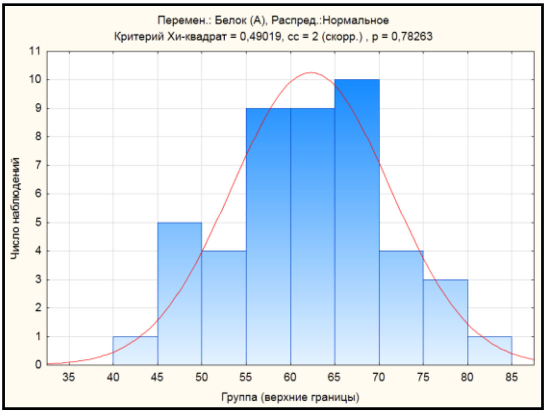
Вывод: рост пациентов мужского и женского пола статистически не различается на уровне значимости 0,05.

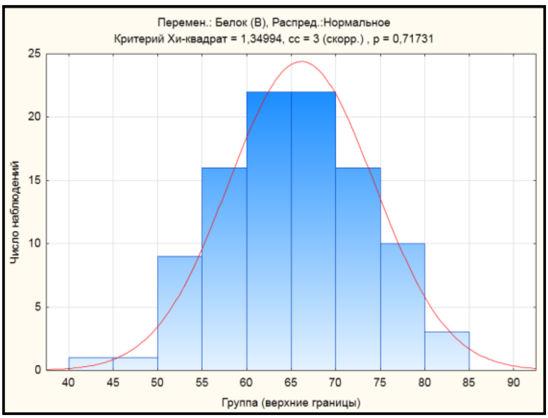
Анализ крови (общий белок) пациентов, прооперированных по методике А (46 чел.) и В (100 чел.). Данные взяты из таблицы № 7 приложения.

Проверка нормальности




Итоговые.: Белок (В)
Шапиро-Уилка W=,98977, p=,64614
Ожидаемое нормальное
Число набл.
Верхние границы (x <= граница)
Нормал.график: Белок (В)
Ожидаемое нормальное
Значение
Итоговые статистики:Белок (В)
N набл.=100,000000
Среднее= 66,150000
Минимум= 45,000000
Максим.= 85,000000
Ст.откл.= 8,172849
Белок (В)
Среднее = 66,15
Среднее±Ст.откл.
= (57,9772, 74,3228)
Среднее±1,96*Ст.откл.
= (50,1312, 82,1688)


Перемен.: Белок (А), Распред.:Нормальное
Критерий Хи-квадрат = 0,49019, сс = 2 (скорр.) , p = 0,78263
Число наблюдений
Группа (верхние границы)


Перемен.: Белок (В), Распред.:Нормальное
Критерий Хи-квадрат = 1,34994, сс = 3 (скорр.) , p = 0,71731
Число наблюдений
Группа (верхние границы)

По критерию Шапиро-Уилка в обоих случаях вероятности нулевой гипотезы (0,86 и 0,65) превышают значение 0,05. Вывод: для обеих переменных принимается гипотеза о нормальности.

По критерию хи-квадрат в обоих случаях вероятности нулевой гипотезы (0,78 и 0,71) превышают значение 0,05. Вывод: для обеих переменных принимается гипотеза о нормальности.

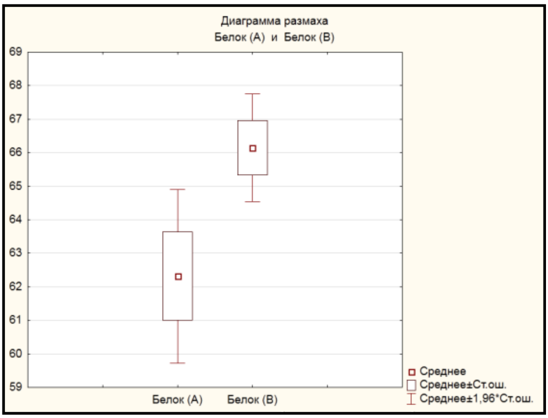
Гипотеза о нормальности принимается по данным обоих критериев.

Анализ. Основные статистики и таблицы, t- критерий для независимых выборок:

	Т-критерий независимых выборок (Таблица данных1) Замечание: Переменные рассм. как независимые выборки							
Группа 1 и Группа 2	**Среднее Группа 1**	Среднее Группа 2	t-знач.	сс	p	N набл. Группа 1	N набл. Группа 2	Ст.откл. Группа 1	Ст.откл. Группа 2
Белок (А) vs. Белок (В)	62,31522	66,15000	-2,55574	144	0,011633	46	100	8,946358	8,172849

Нулевая гипотеза (H_0) — различия нет. Альтернативная гипотеза (H_1) — различие есть. Числа в таблице выделены красным — установлено статистически значимое различие между переменными. Вероятность нулевой гипотезы по критерию Стьюдента $p = 0{,}011633$.

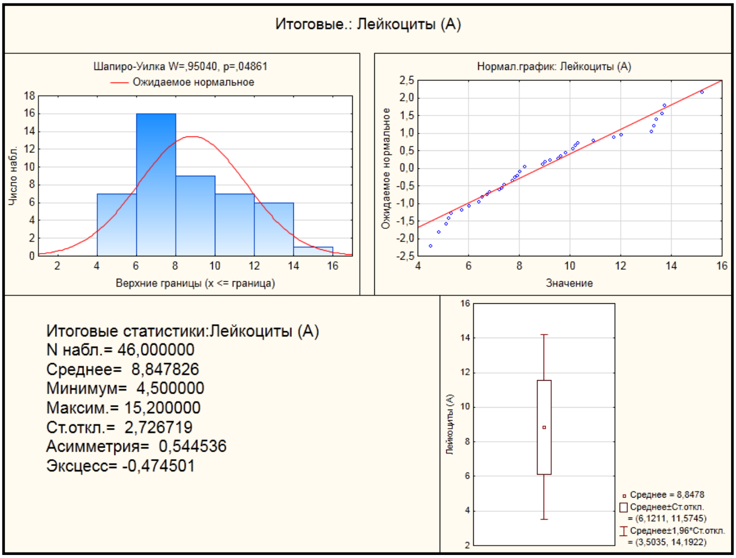
Диаграммы. Быстрый → Дополнительно. Диаграммы размаха.

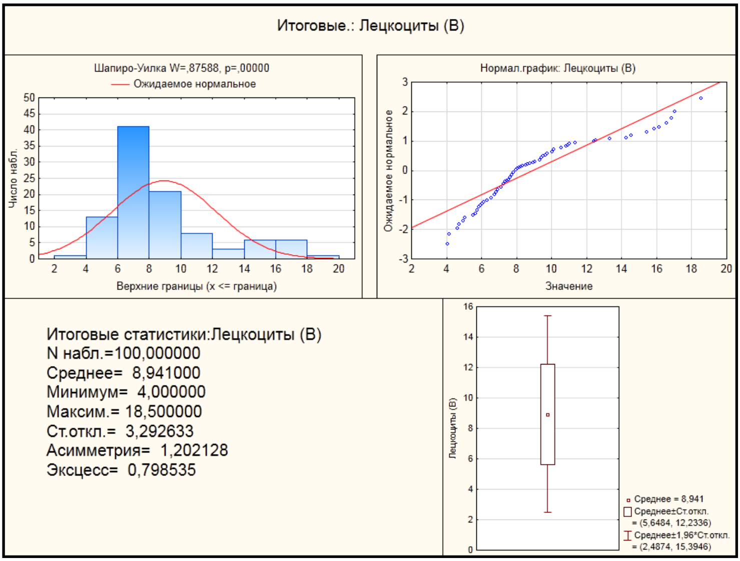


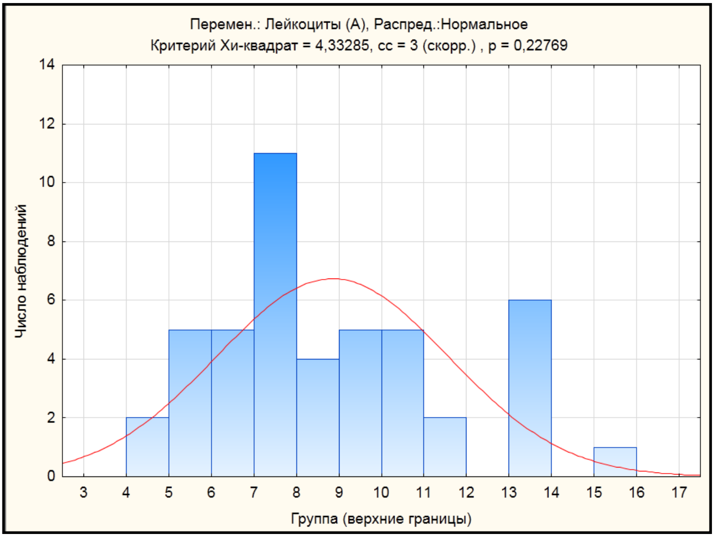
Вывод: общий белок крови у пациентов мужского и женского пола, прооперированных по методике и А (46 чел.), и В (100 чел.), статистически различается на уровне значимости 0,05.

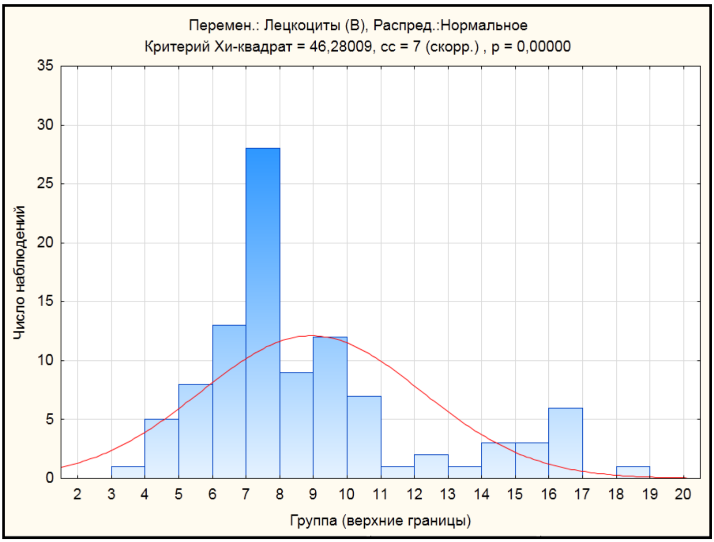
Анализ крови (лейкоциты) пациентов, прооперированных по методике А (46 чел.) и В (100 чел.). Данные взяты из таблицы № 7 приложения.

Проверка нормальности.









По критерию Шапиро-Уилка в обоих случаях вероятности нулевой гипотезы меньше, чем 0,05. Вывод: для обеих переменных гипотеза о нормальности отвергается.

По критерию хи-квадрат в случае А вероятность нулевой гипотезы 0,23, а в случае В — меньше, чем 0,05. Однако при объеме выборки меньше, чем 50, мы отдаем предпочтение критерию Шапиро-Уилка.

Вывод: для обеих переменных гипотеза о нормальности отвергается.

Анализ. Основные статистики и таблицы. t- критерий для независимых выборок:

	Т-критерий независимых выборок (Таблица) Замечание: Переменные рассм. как независимые выборки								
Группа 1 и Группа 2	**Среднее Группа 1**	Среднее Группа 2	t-знач.	сс	p	N набл. Группа 1	N набл. Группа 2	Ст.откл. Группа 1	Ст.откл. Группа 2
Лейкоциты (А) vs. Лецкоциты (В)	8,847826	8,941000	-0,167261	144	0,867399	46	100	2,726719	3,292633

Нулевая гипотеза (H_0) — различия нет. Альтернативная гипотеза (H_1) — различие есть. Числа в таблице не выделены красным — установлено отсутствие статистически значимого различия между переменными. Вероятность нулевой гипотезы по критерию Стьюдента $p = 0{,}877399$.

Диаграммы. Быстрый → Дополнительно. Диаграммы размаха

Вывод: анализы крови (лейкоциты) пациентов, прооперированных по методике А (46 чел.) и В (100 чел.), статистически не различаются на уровне значимости 0,05.

Поскольку данные не подчиняются нормальному закону, следует проверить вывод с помощью непараметрического критерия Манна-Уитни.

Вопросы для обсуждения:

— сравнение различия независимых переменных по величине математического ожидания и стандартного отклонения;

— расчет критического значения критериев Стьюдента и Фишера с использованием вероятностного калькулятора.

Занятие № 11. Критерии Стьюдента и Фишера для независимых выборок

А. Папка «Heartdisease»: Зависимая переменная (2) — систолическое артериальное давление, группирующая переменная (17) — наличие ишемической болезни сердца (ИБС).

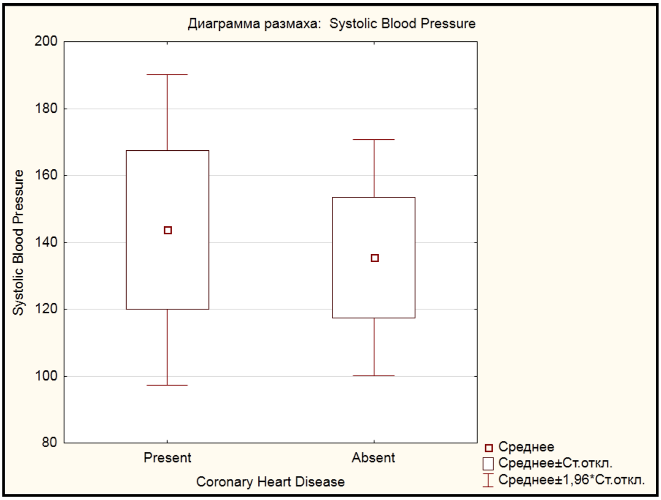
В табл. 2 приложения дана таблица из папки примеров программы STATISTICA в урезанном формате. На самом деле в таблице, кроме столбцов в числовом формате, приведены столбцы с качественными текстовыми метками. Например, столбец «давление» также продублирован столбцом «давление» с метками: «высокое», «среднее», «низкое». Ниже из цитируемой таблицы в качестве группирующей переменной взяты: наличие ИБС, уровень потребления алкоголя и уровень потребления табака.

Анализ. Основные статистики и таблицы. t- критерий для независимых выборок:

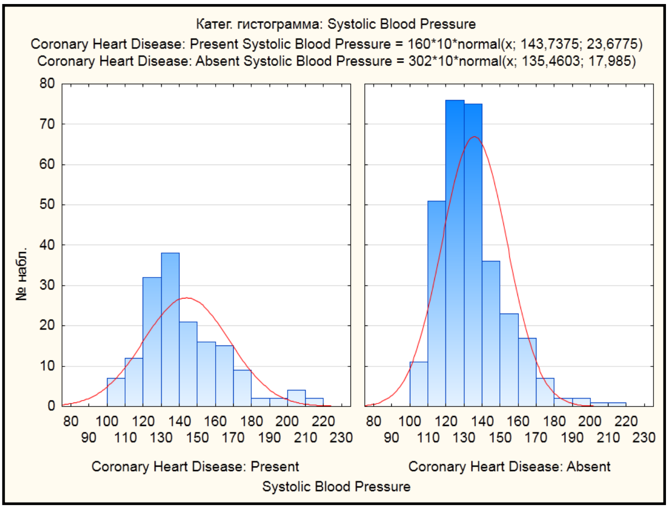
	Т-критерии; Группир.: Coronary Heart Disease (HeartDisease) Группа 1:Present Группа 2:Absent										
Переменная	**Среднее Present**	Среднее Absent	t-знач.	сс	p	N набл. Present	N набл. Absent	Ст.откл. Present	Ст.откл. Absent	F-отн. дисперс.	p дисперс.
Systolic Blood Pressure	143,7375	135,4603	4,204044	460	0,000032	160	302	23,67747	17,98495	1,733213	0,000048

Нулевая гипотеза (H_0) — различия нет. Альтернативная гипотеза (H_1) — различие есть. Числа в таблице выделены красным — установлено статистически значимое различие. Вероятность нулевой гипотезы $p = 0{,}000032$.

Диаграммы. Быстрый → Дополнительно. Диаграммы размаха (задаем: среднее ±σ и среднее ±1,96σ)



Категоризированные гистограммы: в данном подходе это возможность качественно по виду гистограмм оценить соответствие данных нормальному закону.



Вывод: систолическое давление у пациентов с ИБС выше, чем у пациентов без наличия ИБС. Различие статистически значимо на уровне 0,05.

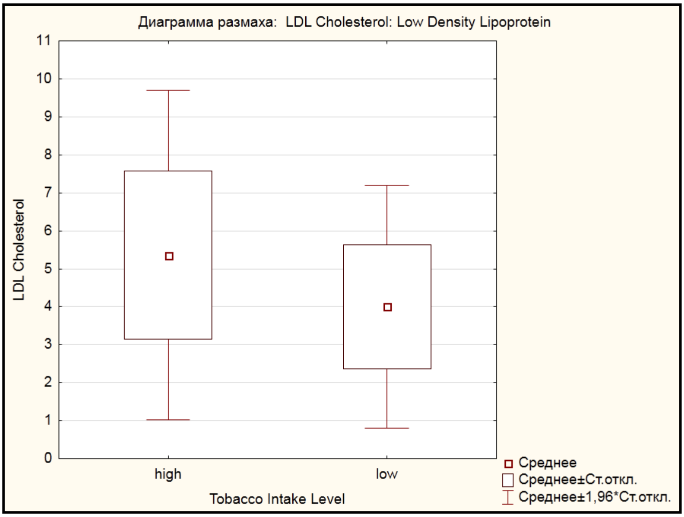
В. Папка «Heartdisease»: Зависимая переменная (6) — ЛПНП холестерол, группирующая переменная (3) — уровень потребления табака.

Анализ. Основные статистики и таблицы. t- критерий для независимых выборок:

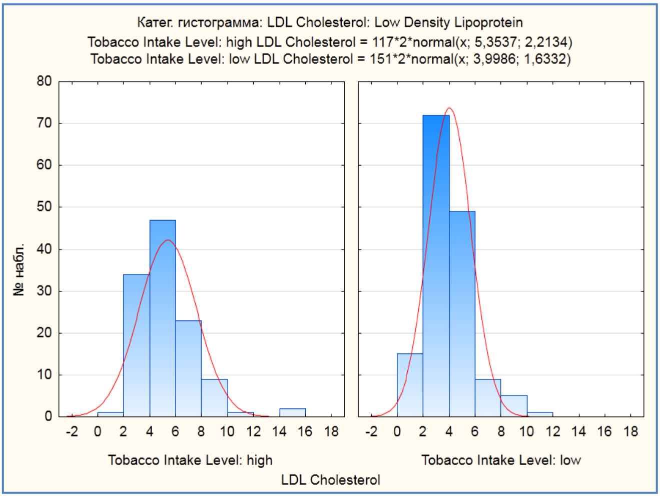
	Т-критерии; Группир.: Tobacco Intake Level: New (HeartDisease) Группа 1:high Группа 2:low										
Переменная	**Среднее high**	Среднее low	t-знач.	сс	p	N набл. high	N набл. low	Ст.откл. high	Ст.откл. low	F-отн. дисперс.	p дисперс.
LDL Cholesterol	5,353675	3,998609	5,766132	266	0,000000	117	151	2,213399	1,633227	1,836649	0,000472

Нулевая гипотеза (H_0) — различия нет. Альтернативная гипотеза (H_1) — различие есть. Числа в таблице выделены красным — установлено статистически значимое различие. Вероятность нулевой гипотезы $p = 0{,}000000$.

Диаграммы. Быстрый → Дополнительно. Диаграммы размаха (задаем: среднее ±σ и среднее ±1,96σ)



Категоризированные гистограммы позволяют качественно по виду гистограмм оценить соответствие данных нормальному закону.



Вывод: содержание холестерина у пациентов с высоким уровнем потребления табака выше, чем у пациентов с низким уровнем потребления. Различие статистически значимо на уровне 0,05.

С. Папка «Heartdisease»: Зависимая переменная (6) — ЛПНП холестерол, группирующая переменная (15) — возраст.

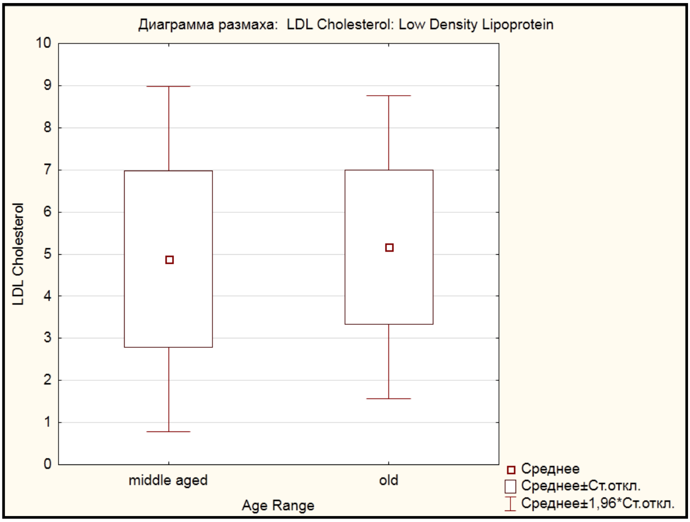
Анализ. Основные статистики и таблицы, t- критерий для независимых выборок:

Т-критерии; Группир.: Age Range: New (HeartDisease)
Группа 1:middle aged
Группа 2:old

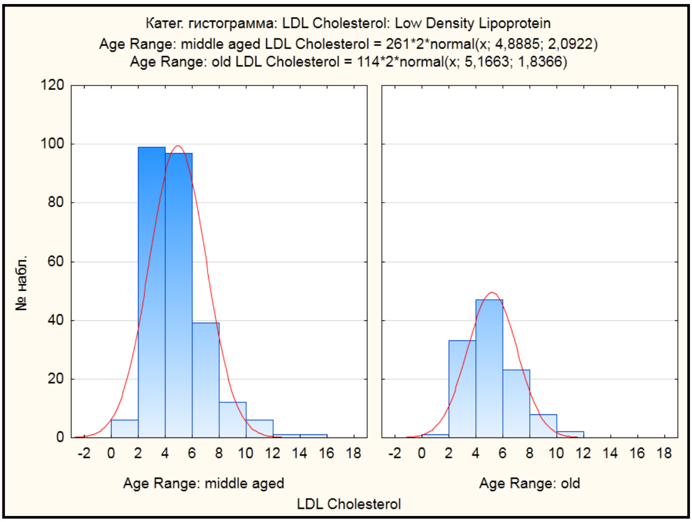
Переменная	Среднее middle aged	Среднее old	t-знач.	cc	p	N набл. middle aged	N набл. old	Ст.откл. middle aged	Ст.откл. old	F-отн. дисперс.	p дисперс.
LDL Cholesterol	4,888506	5,166316	-1,22615	373	0,220915	261	114	2,092180	1,836645	1,297621	0,113556

Нулевая гипотеза (H_0) — различия нет. Альтернативная гипотеза (H_1) — различие есть. Числа в таблице не выделены красным — статистически значимого различия не установлено. Вероятность нулевой гипотезы $p = 0{,}220915$.

Диаграммы. Быстрый → Дополнительно. Диаграммы размаха (задаем: среднее ±σ и среднее ±1,96σ)



Категоризированные гистограммы позволяют качественно по виду гистограмм оценить соответствие данных нормальному закону.



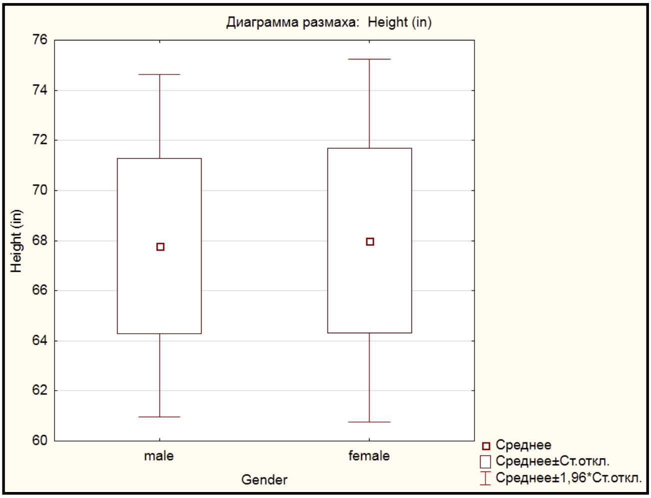
Вывод: содержание холестерина у пациентов среднего и пожилого возраста статистически не различается на уровне значимости 0,05.

D. Папка «Characteristics». Зависимая переменная (4 — рост в дюймах), группирующая переменная (1 — пол).

Анализ. Основные статистики и таблицы, t- критерий для независимых выборок:

	Т-критерии; Группир.: Gender (Characteristics) Группа 1:male Группа 2:female										
Переменная	**Среднее male**	Среднее female	t-знач.	сс	p	N набл. male	N набл. female	Ст.откл. male	Ст.откл. female	F-отн. дисперс.	p дисперс.
Height (in)	67,78846	68,00000	-0,294428	98	0,769054	52	48	3,488561	3,695886	1,122392	0,684868

Быстрый → Дополнительно. Диаграммы размаха (задаем: среднее ±σ и среднее ±1,96σ).



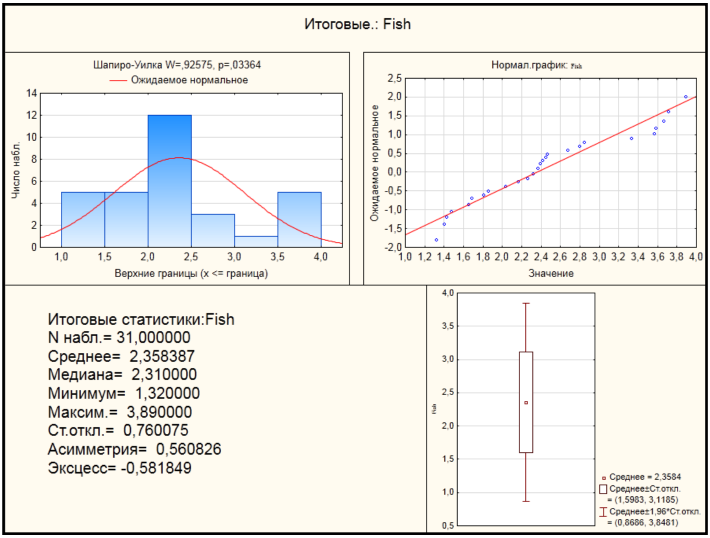
Вывод: рост пациентов мужского и женского пола статистически не различается на уровне значимости 0,05.

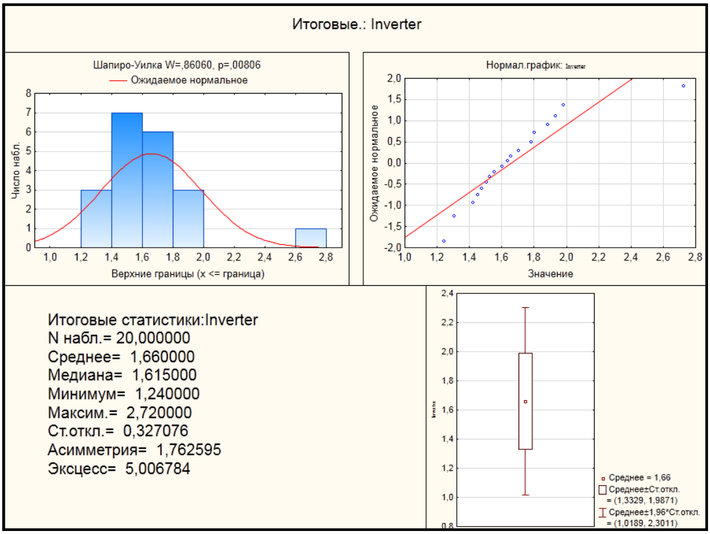
Вопросы для обсуждения:

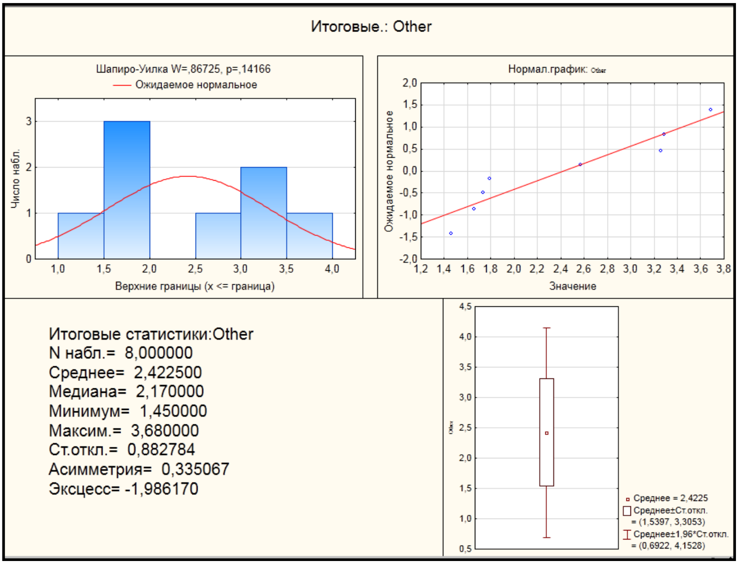
— сравнение различия независимых выборок по величине математического ожидания и стандартного отклонения;

— расчет критического значения критериев Стьюдента и Фишера с использованием вероятностного калькулятора.

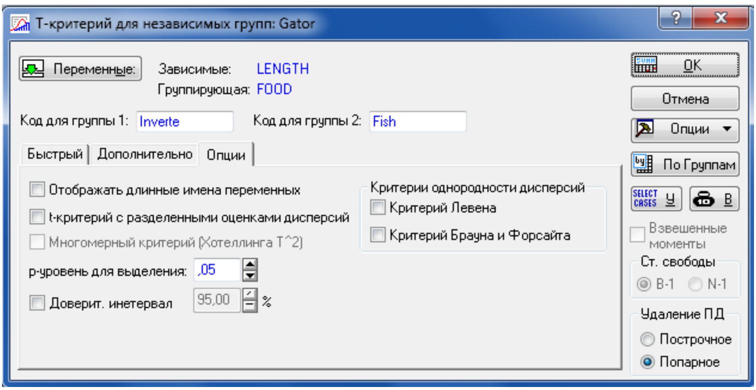
Папка «Gator» (Аллигаторы). Таблица № 3

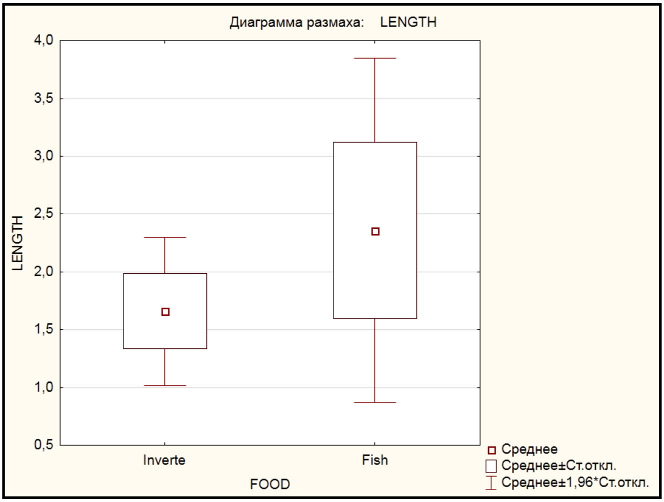






Гипотеза о нормальности отвергается, но начнем мы с критерия Стьюдента.





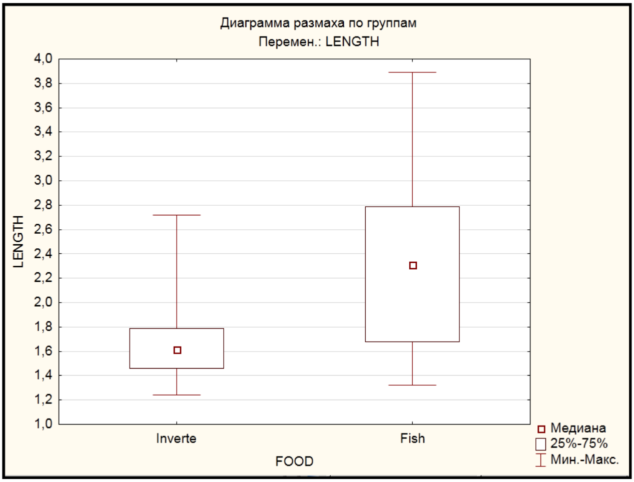
	Т-критерии; Группир.: FOOD (Gator) Группа 1:Inverte: Invertebrates Группа 2:Fish: Fish										
Переменная	**Среднее Inverte**	Среднее Fish	t-знач.	cc	p	N набл. Inverte	N набл. Fish	Ст.откл. Inverte	Ст.откл. Fish	F-отн. дисперс.	p дисперс.
LENGTH	1,660000	2,358387	-3,87353	49	0,000318	20	31	0,327076	0,760075	5,400259	0,000319

Теперь смотрим, что дают непараметрические критерии.

	Критерий серий Вальда-Вольфовица (Gator) По перем. FOOD Отмеченные критерии значимы на уровне p <,05000									
Перем.	**N Inverte**	N Fish	Среднее Inverte	Среднее Fish	Z	p-уров.	Z скорр.	p-уров.	Число серий	Число совп.
LENGTH	20	31	1,660000	2,358387	-2,17216	0,029844	2,023662	0,043006	18	4

	Критерий Колмогорова-Смирнова (Gator) По перем. FOOD Отмеченные критерии значимы на уровне p <,05000								
Перем.	**Макс.отр Разн.**	Макс. по Разн.	p-уров.	Среднее Inverte	Среднее Fish	Ст.откл. Inverte	Ст.откл. Fish	N Inverte	N Fish
LENGTH	-0,627419	0,00	p < .001	1,660000	2,358387	0,327076	0,760075	20	31

	U критерий Манна-Уитни (Gator) По перем. FOOD Отмеченные критерии значимы на уровне p <,05000									
Перем.	**Сум.ранг Inverte**	Сум.ранг Fish	U	Z	p-уров.	Z скорр.	p-уров.	N Inverte	N Fish	2-х стор точное p
LENGTH	338,5000	987,5000	128,5000	-3,49198	0,000480	-3,49301	0,000478	20	31	0,000292



Вывод: длина аллигаторов, которые питаются рыбой, статистически значимо выше.

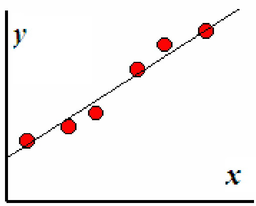
Как видно, все 4 критерия дают возможность сделать один и тот же вывод. Заметим, что ряд авторов указывает на то, что в особо ответственных задачах прикладной статистики следует использовать не один, а несколько критериев.

Занятие № 12. Метод наименьших квадратов

Метод наименьших квадратов основан на применении производной для построения оптимальной линии через ряд точек на графике, в частности — в задаче о корреляции. Линия может быть прямой либо кривой, заданной какой-либо элементарной функцией. Алгоритм схож. Мы рассмотрим метод в применении к прямой линии.

Пусть x — независимая, а y — зависимая переменная:

Задача. По имеющимся наблюдениям — $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ — найти коэффициенты a и b в уравнении линейной регрессии $y = ax + b$.



$y_i = ax_i + b + \varepsilon_i$, где ε_i— невязка (расхождение). Решается задача на нахождение минимума суммы:

$$T = \sum_{i=1}^{n} [y_i - (ax_i + b)^2].$$

Нахождение частных производных по a и b позволяет получить систему двух

уравнений с двумя неизвестными:

$$\begin{cases} \dfrac{\partial T}{\partial a} = -2\sum\limits_{i=1}^{n}(y_i - ax_i)\cdot x_i = 0, \\ \dfrac{\partial T}{\partial b} = -2\sum\limits_{i=1}^{n}(y_i - ax_i - b) = 0. \end{cases}$$

После сокращения получим:

$$\begin{cases} \sum\limits_{i=1}^{n}(y_i - ax_i)\cdot x_i = 0, \\ \sum\limits_{i=1}^{n}(y_i - ax_i - b) = 0. \end{cases}$$

Перепишем уравнения в виде:

$$\begin{cases} \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i y_i - a\dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^2 - b\dfrac{1}{n}\sum\limits_{i=1}^{n} x_i = 0, \\ \dfrac{1}{n}\sum\limits_{i=1}^{n} y_i - a\dfrac{1}{n}\sum\limits_{i=1}^{n} x_i - b = 0. \end{cases}$$

Решение системы дает значения a и b:

$$a = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\langle x^2\rangle - \langle x\rangle^2}, b = \langle y\rangle - a\langle x\rangle,$$

где использованы обозначения:

$$\langle x\rangle = \frac{1}{n}\sum_{i=1}^{n} x_i;\ \langle y\rangle = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

$$\langle x^2\rangle = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \text{ и } \langle xy\rangle = \frac{1}{n}\sum_{i=1}^{n} x_i y_i.$$

Вопросы для обсуждения:

- метод наименьших квадратов для нахождения значений параметров уравнения регрессии;

- построение графика уравнения регрессии.

Занятие № 13. Коэффициент линейной корреляции Пирсона

Краткая теория

1. Проверка нормальности (критерий Пирсона применяется только для нормально распределенных случайных величин).

2. Коэффициент корреляции:

$$\boldsymbol{r} = \frac{\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{y_i} - \overline{\boldsymbol{y}})}{\sqrt{\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2 \sum(\boldsymbol{y_i} - \overline{\boldsymbol{y}})^2}}.$$

3. Статистика критерия:

$$\boldsymbol{T} = \frac{\boldsymbol{r}\sqrt{\boldsymbol{n}-\boldsymbol{2}}}{\sqrt{\boldsymbol{1}-\boldsymbol{r}^2}}.$$

Если гипотеза H_0 (по которой корреляции нет) верна, тогда T имеет распределение Стьюдента с $n - 2$ степенями свободы. На уровне значимости α берем значение $t_{кр.}(\alpha, n - 2)$ для двусторонней области:

Если $T < t_{кр.}$ → принимается гипотеза H_0— корреляции нет.

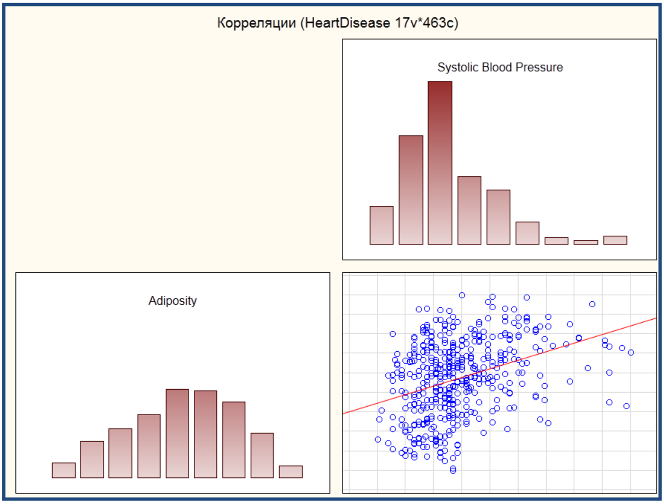
Если $T > t_{кр.}$ → принимается гипотеза H_1— корреляции есть.

Практическое занятие:

А. Папка: **«Heart Disease».**

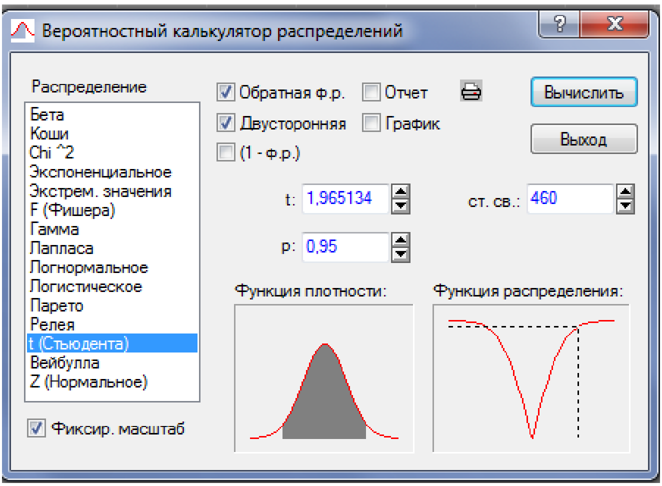
(X): *Systolic Blood Pressure*. (Y): *Adiposity* (уровень жира в тканях).

	Корреляции (HeartDisease) Отмеченные корреляции значимы на уровне p <,05000 N=462 (Построчное удаление ПД)
Переменная	**Adiposity**
Systolic Blood Pressure	0,356500

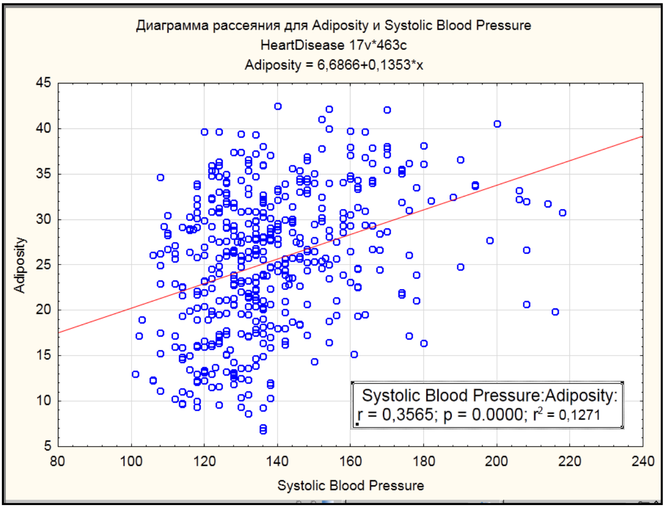


Проверка значимости коэффициента корреляции: $T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = 8{,}3$.

Критическое значение коэффициента Стьюдента ($t_{\text{кр.}}$) можно найти с помощью вероятностного калькулятора.

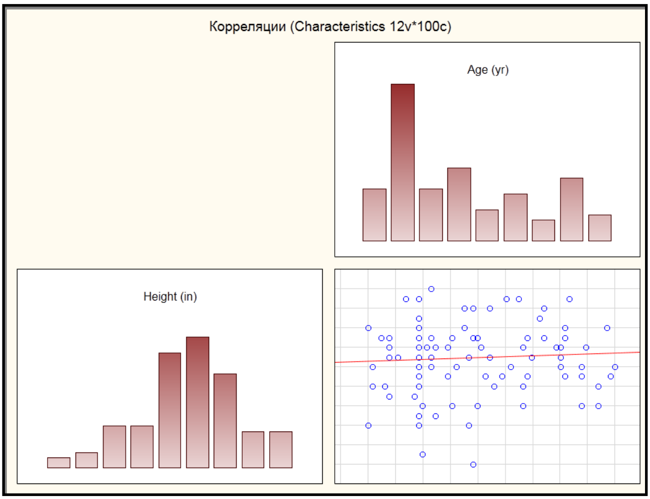


$t_{\text{кр.}} = 1{,}96$. Поскольку $T > t_{\text{кр.}}$, принимается гипотеза H_1— корреляция есть. Коэффициент корреляции r = 0,36. Вывод: установлена средняя по величине, положительная, статистически значимая корреляция между переменными на уровне 0,05.

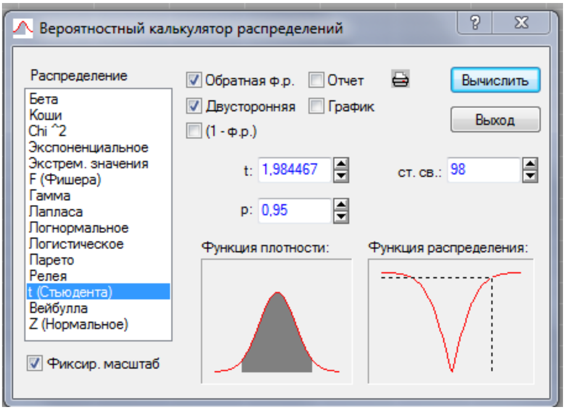


А. Папка **«Characteristics».** (X): возраст, год. (Y): рост, дюймы.

	Корреляции (Characteristics) Отмеченные корреляции значимы на уровне p <,05000 N=100 (Построчное удаление ПД)
Переменная	**Height (in)**
Age (yr)	0,061739



Проверка значимости коэффициента корреляции: $T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = 0{,}61$.



$t_{кр.} = 1{,}98$. Поскольку $T < t_{кр.}$, мы принимаем гипотезу H_0 — корреляции нет. Коэффициент корреляции $r = 0{,}062$.

Вопросы для обсуждения:

— метод наименьших квадратов для нахождения значений параметров уравнения регрессии;

— построение графика уравнения регрессии.

Занятие № 14. Непараметрические критерии в STATISTICA

Краткая теория

Использование параметрических критериев статистики без предварительной проверки вида распределения может привести к определенным ошибкам в ходе проверки рабочей гипотезы.

Для преодоления указанных трудностей в практике исследований следует использовать **непараметрические критерии статистики**, такие как критерий знаков, критерий Манна-Уитни, критерий Спирмена и др. Непараметрические критерии статистики свободны от допущения о законе распределения выборок и базируются на предположении о независимости наблюдений.

Критерий знаков (G-критерий)

Критерий знаков — это непараметрический критерий, который основан на оценке разности попарно сопряженных вариантов (например, до и после лечения). Учитывается не величина, а направленность сдвигов.

Применение критерия знаков не зависит от характера распределения данных. Изменения оценивают в альтернативной форме (увеличение-уменьшение и т.п., что обозначают знаками + или –, откуда и произошло название критерия). Случаи, когда парные наблюдения не имеют разницы, в расчет не принимаются. Следует стремиться, чтобы количество нулевых разностей было минимальным. Для этого необходимо повышать точность измерения показателей, что обеспечивает непрерывность выборочных данных.

Практическое применение критерия знаков включает следующие этапы:

1. Определяется направленность изменений в сравниваемых наблюдениях.
2. Подсчитывается общее число парных наблюдений, имеющих различия (**n**).
3. Подсчитывается меньшее число однозначных результатов сравнения, обзначаемых как **Z**.
4. **Z** сравнивается по специальной таблице с критическими значениями для данного **n**.

Критерий знаков приложим как к совокупностям непрерывных признаков, так и для оценки различия полуколичественных признаков (баллы и т.п.) при достаточном числе их градаций.

Нулевая гипотеза формулируется следующим образом: в состоянии изучаемого свойства нет значимых различий при первичном и вторичном измерениях. **Альтернативная гипотеза**: состояния изучаемого свойства значимо различны в одной и той же совокупности при первичном и вторичном измерениях этого свойства.

Нулевая гипотеза принимается на уровне значимости 0,05, если наблюдаемое значение $Z < Z_{\text{кр.}}$, где значение $Z_{\text{кр.}}$, определяется из статистических таблиц для критерия знаков.

Критерий Манна-Уитни

U-критерий Манна-Уитни — непараметрический статистический критерий, используемый для сравнения двух независимых выборок по уровню какого-либо признака, измеренного количественно. Метод основан на определении того, достаточно ли мала зона перекрещивающихся значений между двумя вариационными рядами (ранжированным рядом значений параметра в первой выборке и таким же во второй выборке). Чем меньше значение критерия, тем вероятнее, что различия между значениями параметра в выборках достоверны.

Расчет критерия Манна-Уитни

Сначала из обеих сравниваемых выборок составляется единый ранжированный ряд, путем ранжирования единиц наблюдения по степени возрастания признака и присвоения меньшему значению меньшего ранга. В случае равных значений признака у нескольких единиц каждой из них присваивается среднее арифметическое последовательных значений рангов.

Например, две единицы, занимающие в едином ранжированном ряду 2 и 3 место (ранг), имеют одинаковые значения. Следовательно, каждой из них присваивается ранг, равный (3 + 2)/2 = 2,5.

В составленном едином ранжированном ряду общее количество рангов получится равным: $N = n_1 + n_2$, где n_1 — количество элементов в первой выборке, а n_2 — количество элементов во второй выборке.

Далее вновь разделяем единый ранжированный ряд на два, состоящие, соответственно, из единиц первой и второй выборок, запоминая при этом значения рангов для каждой единицы. Подсчитываем отдельно сумму рангов, пришедшихся на долю элементов первой выборки, и отдельно — на долю элементов второй выборки. Определяем большую из двух ранговых сумм (T_x) соответствующую выборке с n_x элементами, n_x — количество испытуемых в группе с большей ранговой суммой.

Наконец, находим значение U-критерия Манна-Уитни по формуле:

$$U = n_1 \cdot n_2 + \frac{n_x \cdot (n_x + 1)}{2} - T_x$$

Полученное значение U-критерия сравниваем по таблице для избранного уровня статистической значимости ($p=0{,}05$ или $p=0{,}01$) с критическим значением U при заданной численности сопоставляемых выборок:

Если полученное значение U **меньше табличного или равно** ему, то признается статистическая значимость различий между уровнями признака в рассматриваемых выборках (принимается альтернативная гипотеза). Достоверность различий тем выше, чем меньше значение U.

Если же полученное значение U **больше табличного**, принимается нулевая гипотеза об отсутствии статистически значимых различий между выборками.

Коэффициент корреляции Спирмена

Коэффициент ранговой корреляции Спирмена — это непараметрический метод, который используется с целью статистического изучения связи между случайными величинами. В этом случае определяется фактическая степень параллелизма между двумя количественными рядами изучаемых признаков и дается оценка тесноты установленной связи с помощью количественно выраженного коэффициента.

Коэффициент ранговой корреляции Спирмена используется для выявления и оценки тесноты связи между двумя рядами сопоставляемых количественных показателей. В том случае, если ранги показателей, упорядоченных по степени возрастания или убывания, в большинстве случаев совпадают (большему значению одного показателя соответствует большее значение другого показателя — например, при сопоставлении роста пациента и его массы тела), делается вывод о наличии **прямой** корреляционной связи. Если ранги показателей имеют противоположную направленность (большему значению одного показателя соответствует меньшее значение другого — например, при сопоставлении возраста и частоты сердечных сокращений), то говорят об **обратной** связи между показателями.

Коэффициент корреляции Спирмена обладает следующими свойствами:

1. Коэффициент корреляции может принимать значения от минус единицы до плюс единицы, причем при $r(S) = 1$ имеет место строго прямая связь, а при $r(S) = -1$ — строго обратная связь.
2. Если коэффициент корреляции отрицательный, то имеет место обратная связь, если положительный — то прямая связь.
3. Если коэффициент корреляции равен нулю, то связь между величинами практически отсутствует.
4. Чем ближе модуль коэффициента корреляции к единице, тем более сильной является связь между измеряемыми величинами.

Расчет коэффициента ранговой корреляции Спирмена включает следующие этапы:

— сопоставить каждому из признаков их порядковый номер (ранг) по возрастанию или убыванию;

— определить разности рангов каждой пары сопоставляемых значений (d);

— возвести в квадрат каждую разность и суммировать полученные результаты;

— вычислить коэффициент корреляции рангов по формуле:

$$r = 1 - \frac{6 \cdot \sum d^2}{n(n^2 - n)};$$

— определить **статистическую значимость** коэффициента при помощи **t-критерия**, рассчитанного по формуле:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

При использовании коэффициента ранговой корреляции условно оценивают тесноту связи между признаками, считая значения коэффициента, равные **0,3 и менее**, показателями **слабой тесноты связи**; значения **более 0,3, но менее 0,7** — показателями **средней тесноты связи**, а значения **0,7 и более** — показателями **высокой тесноты связи.**

Статистическая значимость полученного коэффициента оценивается при помощи t-критерия Стьюдента. Если рассчитанное значение t-критерия меньше табличного при заданном числе степеней свободы, статистическая значимость наблюдаемой взаимосвязи отсутствует. Если больше, то корреляционная связь считается статистически значимой.

Практическое занятие

А. Ранее (занятие №9) мы сравнивали показатели здоровья с применением параметрического критерия Стьюдента (t). Было установлено, что для обоих показателей (по критерию хи-квадрат) можно принять гипотезу о нормальном законе распределения.

Теперь проверим выводы работы №9 с помощью непараметрического критерия знаков.

1. Открыть папку примеров «Characteristics» (таблица №4 приложения).
2. Выделить интересующие нас колонки с переменными W1 и W2.
3. Нажать вкладку **Анализ** и выбрать **Непараметрическую статистику.**
4. Выбрать команду **Сравнение двух зависимых переменных.**
5. Выбрать **Критерий знаков**.

6. Результат:

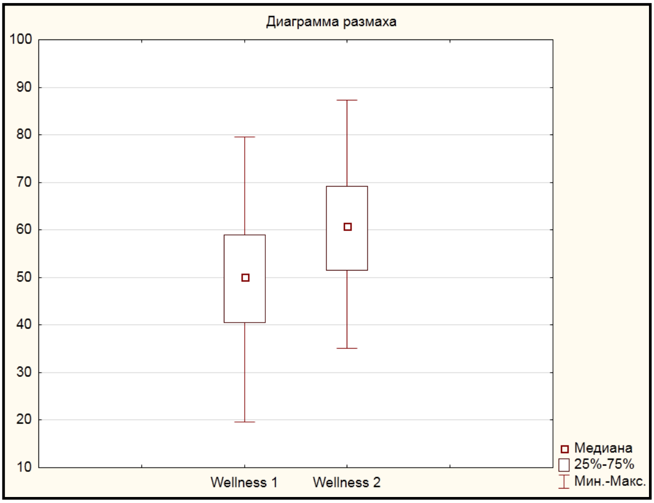
Пара перем.	Критерий знаков (Characteristics) Отмеченные критерии значимы на уровне p <,05000 **Число несовп.**	Процент v < V	Z	p-уров.
Wellness 1 & Wellness 2	100	72,00000	4,300000	0,000017

a) значение критерия знаков (Z) =4,3 выделено красным, что означает, что показатель здоровья W2 статистически значимо выше, чем W1; уровень значимости составляет 0,000017, что существенно ниже заданного по умолчанию уровня 0,05;

b) полученный результат совпадает с результатами работы №10.

Критерий Вилкоксона

Пара перем.	Критерий Вилкоксона (Characteristics) Отмеченные критерии значимы на уровне p <,05000 **Число набл.**	T	Z	p-уров.
Wellness 1 & Wellness 2	100	925,0000	5,501320	0,000000



Результаты использования критерия знаков и критерия Вилкоксона согласуются между собой и с выводом занятия №10, где был использован параметрический t-критерий Стьюдента.

U-критерий Манна-Уитни как критерий сравнения двух независимых групп (папка примеров «Characteristic Height», переменные: «Male Height» — «Female Height»).

В таблице приведено значение U-критерия Манна-Уитни и соответствующее ему значение критерия стандартного нормального распределения (Z). С критическим сравнивается значение Z, которое красным цветом не выделено, вероятность нулевой гипотезы об отсутствии различий p=0,20. Статистически значимого различия роста юношей и девушек на уровне значимости 0,05 не выявлено.

— Открыть папку примеров «Characteristic Height».

— Выделить интересующие нас колонки с переменными «Male Height» и «Female Height».

— Нажать вкладку **Анализ** и выбрать **Непараметрическую статистику.**

— Выбирать команду **Сравнение двух независимых групп.**

— Выбирать **U-критерий Манна-Уитни**.

— Результат:

	U критерий Манна-Уитни (CharacteristicsHeight) По перем. Female Height Отмеченные критерии значимы на уровне p <,05000				
Перем.	**Сум.ранг Группа 1**	Сум.ранг Группа 2	U	Z	p-уров.
Male Height	30,50000	47,50000	9,500000	-1,28103	0,200186

— значение критерия Манна-Уитни U = 9,5 красным цветом не выделено, что означает отсутствие статистически значимых отличий роста юношей и девушек для выбранного уровня значимости 0,05.

Коэффициент корреляции Спирмена как показатель наличия корреляции (папка примеров «Heart Diseases»). Переменные: «систолическое давление» — «холестерин».

В таблице приведены данные о систолическом артериальном давлении у 200 пациентов сердечно-сосудистого отделения госпиталя. В работе №6 мы с помощью критерия хи-квадрат выяснили, что для обеих переменных («давление» и «холестерин») нельзя принять гипотезу о нормальном законе распределения. Поэтому для установления факта наличия либо отсутствия корреляции использование параметрического коэффициента корреляции Пирсона неправомерно.

Применить для проверки гипотезы о наличии корреляции непараметрический критерий Спирмена.

— Открыть папку примеров «Heart Diseases».

— Нажать вкладку **Анализ** и выбрать **Непараметрическую статистику.**

— Выбирать команду **Корреляции Спирмена.**

— Выбирать переменные со второй по шестую (2-я и 6-я переменные соответствуют нашей задаче).

— Выбирать: **Удалить неподходящие переменные.**

— Остаются переменные: «давление» (2), «потребление табака» (4) и «холестерин» (6).

— Выбрать **Коэффициент корреляции Спирмена.**

— Результат:

	Ранговые корреляции Спирмена (HeartDisease) ПД попарно удалены Отмеченные корреляции значимы на уровне p <,05		
Перем.	**Systolic Blood Pressure**	Tobacco Intake (kg)	LDL Cholesterol
Systolic Blood Pressure	1,000000	0,221805	0,189325
Tobacco Intake (kg)	0,221805	1,000000	0,274268
LDL Cholesterol	0,189325	0,274268	1,000000

— значение коэффициента корреляции Спирмена составило 0,189325 и выделено красным цветом как статистически значимая корреляция на уровне 0,05.

Вывод: установлена слабая, положительная, статистически значимая на уровне 0,05 корреляция (взаимосвязь) между переменными «**давление»** и «**холестерин».**

Вопросы для обсуждения:

— критерий знаков;
— критерий Манна-Уитни;
— коэффициент корреляции Спирмена.

Занятие № 15. Таблицы вида 2х2. Критерий хи-квадрат и точный критерий Фишера

Часто возникают ситуации, когда сравнивают две группы значений переменных с качественными признаками, относящимися к номинальной шкале. В таких случаях составляют таблицы частот.

Сравнение групп А и В по полу (см. табл. 7). Предстоит выяснить, есть ли различие в двух группах пациентов, которые лечились по разным методикам, по соотношению лиц мужского и женского пола. Используем непараметрические критерии в применении к таблице 2×2.

	М	Ж	Всего:
Группа А	32	14	32+14=46
Группа В	66	34	66+34=100

Нулевая гипотеза H_0: соотношение мужчин и женщин в группах А и В не различается. Альтернативная гипотеза H_1 : соотношение мужчин и женщин в группах А и В различается.

Критерий хи-квадрат используется в тех же случаях, когда применяется критерий точной вероятности Фишера. χ^2 является приближенным критерием, причем приближение к точным вероятностям Фишера улучшается по мере увеличения суммарного числа пациентов n. χ^2 можно использовать при $n > 30$. Критерий χ^2 не рекомендуется использовать, если ожидаемая частота хотя бы в одной из клеток меньше 5. В нашем случае можно использовать оба критерия, и они дают одинаковый ответ на поставленный вопрос.

Нажимаем вкладку Анализ, далее — Непараметрическая статистика и выбираем первую опцию: Таблицы 2х2 … Макнемара, Фишера. Создаем таблицу 2х2 и заносим в нее данные — количество мужчин и женщин в каждой группе (А и В). На результат не влияет то, как у нас скомпонована таблица, то есть что мы взяли в качестве столбцов и что в качестве строк. Нажимаем ОК и получим следующую таблицу:

	Таблица 2x2 (Таблица данных		
	Столб. 1	Столб. 2	Сумма строки
Частоты, строка 1	32	14	46
Процент от общего	21,918%	9,589%	31,507%
Частоты, строка 2	66	34	100
Процент от общего	45,205%	23,288%	68,493%
Сумма по столбцам	98	48	146
Процент от общего	67,123%	32,877%	
Хи-квадрат (ст.св.=1)	,18	p= ,6701	
V-квадрат (ст.св.=1)	,18	p= ,6712	
С поправкой Йетса	,06	p= ,8131	
Фи коэффициент	,00124		
Фишера p, односторонний		p= ,4097	
двусторонний		p= ,7086	
Макнемара Хи-квадрат (A/D)	,02	p= ,9020	
Хи-квадрат (B/C)	32,51	p= ,0000	

Для этого можно использовать: 1) критерий хи-квадрат с поправкой Йетса. В таблице это p=0,8131; 2) точный критерий Фишера. В таблице это p=0,7086. В обоих случаях p показывает значение вероятности того, что справедлива нулевая гипотеза.

Вывод: соотношение мужчин и женщин в группах А и В статистически не различается.

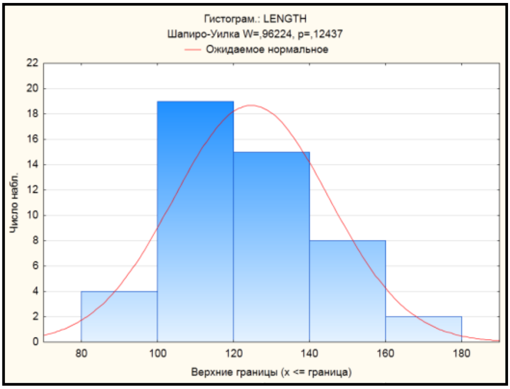
К этому можно добавить, что количество больных мужчин в группе А составляет 70% от общего числа пациентов, а в группе В — 66%.

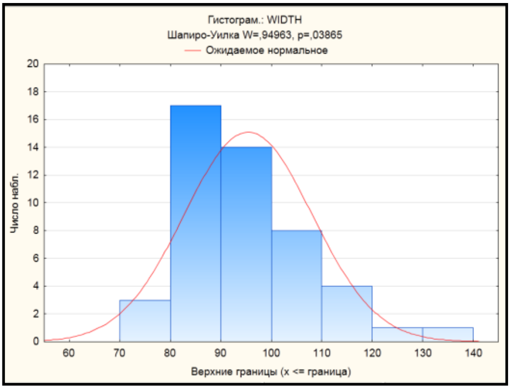
Вопросы для обсуждения:

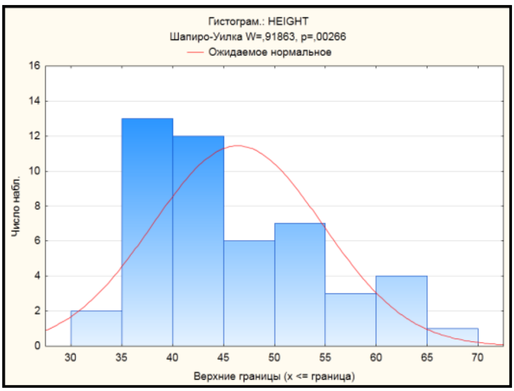
- критерий хи-квадрат;
- точный критерий Фишера.

Занятие № 16. Самостоятельная работа
Папка «Turtles» (Таблица № 6 приложения)

1. Проверка нормальности с помощью критерия Шапиро-Уилка.







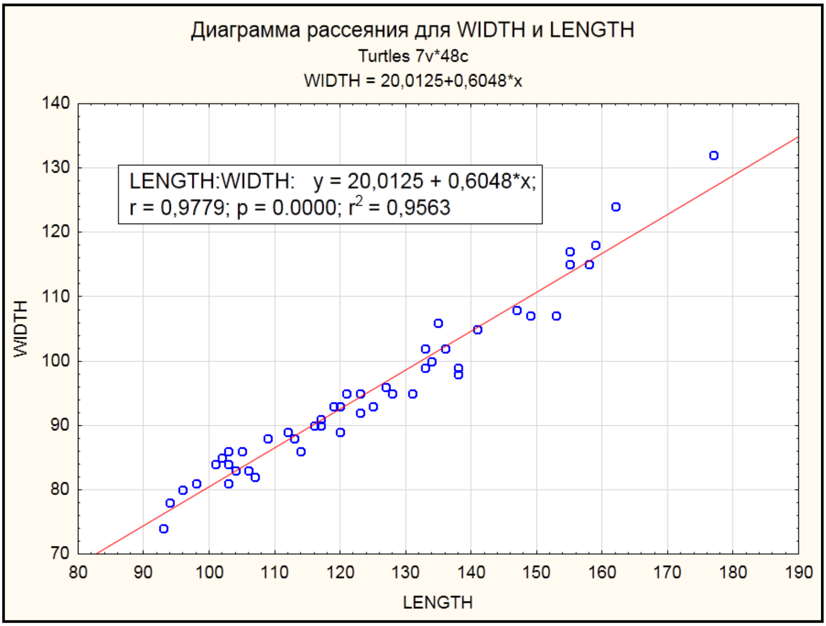
На уровне значимости 0,05 для переменной «длина» принимается гипотеза о нормальности, а для переменных «ширина» и «высота» гипотеза о нормальности отвергается.

2. Проверка наличия корреляции между переменными.
3.

Variable	Correlations (Turtles.sta) Marked correlations are significant at p < ,05000 N=48 (Casewise deletion of missing data) **Means**	Std.Dev.	LENGTH	WIDTH	HEIGHT
LENGTH	124,7083	20,49490	1,000000	0,977887	0,962890
WIDTH	95,4375	12,67584	0,977887	1,000000	0,959905
HEIGHT	46,3750	8,36565	0,962890	0,959905	1,000000

Между каждой из трех переменных имеется положительная, сильная, статистически значимая корреляция (на уровне значимости 0,05).

4. Графики (поле корреляции).



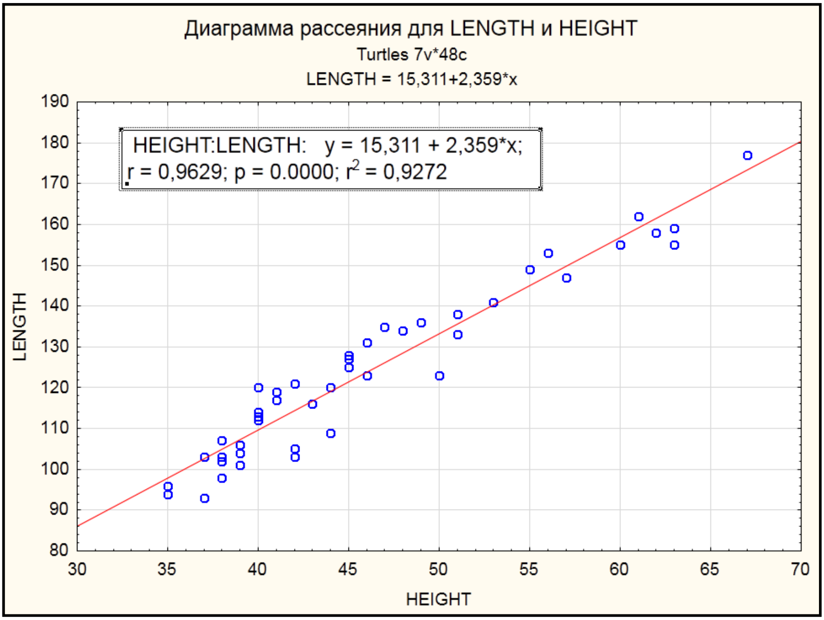


Диаграмма рассеяния для LENGTH и HEIGHT
Turtles 7v*48c
LENGTH = 15,311+2,359*x
HEIGHT:LENGTH: y = 15,311 + 2,359*x;
r = 0,9629; p = 0.0000; r2 = 0,9272
LENGTH
HEIGHT

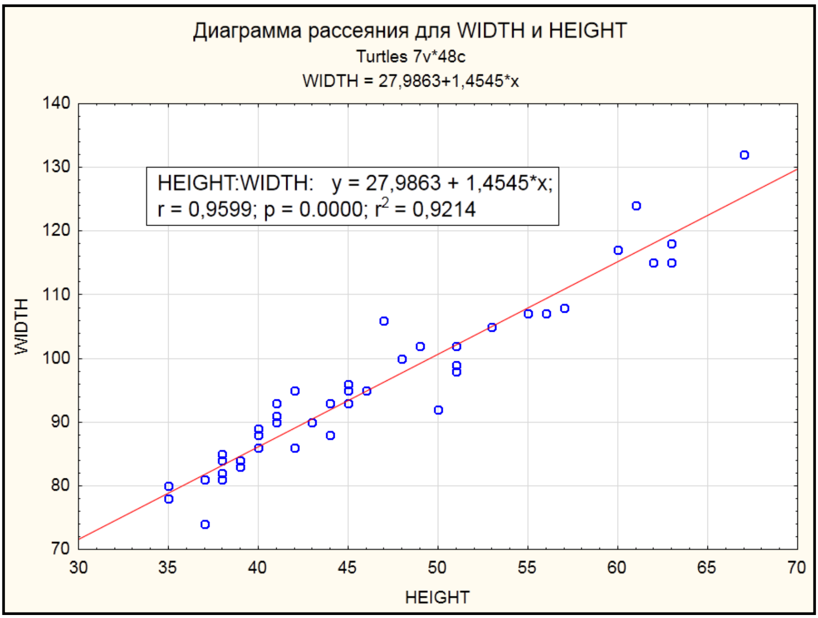


Диаграмма рассеяния для WIDTH и HEIGHT
Turtles 7v*48c
WIDTH = 27,9863+1,4545*x
HEIGHT:WIDTH: y = 27,9863 + 1,4545*x;
r = 0,9599; p = 0.0000; r2 = 0,9214
WIDTH
HEIGHT

5. Корреляция по Спирмену (непараметрический критерий).
6.

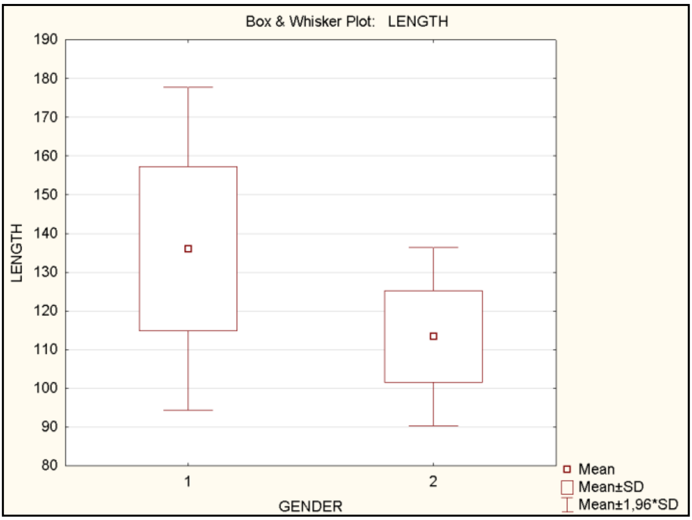
Variable	Spearman Rank Order Correlations (Turtles.sta) MD pairwise deleted Marked correlations are significant at p <,05000 LENGTH	WIDTH	HEIGHT
LENGTH	1,000000	0,977856	0,956709
WIDTH	0,977856	1,000000	0,961095
HEIGHT	0,956709	0,961095	1,000000

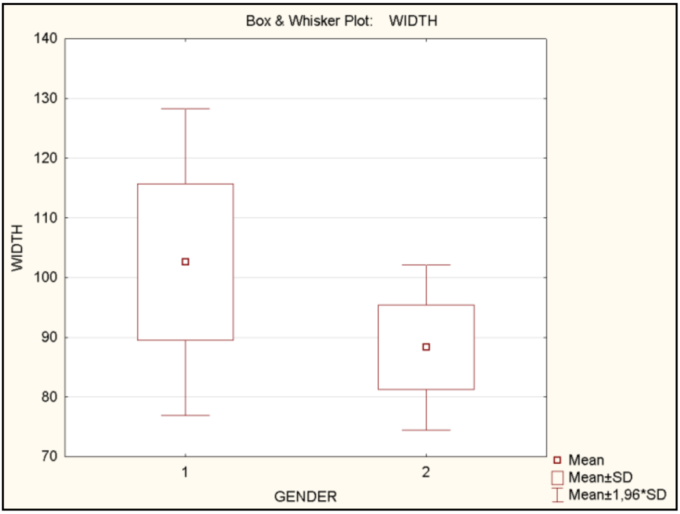
7. Различие длины, ширины и высоты для особей женского (1) и мужского (2) пола (критерии Стьюдента и Фишера).
8.

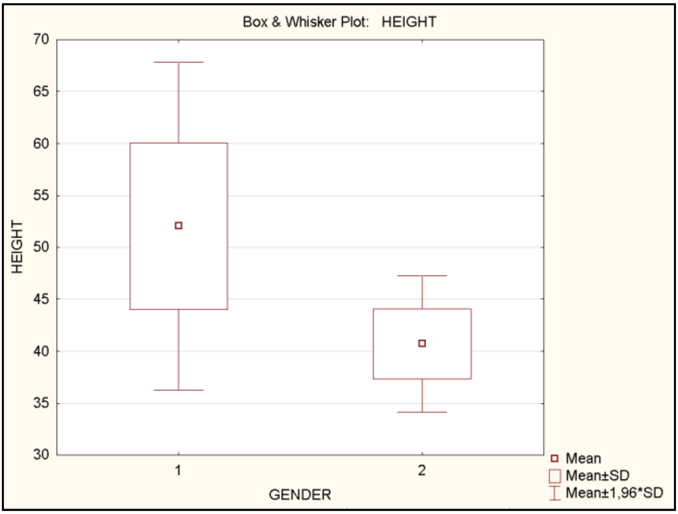
Переменная	Т-критерии; Группир.: GENDER (Turtles) Группа 1:1 Группа 2:2 Среднее 1	Среднее 2	t-знач.	cc	p	N набл. 1	N набл. 2	Ст.откл. 1	Ст.откл. 2	F-отн. дисперс.	p дисперс.
LENGTH	136,0417	113,3750	4,570481	46	0,000037	24	24	21,24900	11,77991	3,253815	0,006449
WIDTH	102,5833	88,2917	4,701470	46	0,000024	24	24	13,10465	7,07401	3,431778	0,004507
HEIGHT	52,0417	40,7083	6,368935	46	0,000000	24	24	8,04595	3,35545	5,749799	0,000087

Все три переменные («длина», «ширина» и «высота») различаются: особи женского пола имеют большие размеры по сравнению с особями мужского пола. Различие статистически значимо на уровне 0,05.

9. Графики, иллюстрирующие отмеченные закономерности.




Box & Whisker Plot: WIDTH
140
130
120
110
100
90
80
70
WIDTH
1
2
GENDER
Mean
Mean±SD
Mean±1,96*SD


Box & Whisker Plot: HEIGHT
70
65
60
55
50
45
40
35
30
HEIGHT
1
2
GENDER
Mean
Mean±SD
Mean±1,96*SD

10. Непараметрический критерий Манна-Уитни.

U критерий Манна-Уитни (Turtles)
По перем. GENDER
Отмеченные критерии значимы на уровне p <,05000

Перем.	Сум.ранг Группа 1	Сум.ранг Группа 2	U	Z	p-уров.	Z скорр.	p-уров.	N Группа 1	N Группа 2	2-х стор точное p
LENGT	772,0000	404,0000	104,0000	3,783706	0,000155	3,785042	0,000154	24	24	0,000079
WIDTH	777,0000	399,0000	99,0000	3,886804	0,000102	3,890079	0,000100	24	24	0,000047
HEIGH	816,0000	360,0000	60,0000	4,690971	0,000003	4,698756	0,000003	24	24	0,000000

Различие особей мужского и женского пола подтверждается значениями непараметрического критерия Манна-Уитни.

11. Выводы:

a) установлена тесная (сильная) корреляция между тремя параметрами (размерами) черепашек — ростом, длиной и шириной;

b) размеры особей женского пола превышают таковые для особей мужского пола;

c) полученные выводы статистически значимы на уровне 0,05.

Вопросы для обсуждения:

— метод наименьших квадратов для нахождения значений параметров уравнения регрессии;

— построение графика уравнения регрессии.

Контрольная работа № 1
по теме «Теория вероятностей и случайные величины»

Вариант № 1

1. Формулу сложения вероятностей $P(A+B) = P(A) + P(B)$ применяют, если события A и B:

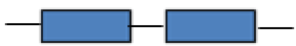
a) независимы;
b) несовместны;
c) достоверны;
d) невозможны;
e) противоположные.

2. Формула $F(x) = P(X < x)$ определяет:

a) условную вероятность;
b) нормальный закон распределения;
c) функцию распределения случайной величины;
d) плотность вероятности случайной величины;
e) ничего не определяет, так как икс большой не может быть меньше икса маленького.

3. В ящике 30 шаров: 10 красных, 5 синих и 15 белых. Найти вероятность того, что при единственном опыте из корзины достанут цветной шар.

4. Определить надежность схемы, состоящей из одинаковых, последовательно соединенных элементов, если вероятность отказа каждого элемента $q = 0{,}2$.



5. Дискретная случайная величина X задана законом распределения

X	-1	0	1	2
P	0,2	0,1	0,3	0,4

Найти $M(X), D(X), \sigma$, построить полигон.

6. Математическое ожидание и стандартное отклонение нормально распределенной случайной величины равны, соответственно, 9 и 2. Найти вероятность того, что в результате испытания X примет значение, заключенное в интервале [11,13].

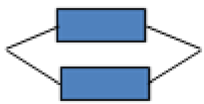
Вариант № 2

7. Формулу умножения вероятностей $P(AB) = P(A)P(B)$ применяют, если события А и В:
   a) независимы;
   b) несовместны;
   c) достоверны;
   d) невозможны;
   e) противоположные.

8. Формула $P(a < X < b) = F(b) - F(a)$ определяет:
   a) нормальный закон распределения;
   b) плотность вероятности случайной величины;
   c) вероятность того, что случайная величина примет значения из промежутка [a, b];
   d) ничего не определяет, но является точной формулой Ньютона-Лейбница;
   e) доверительный интервал.

9. Известно, что вероятность того, что лампочка сгорит при включении, равна 0,01. Найти вероятность того, что лампочка сгорит при втором включении.

10. Определить надежность схемы, состоящей из двух одинаковых параллельно соединенных элементов, если вероятность срабатывания каждого $p = 0{,}8$.



11. Дискретная случайная величина X задана законом распределения

X	-2	-1	0	1
P	0,2	0,2	0,5	0,1

Найти $M(X), D(X), \sigma$, построить полигон.

12. Математическое ожидание и среднее квадратичное отклонение нормально распределенной случайной величины равны соответственно 8 и 3; найти вероятность того, что в результате испытания X примет значение, заключенное в интервале [10, 12].

Контрольная работа № 2
по теме статистика

Вариант № 1

1. Разброс значений случайной величины X характеризуется:
a) математическим ожиданием M(X);
b) формулой Бернулли;
c) формулой Пуассона;
d) средним квадратичным отклонением σ;
e) функцией Лапласа.

2. Формула $F(x) = P(X < x)$ определяет:
a) условную вероятность;
b) нормальный закон распределения;
c) плотность вероятности случайной величины;
d) функцию распределения случайной величины;
e) нет правильного ответа.

3. Плотность вероятности случайной величины X задана формулой: $f(x) = \frac{1}{\sqrt{8\pi}} e^{-\frac{x^2}{8}}$.
Найти математическое ожидание M(X) и дисперсию D(x).

4. Какая из статистических совокупностей является частью другой?
a) выборочная — часть генеральной совокупности;
b) генеральная — часть выборочной совокупности;
c) выборочная и генеральная совокупности равны по численности;
d) генеральная — часть выборочной совокупности, если совокупности дискретные;
e) генеральная — часть выборочной совокупности, если совокупности непрерывные.

5. Статистическая гипотеза — это:
a) предположение о необходимом соотношении генеральной и выборочной совокупностей;
b) предположение о способах расчета параметров выборочной совокупности;
c) предположение о законе распределения генеральной совокупности;
d) предположение о возможных ошибках выборки.

6. Какой из критериев применяют как непараметрический для сравнения независимых выборок?

a) Стьюдента;
b) Манна-Уитни;
c) χ^2;
d) Фишера;
e) никакой из перечисленных выше.

7. Статистические гипотезы могут называться следующим образом:

a) нулевая;
b) неправильная;
c) подходящая;
d) правильная;
e) необходимая.

8. Если переменные связаны сильной отрицательной корреляционной связью, то:

a) они подчиняются нормальному закону распределения;
b) коэффициент корреляции $r > 0{,}5$;
c) коэффициент корреляции $r = 0{,}5$;
d) коэффициент корреляции $|r| > 0{,}7$;
e) они не подчиняются нормальному закону.

9. Критическое значение критерия — это:

a) значение, при сравнении с которым эмпирического критерия формулируется вывод относительно выдвинутых гипотез;
b) максимально возможное значение случайной величины;
c) значение, которое всегда меньше эмпирического критерия, полученного по данным генеральной совокупности;
d) минимально возможное значение случайной величины.

10. Для проверки значимости коэффициента корреляции находят значение:

a) параметра $T = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$;
b) величины коэффициента корреляции;
c) модуля коэффициента корреляции;
d) критерия χ^2;
e) коэффициентов линии регрессии.

Вариант № 2

11. Формулу умножения вероятностей $P(AB) = P(A) \cdot P(B)$ применяют, если события A и B:

a) независимы;

b) несовместны;

c) достоверны;

d) невозможны;

e) противоположные.

12. Формула $P(a < X < b) = F(b) - F(a)$ определяет:

f) нормальный закон распределения;

g) плотность вероятности случайной величины;

h) вероятность того, что случайная величина примет значения из промежутка [a, b];

i) ничего не определяет, но является точной формулой Ньютона-Лейбница;

j) доверительный интервал.

13. Плотность вероятности случайной величины X задана формулой:

$$f(x) = \frac{1}{\sqrt{8\pi}} e^{-\frac{(x-2)^2}{8}}$$

Найти математическое ожидание M(X) и дисперсию D(x).

14. От объема выборки зависит:

a) значимость полученных статистических выводов;

b) величина случайной ошибки;

c) величина систематической ошибки;

d) размер оплаты услуг сторонних организаций.

15. Статистической гипотезой может быть:

a) любое грамотно сформулированное предположение;

b) предположение, которое невозможно доказать;

c) предположение, которое невозможно опровергнуть;

d) предположение о нормальном законе распределения генеральной совокупности;

e) предположение о возможных ошибках при отборе данных.

16. Какой из критериев применяют как непараметрический для срвнения зависимых выборок?

a) Стьюдента;

b) критерий знаков;

c) χ^2;

d) Фишера;

e) никакой из перечисленных выше.

17. Статистические гипотезы могут называться следующим образом:

a) альтернативная;

b) неправильная;

c) подходящая;

d) правильная;

e) необходимая.

18. При наличии слабой связи между переменными коэффициент корреляции:

a) $r < 0$;

b) $r > 0$;

c) $r = 0$;

d) $|r| < 0{,}3$;

e) $r = 1$.

19. Гистограмма — это:

a) диаграмма рассеяния, на которой каждой точке соответствует два числа;

b) ступенчатая фигура, состоящая из прямоугольников, основанием которых служат интервалы значений изучаемой переменной величины;

c) несимметричная кривая линия, не имеющая максимума;

d) диаграмма рассеяния, на которой каждая точка соответствует одному измерению;

e) симметричная кривая линия с максимумом в центре.

20. При регрессионном анализе компьютер может провести оптимальную линию зависимости одной переменной от другой, при этом используемый алгоритм основан на методе:

a) максимального правдоподобия;

b) минимального расстояния;

c) наименьших квадратов;

d) линейного программирования;

e) нелинейного программирования.

Вариант № 3

21. Разброс значений случайной величины X характеризуется:

a) математическим ожиданием M(X);

b) формулой Бернулли;

c) дисперсией D (X);

d) формулой Пуассона;

e) функцией Лапласа.

22. Формула $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ определяет:

a) функцию распределения непрерывной случайной величины;

b) функцию распределения дискретной случайной величины;

c) плотность вероятностей случайной величины, подчиняющейся нормальному закону распределения;

d) показательную функцию;

e) закон распределения дискретной случайной величины.

23. Математическое ожидание нормально распределенной случайной величины X равно 4, а среднее квадратичное отклонение равно 3. Записать формулу для плотности вероятностей величины X.

24. Положительный коэффициент корреляции соответствует случаю, когда:

a) корреляция значима;

b) рост значений одной переменной сопровождается уменьшением значений другой переменной;

c) рост значений одной переменной сопровождается ростом значений другой переменной;

d) нулевая гипотеза верна;

e) альтернативная гипотеза верна.

25. В результате опыта случайная величина может принять то или иное значение, причем:

a) результат предсказуем;

b) заранее известно, какое именно;

c) результат предопределен;

d) заранее неизвестно, какое именно;

e) результат непредсказуем.

26. Статистической гипотезой называется:

a) предположение относительно генеральной совокупности;

b) предположение относительно выборки;

c) предсказание гадалки;

d) любое разумное предположение;

e) только то, что можно доказать.

27. Какой из критериев относится к числу непараметрических?

a) Стьюдента;

b) Розенбаума;

c) χ^2;

d) Фишера;

e) никакой из перечисленных выше.

28. Доверительная вероятность γ связана с уровнем значимости α простой формулой:

a) $\gamma + \alpha = 1$;

b) $\gamma - \alpha = 1$;

c) $\alpha - \gamma = 1$;

d) $\gamma + \alpha = 2$.

e) $\gamma + \alpha = 0$.

29. Гистограмма — это:

a) диаграмма рассеяния, на которой каждой точке соответствует два числа;

b) ступенчатая фигура, состоящая из прямоугольников, основанием которых служат интервалы значений изучаемой переменной величины;

c) несимметричная кривая линия, не имеющая максимума;

d) диаграмма рассеяния, на которой каждая точка соответствует одному измерению;

e) симметричная кривая линия с максимумом в центре.

30. Какая из статистических совокупностей является частью другой?

a) выборочная — часть генеральной совокупности;

b) генеральная — часть выборочной совокупности;

c) выборочная и генеральная совокупности равны по численности;

d) генеральная — часть выборочной совокупности, если совокупности дискретные;

e) генеральная — часть выборочной совокупности, если совокупности непрерывные.

Вариант № 4

31. Для оценки среднего значения случайной величины X используется:
a) дисперсия D (X);
b) формула Бернулли;
c) математическое ожидание M (X);
d) формула Пуассона;
e) среднее квадратичное отклонение σ.

32. Полигон — это:
a) график зависимости непрерывной случайной величины от ее вероятности;
b) график зависимости дискретной случайной величины от ее вероятности;
c) ломаная линия, соединяющая точки с координатами (x_i, p_i), где x_i — значения дискретной случайной величины, а p_i — соответствующие этим значениям вероятности;
d) плавная линия, соединяющая точки, координаты которых подчиняются нормальному закону;
e) зависимость дискретной случайной величины от ее математического ожидания;
f) график зависимости непрерывной случайной величины от ее дисперсии.

33. Плотность вероятности случайной величины X задана формулой:

$$f(x) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{(x-3)^2}{32}}$$

Найти математическое ожидание M(X) и σ.

34. Дискретная случайная величина считается заданной, если:
a) указаны все ее возможные значения;
b) указаны все ее вероятности;
c) указаны все ее возможные значения и соответствующие им вероятности;
d) ее значения взяты из опыта;
e) ее вероятности получены путем правильных логических подходов.

35. Статистической гипотезой может быть:
a) любое грамотно сформулированное предположение;

b) предположение, которое невозможно доказать;

c) предположение, которое невозможно опровергнуть;

d) предположение о нормальном законе распределения генеральной совокупности;

e) предположение о возможных ошибках при отборе данных.

36. Какой критерий применяют для решения вопроса о соответствии эмпирического распределения нормальному закону?

a) Фишера;
b) Розенбаума;
c) Стьюдента;
d) Манна-Уитни;
e) Хи-квадрат Пирсона.

37. При проверке наличия корреляции двух случайных переменных нулевой гипотезой H_0 является утверждение:

a) корреляции нет;
b) корреляция невозможна в принципе;
c) выборки зависимы;
d) выборки независимы;
e) объем выборок адекватен закону распределения.

38. Если переменные связаны слабой отрицательной корреляционной связью, то:

a) они не подчиняются нормальному закону распределения;

b) коэффициент корреляции $r > 0{,}5$;

c) коэффициент корреляции $r = 0{,}7$;

d) коэффициент корреляции $|r| < 0{,}3$;

e) необходимы дополнительные опыты или наблюдения — лучше и то и другое.

39. Какой из критериев является непараметрическим?

a) t — Стьюдента;
b) F — Фишера;
c) критерий Розенбаума;
d) хи-квадрат Пирсона;
e) ни один из перечисленных.

40. При регрессионном анализе компьютер может провести оптимальную линию зависимости одной переменной от другой, при этом используемый алгоритм основан на методе:

a) максимального правдоподобия;

b) минимального расстояния;

c) наименьших квадратов;

d) линейного программирования;

e) нелинейного программирования.

Образец задания и пример выполнения самостоятельной работы
Учебно-исследовательская работа студента (УИРС)

Министерство здравоохранения Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Уральский государственный медицинский университет»

Кафедра медицинской физики, информатики и математики

УИРС по статистике
«Анализ различия двух
независимых выборок»

Выполнили: студент
группы ОМП-300
Герасимов А.П.

Проверил: доцент
Крохалев В.Я.

г. Екатеринбург, 2017 г.

УИРС по дисциплине «Статистика»: Задание № 1

Выполнить анализ различия для двух независимых выборок по таблице: Папка примеров/Папка Datasets/**"Heart Disease"**. Переменные: **"Systolic Blood Pressure"** (зависимая переменная) и **"Tobacco Intake Level"** (группирующая переменная).

— Сделать титульный лист. Скопировать задание (на 2-й лист) и таблицу с данными.

— Проверить нормальность данных для зависимой переменной «качественно» и с помощью критерия χ^2.

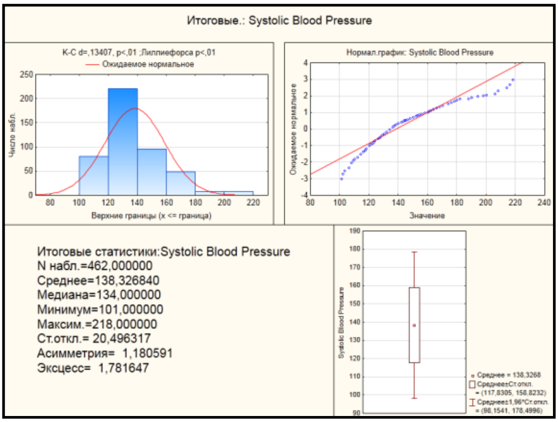
— Сравнить независимые выборки с помощью t-критерия Стьюдента. В отчете привести категоризованные гистограммы и нормальные вероятностные графики для каждой группы пациентов в отдельности.

— Если переменные не подчиняются нормальному закону распределения, сравнить переменные с помощью непараметрического критерия Манна-Уитни и проверить вывод п. 2.

— Сделать вывод о различии указанных переменных.

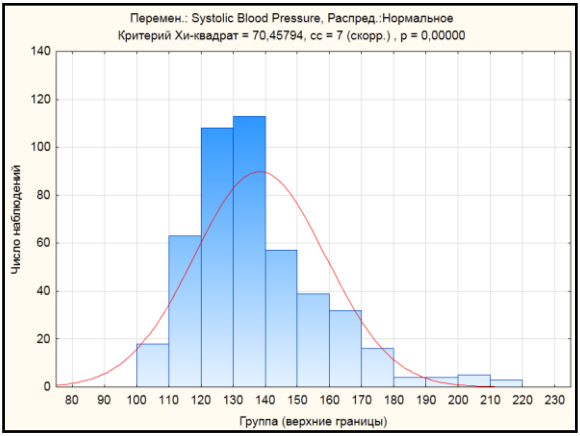
— Отчет в виде файла Word поместить в базе данных «Тандем».

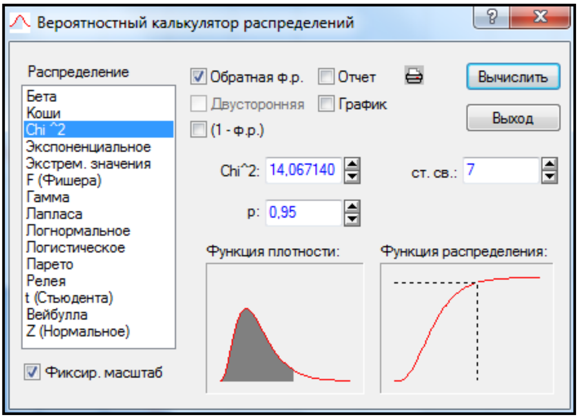
1. Проверка гипотезы о нормальном законе распределения (качественно) для зависимой переменной всех пациентов таблицы.



Среднее и медиана мало отличаются друг от друга, но асимметрия и эксцесс больше единицы значительно отличаются от значений для нормального закона (ноль). Нормальный вероятностный график показывает неудовлетворительное наложение точек на прямую линию.

1. Проверка гипотезы о нормальном законе распределения с помощью критерия согласия хи-квадрат для зависимой переменной всех пациентов таблицы.
2.





Хи^2 (эмпирический) = 70,5 > Хи^2 (критический) = 14,1. Гипотезу о нормальном законе для нашей переменной (систолическое кровяное давление) отвергаем.

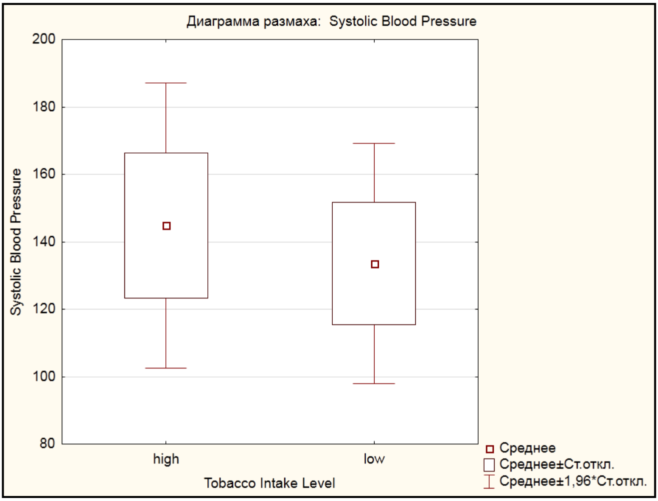
3. Сравнение независимых выборок с помощью t-критерия (Стьюдента).

	Т-критерии; Группир.: Tobacco Intake Level: New (HeartDisease) Группа 1:high Группа 2:low										
Переменная	**Среднее high**	Среднее low	t-знач.	cc	p	N набл. high	N набл. low	Ст.откл. high	Ст.откл. low	F-отн. дисперс.	p дисперс.
Systolic Blood Pressure	144,8547	133,5033	4,669394	266	0,000005	117	151	21,59073	18,17613	1,411016	0,047426

Значение эмпирического коэффициента t = 4.67 (много больше единицы) и выделено красным, что позволяет сделать вывод о статистически значимом различии. Приведенное в таблице значение *p* намного меньше (общепринятого в медицине) уровня значимости (0,05). Это позволяет сделать вывод: у курящих (группа 1) систолическое давление выше, чем у некурящих (группа 2).

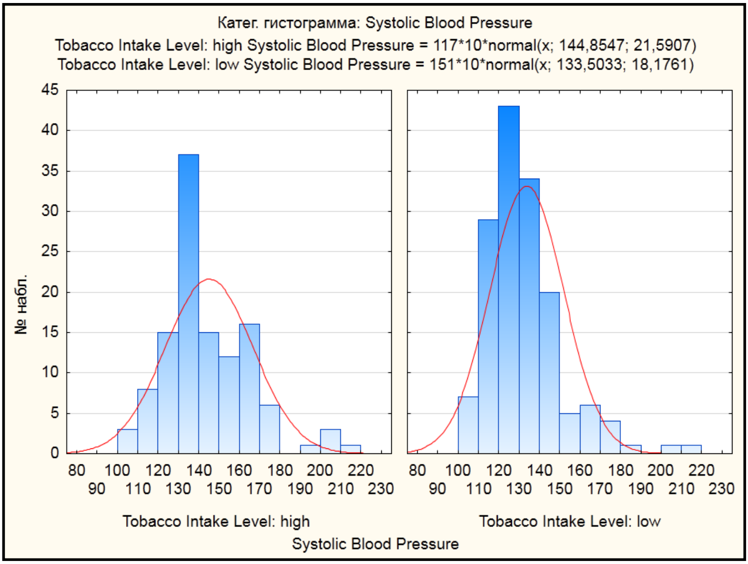
Вывод подтверждается и значениями F-критерия Фишера.

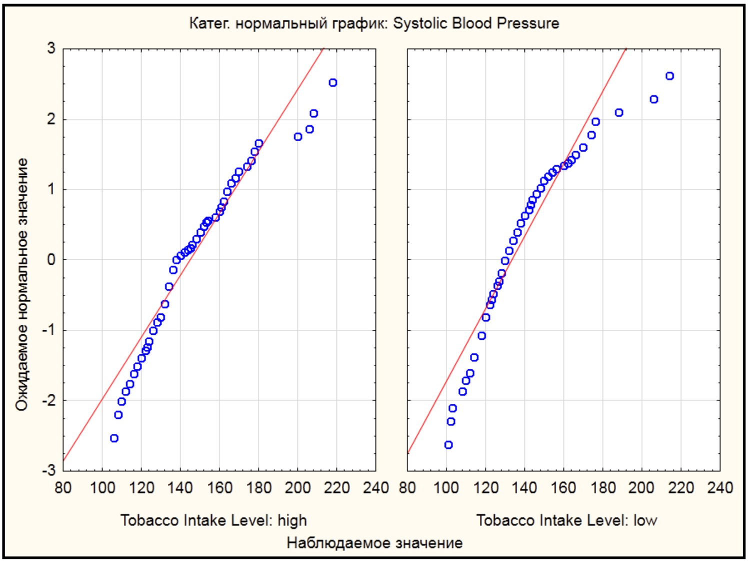
4. Диаграммы размаха (графическая иллюстрация закономерностей).



Как видно из диаграммы, области значений переменных перекрываются.

5. Построим категоризованные гистограммы и нормальные вероятностные графики для каждой группы пациентов в отдельности.





На «качественном уровне» мы не можем принять гипотезу о нормальном законе для наших переменных.

6. Сравнение переменных с помощью критерия Манна-Уитни.

Перем.	U критерий Манна-Уитни (HeartDisease) По перем. Tobacco Intake Level Отмеченные критерии значимы на уровне p <,05000								
	Сум.ранг high	Сум.ранг low	U	Z	p-уров.	Z скорр.	p-уров.	N high	N low
Systolic Blood Pressure	18811,50	17234,50	5758,500	4,885484	0,000001	4,889703	0,000001	117	151

Значение параметра Z выделено красным, что позволяет сделать вывод о статистически значимом различии. Приведенное в таблице значение p намного меньше чем 0,05. Систолическое давление у пациентов 1-ой группы (много курящие) выше, чем у пациентов 2-ой группы (мало курящие)

Выводы:

— Сравнение систолического давления у пациентов группы 1 (сильно курящих) и группы 2 (некурящих) показало, что пациенты первой группы отличаются большим систолическим давлением по сравнению с пациентами второй группы.

— Установленное различие статистически значимо на уровне 0,05 (фактически на уровне, значительно меньшем).

— Результат получен методами параметрической статистики (t и F критерии) и методом непараметрической статистики (критерий Манна-Уитни), поскольку гипотеза о нормальном законе распределения для переменных не может быть принята.

Вопросы к зачету

Основные понятия теории вероятностей. Понятие события. Достоверные и невозможные события. Случайное событие. Способы определения вероятностей: статистический, классический, геометрический. Диаграммы Эйлера. Правило сложения вероятностей для несовместных событий. Противоположные события. Условная вероятность и правило умножения вероятностей для независимых событий. Повторные независимые испытания и формула Бернулли.

Случайные величины и законы их распределения. Определение случайной величины. Типы и примеры случайных величин. Дискретные случайные величины. Закон распределения дискретной случайной величины (таблица), ее числовые (математическое ожидание, дисперсия, стандартное отклонение) и графические характеристики (полигон). Способы задания непрерывной случайной величины; функция распределения и плотность распределения вероятностей. Числовые характеристики непрерывной случайной величины. Примеры распределений (равномерное, биномиальное, показательное). Нормальный закон распределения: формула для плотности вероятностей, функция Лапласа и способ нахождения вероятности попадания нормально распределенной случайной величины в заданный интервал. Правило «трех сигм».

Основы математической статистики. Измерительные шкалы. Понятие генеральной и выборочной совокупностей. Описательные статистики. Методы группировки экспериментальных данных. Построение вариационных рядов. Расчет статистических характеристик. Точечные и интервальные оценки числовых характеристик генеральной совокупности. Построение доверительного интервала. Проверка статистических гипотез. Понятие критерий согласия.

Представление экспериментальных данных. Возможности прикладных статистических программ Excel и STATISTICA. Критерии согласия (критерии асимметрии и эксцесса, критерий $\chi 2$ Пирсона). Дисперсионный анализ. Параметрические критерии (Фишера — Снедекора, t-критерий Стьюдента для зависимых и независимых выборок).

Корреляционный анализ. Понятие о корреляционной зависимости. Параметрические и непараметрические показатели связи (коэффициент линейной корреляции Пирсона). Проверка значимости коэффициента корреляции с помощью t-критерия Стьюдента. Метод наименьших квадратов для нахождения линий регрессии.

Непараметрические методы статистики: а) для зависимых совокупностей (G-критерий знаков); б) для независимых совокупностей (U-критерий Манна-Уитни, Q-критерий Розенбаума). Применение коэффициента ранговой корреляции Спирмена для установления взаимосвязей между переменными. Проверка значимости коэффициента корреляции.

ЛИТЕРАТУРА

Список основной литературы

1. Основы высшей математики и статистики: учебник / И. В. Павлушков и др.: М.: ГЭОТАР-Медиа, 2012. — 432 с.
2. Морозов, Ю. В. Основы высшей математики и статистики / Ю. В. Морозов. — М.: Медицина, 2001. — 232 с.
3. Фадеева, Л. Н. Теория вероятностей и математическая статистика (задачи и упражнения) / Л. Н. Фадеева, Ю. В. Жуков, А. В. Лебедев. — М.: ЭКСМО, 2007. — 336 с.
4. Богинич, А. В. Учебное пособие по высшей математике / А. В. Богинич, М. А. Двинина, В. А. Телешев. — Екатеринбург: Изд. УГМА, 2007. — 82 с.
5. Боровиков, В. Популярное введение в программу STATISTICA / В. Боровиков. — М., 1998 (доступно для скачивания в Интернете).

Список дополнительной литературы

1. Банерджи, А. Медицинская статистика понятным языком: вводный курс / А. Банерджи, пер. с англ.; под ред. В. П. Леонова. — М.: Практическая медицина, 2014. — 287 с.
2. Герасимов, А. Н. Медицинская статистика: Учебное пособие / А. Н. Герасимов. — М.: Медицинское информационное агентство, 2007. — 476 с.
3. Гланц, С. Медико-биологическая статистика; пер. с англ. [Электронный ресурс]. — М.: Практика, 1998. — 459 с.
4. Ланг, Т. А. Как описывать статистику в медицине. Аннотированное руководство для авторов, редакторов и рецензентов / Т. А. Ланг, М. Сесик. — М.: Практическая медицина, 2011. — 480 с.
5. Письменный, Д. Т. Конспект лекций по теории вероятностей, математической статистике и случайным процессам / Д. Т. Письменный. — М.: Айрис-пресс, 2006. — 288 с.
6. Трухачева, Н. В. Математическая статистика в медико-биологических исследованиях с применением пакета STATISTICA / Н. В. Трухачева. — М.: ГЭОТАР-Медиа, 2013. — 379 с.
7. Воскобойников, Ю. Е. Математическая статистика (с примерами в Excel): уч. пособие / Ю. Е. Воскобойников, Е. И. Тимошенко. — Новосибирск: НГА-СУ, 2006. — 152 с.
8. Реброва, О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA / О. Ю. Реброва. — М.: Медиа Сфера, 2002. — 312 с.
9. Лапач, С. Н. Статистические методы в медико-биологических исследованиях с использованием Excel / С. Н. Лапач, А. В. Чубенко, П. Н. Бабич. — Киев: МОРИОН, 2001. — 408 с.

ПРИЛОЖЕНИЕ

Таблицы, используемые при выполнении заданий.
Таблицы взяты из папки примеров, прилагаемых к программе STATISTICA

Таблица 1

Surgical units

№	CLOTTING	ENZYME	LIVER	№	CLOTTING	ENZYME	LIVER
1	6,7	81	2,59	26	5,8	73	3,50
2	5,1	66	1,70	27	5,2	86	2,45
3	7,4	83	2,16	28	11,2	90	5,59
4	6,5	41	2,01	29	5,2	56	2,71
5	7,8	115	4,30	30	5,8	59	2,58
6	5,8	72	1,42	31	3,2	65	0,74
7	5,7	63	1,91	32	8,7	23	2,52
8	3,7	81	2,57	33	5,0	73	3,50
9	6,0	93	2,50	34	5,8	93	3,30
10	3,7	94	2,40	35	5,4	70	2,64
11	6,3	83	4,13	36	5,3	99	2,60
12	6,7	43	1,86	37	2,6	86	2,05
13	5,8	114	3,95	38	4,3	119	2,85
14	5,8	88	3,95	39	4,8	76	2,45
15	7,7	67	3,40	40	5,4	88	1,81
16	7,4	68	2,40	41	5,2	72	1,84
17	6,0	28	2,98	42	3,6	99	1,30
18	3,7	41	1,55	43	8,8	88	6,40
19	7,3	74	3,56	44	6,5	77	2,85
20	5,6	87	3,02	45	3,4	93	1,48
21	5,2	76	2,85	46	6,5	84	3,00
22	3,4	53	1,12	47	4,5	106	3,05
23	6,7	68	2,10	48	4,8	101	4,10
24	5,8	86	3,40	49	5,1	77	2,86
25	6,3	100	2,95	50	3,9	103	4,55

Таблица 2

Пациенты с сердечно-сосудистыми заболеваниями

№	Систолическое давление	Потребление табака	ЛПНП холестерин	Уровень жира в тканях	Уровень стресса	Индекс массы тела	Потребление алкоголя	Возраст заболевания
1	160	12	5,73	23,11	49	25,3	97,2	52
2	144	0,01	4,41	28,61	55	28,87	2,06	63
3	118	0,08	3,48	32,28	52	29,14	3,81	46
4	170	7,5	6,41	38,03	51	31,99	24,26	58
5	134	13,6	3,5	27,78	60	25,99	57,34	49
6	132	6,2	6,47	36,21	62	30,77	14,14	45
7	142	4,05	3,38	16,2	59	20,81	2,62	38
8	114	4,08	4,59	14,6	62	23,11	6,72	58
9	114	0	3,83	19,4	49	24,86	2,49	29
10	132	0	5,8	30,96	69	30,11	0	53
11	206	6	2,95	32,27	72	26,81	56,06	60
12	134	14,1	4,44	22,39	65	23,09	0	40
13	118	0	1,88	10,05	59	21,57	0	17
14	132	0	1,87	17,21	49	23,63	0,97	15
15	112	9,65	2,29	17,2	54	23,53	0,68	53
16	117	1,53	2,44	28,95	35	25,89	30,03	46
17	120	7,5	15,33	22	60	25,31	34,49	49
18	146	10,5	8,29	35,36	78	32,73	13,89	53
19	158	2,6	7,46	34,07	61	29,3	53,28	62
20	124	14	6,23	35,96	45	30,09	0	59
21	106	1,61	1,74	12,32	74	20,92	13,37	20
22	132	7,9	2,85	26,5	51	26,16	25,71	44
23	150	0,3	6,38	33,99	62	24,64	0	50
24	138	0,6	3,81	28,66	54	28,7	1,46	58
25	142	18,2	4,34	24,38	61	26,19	0	50
26	124	4	12,42	31,29	54	23,23	2,06	42
27	118	6	9,65	33,91	60	38,8	0	48
28	145	9,1	5,24	27,55	59	20,96	21,6	61
29	144	4,09	5,55	31,4	60	29,43	5,55	56
30	146	0	6,62	25,69	60	28,07	8,23	63
31	136	2,52	3,95	25,63	51	21,86	0	45
32	158	1,02	6,33	23,88	66	22,13	24,99	46
33	122	6,6	5,58	35,95	53	28,07	12,55	59
34	126	8,75	6,53	34,02	49	30,25	0	41
35	148	5,5	7,1	25,31	56	29,84	3,6	48
36	122	4,26	4,44	13,04	57	19,49	48,99	28

37	140	3,9	7,32	25,05	47	27,36	36,77	32
38	110	4,64	4,55	30,46	48	30,9	15,22	46
39	130	0	2,82	19,63	70	24,86	0	29
40	136	11,2	5,81	31,85	75	27,68	22,94	58
41	118	0,28	5,8	33,7	60	30,98	0	41
42	144	0,04	3,38	23,61	30	23,75	4,66	30
43	120	0	1,07	16,02	47	22,15	0	15
44	130	2,61	2,72	22,99	51	26,29	13,37	51
45	114	0	2,99	9,74	54	46,58	0	17
46	128	4,65	3,31	22,74	62	22,95	0,51	48
47	162	7,4	8,55	24,65	64	25,71	5,86	58
48	116	1,91	7,56	26,45	52	30,01	3,6	33
49	114	0	1,94	11,02	54	20,17	38,98	16
50	126	3,8	3,88	31,79	57	30,53	0	30
51	122	0	5,75	30,9	46	29,01	4,11	42
52	134	2,5	3,66	30,9	52	27,19	23,66	49
53	152	0,9	9,12	30,23	56	28,64	0,37	42
54	134	8,08	1,55	17,5	56	22,65	66,65	31
55	156	3	1,82	27,55	60	23,91	54	53
56	152	5,99	7,99	32,48	45	26,57	100,32	48
57	118	0	2,99	16,17	49	23,83	3,22	28
58	126	5,1	2,96	26,5	55	25,52	12,34	38
59	103	0,03	4,21	18,96	48	22,94	2,62	18
60	121	0,8	5,29	18,95	47	22,51	0	61
61	142	0,28	1,8	21,03	57	23,65	2,93	33
62	138	1,15	5,09	27,87	61	25,65	2,34	44
63	152	10,1	4,71	24,65	65	26,21	24,53	57
64	140	0,45	4,3	24,33	41	27,23	10,08	38
65	130	0	1,82	10,45	57	22,07	2,06	17
66	136	7,36	2,19	28,11	61	25	61,71	54
67	124	4,82	3,24	21,1	48	28,49	8,42	30
68	112	0,41	1,88	10,29	39	22,08	20,98	27
69	118	4,46	7,27	29,13	48	29,01	11,11	33
70	122	0	3,37	16,1	67	21,06	0	32
71	118	0	3,67	12,13	51	19,15	0,6	15
72	130	1,72	2,66	10,38	68	17,81	11,1	26
73	130	5,6	3,37	24,8	58	25,76	43,2	36
74	126	0,09	5,03	13,27	50	17,75	4,63	20
75	128	0,4	6,17	26,35	64	27,86	11,11	34
76	136	0	4,12	17,42	52	21,66	12,86	40
77	134	0	5,9	30,84	49	29,16	0	55
78	140	0,6	5,56	33,39	58	27,19	0	55
79	168	4,5	6,68	28,47	43	24,25	24,38	56

80	108	0,4	5,91	22,92	57	25,72	72	39
81	114	3	7,04	22,64	55	22,59	0	45
82	140	8,14	4,93	42,49	53	45,72	6,43	53
83	148	4,8	6,09	36,55	63	25,44	0,88	55
84	148	12,2	3,79	34,15	57	26,38	14,4	57
85	128	0	2,43	13,15	63	20,75	0	17
86	130	0,56	3,3	30,86	49	27,52	33,33	45
87	126	10,5	4,49	17,33	67	19,37	0	49
88	140	0	5,08	27,33	41	27,83	1,25	38
89	126	0,9	5,64	17,78	55	21,94	0	41
90	122	0,72	4,04	32,38	34	28,34	0	55
91	116	1,03	2,83	10,85	45	21,59	1,75	21
92	120	3,7	4,02	39,66	61	30,57	0	64
93	143	0,46	2,4	22,87	62	29,17	15,43	29
94	118	4	3,95	18,96	54	25,15	8,33	49
95	194	1,7	6,32	33,67	47	30,16	0,19	56
96	134	3	4,37	23,07	56	20,54	9,65	62
97	138	2,16	4,9	24,83	39	26,06	28,29	29
98	136	0	5	27,58	49	27,59	1,47	39
99	122	3,2	11,32	35,36	55	27,07	0	51
100	164	12	3,91	19,59	51	23,44	19,75	39
101	136	8	7,85	23,81	51	22,69	2,78	50
102	166	0,07	4,03	29,29	53	28,37	0	27
103	118	0	4,34	30,12	52	32,18	3,91	46
104	128	0,42	4,6	26,68	41	30,97	10,33	31
105	118	1,5	5,38	25,84	64	28,63	3,89	29
106	158	3,6	2,97	30,11	63	26,64	108	64
107	108	1,5	4,33	24,99	66	22,29	21,6	61
108	170	7,6	5,5	37,83	42	37,41	6,17	54
109	118	1	5,76	22,1	62	23,48	7,71	42
110	124	0	3,04	17,33	49	22,04	0	18
111	114	0	8,01	21,64	66	25,51	2,49	16
112	168	9	8,53	24,48	69	26,18	4,63	54
113	134	2	3,66	14,69	52	21,03	2,06	37
114	174	0	8,46	35,1	35	25,27	0	61
115	116	31,2	3,17	14,99	47	19,4	49,06	59
116	128	0	10,58	31,81	46	28,41	14,66	48
117	140	4,5	4,59	18,01	63	21,91	22,09	32
118	154	0,7	5,91	25	13	20,6	0	42
119	150	3,5	6,99	25,39	50	23,35	23,48	61
120	130	0	3,92	25,55	68	28,02	0,68	27
121	128	2	6,13	21,31	66	22,86	11,83	60
122	120	1,4	6,25	20,47	60	25,85	8,51	28

123	120	0	5,01	26,13	64	26,21	12,24	33
124	138	4,5	2,85	30,11	55	24,78	24,89	56
125	153	7,8	3,96	25,73	54	25,91	27,03	45
126	123	8,6	11,17	35,28	70	33,14	0	59
127	148	4,04	3,99	20,69	60	27,78	1,75	28
128	136	3,96	2,76	30,28	50	34,42	18,51	38
129	134	8,8	7,41	26,84	35	29,44	29,52	60
130	152	12,18	4,04	37,83	63	34,57	4,17	64
131	158	13,5	5,04	30,79	54	24,79	21,5	62
132	132	2	3,08	35,39	45	31,44	79,82	58
133	134	1,5	3,73	21,53	41	24,7	11,11	30
134	142	7,44	5,52	33,97	47	29,29	24,27	54
135	134	6	3,3	28,45	65	26,09	58,11	40
136	122	4,18	9,05	29,27	44	24,05	19,34	52
137	116	2,7	3,69	13,52	55	21,13	18,51	32
138	128	0,5	3,7	12,81	66	21,25	22,73	28
139	120	0	3,68	12,24	51	20,52	0,51	20
140	124	0	3,95	36,35	59	32,83	9,59	54
141	160	14	5,9	37,12	58	33,87	3,52	54
142	130	2,78	4,89	9,39	63	19,3	17,47	25
143	128	2,8	5,53	14,29	64	24,97	0,51	38
144	130	4,5	5,86	37,43	61	31,21	32,3	58
145	109	1,2	6,14	29,26	47	24,72	10,46	40
146	144	0	3,84	18,72	56	22,1	4,8	40
147	118	1,05	3,16	12,98	46	22,09	16,35	31
148	136	3,46	6,38	32,25	43	28,73	3,13	43
149	136	1,5	6,06	26,54	54	29,38	14,5	33
150	124	15,5	5,05	24,06	46	23,22	0	61
151	148	6	6,49	26,47	48	24,7	0	55
152	128	6,6	3,58	20,71	55	24,15	0	52
153	122	0,28	4,19	19,97	61	25,63	0	24
154	108	0	2,74	11,17	53	22,61	0,95	20
155	124	3,04	4,8	19,52	60	21,78	147,19	41
156	138	8,8	3,12	22,41	63	23,33	120,03	55
157	127	0	2,81	15,7	42	22,03	1,03	17
158	174	9,45	5,13	35,54	55	30,71	59,79	53
159	122	0	3,05	23,51	46	25,81	0	38
160	144	6,75	5,45	29,81	53	25,62	26,23	43
161	126	1,8	6,22	19,71	65	24,81	0,69	31
162	208	27,4	3,12	26,63	66	27,45	33,07	62
163	138	0	2,68	17,04	42	22,16	0	16
164	148	0	3,84	17,26	70	20	0	21
165	122	0	3,08	16,3	43	22,13	0	16

166	132	7	3,2	23,26	77	23,64	23,14	49
167	110	12,16	4,99	28,56	44	27,14	21,6	55
168	160	1,52	8,12	29,3	54	25,87	12,86	43
169	126	0,54	4,39	21,13	45	25,99	0	25
170	162	5,3	7,95	33,58	58	36,06	8,23	48
171	194	2,55	6,89	33,88	69	29,33	0	41
172	118	0,75	2,58	20,25	59	24,46	0	32
173	124	0	4,79	34,71	49	26,09	9,26	47
174	160	0	2,42	34,46	48	29,83	1,03	61
175	128	0	2,51	29,35	53	22,05	1,37	62
176	122	4	5,24	27,89	45	26,52	0	61
177	132	2	2,7	21,57	50	27,95	9,26	37
178	120	0	2,42	16,66	46	20,16	0	17
179	128	0,04	8,22	28,17	65	26,24	11,73	24
180	108	15	4,91	34,65	41	27,96	14,4	56
181	166	0	4,31	34,27	45	30,14	13,27	56
182	152	0	6,06	41,05	51	40,34	0	51
183	170	4,2	4,67	35,45	50	27,14	7,92	60
184	156	4	2,05	19,48	50	21,48	27,77	39
185	116	8	6,73	28,81	41	26,74	40,94	48
186	122	4,4	3,18	11,59	59	21,94	0	33
187	150	20	6,4	35,04	53	28,88	8,33	63
188	129	2,15	5,17	27,57	52	25,42	2,06	39
189	134	4,8	6,58	29,89	55	24,73	23,66	63
190	126	0	5,98	29,06	56	25,39	11,52	64
191	142	0	3,72	25,68	48	24,37	5,25	40
192	128	0,7	4,9	37,42	72	35,94	3,09	49
193	102	0,4	3,41	17,22	56	23,59	2,06	39
194	130	0	4,89	25,98	72	30,42	14,71	23
195	138	0,05	2,79	10,35	46	21,62	0	18
196	138	0	1,96	11,82	54	22,01	8,13	21
197	128	0	3,09	20,57	54	25,63	0,51	17
198	162	2,92	3,63	31,33	62	31,59	18,51	42
199	160	3	9,19	26,47	39	28,25	14,4	54
200	148	0	4,66	24,39	50	25,26	4,03	27

Таблица 3

Alligator size and primary food choices

№	LENGTH	FOOD	№	LENGTH	FOOD
1	1,240	Inverte	31	1,880	Inverte
2	1,300	Inverte	32	1,930	Inverte
3	1,300	Inverte	33	1,980	Inverte
4	1,320	Fish	34	2,030	Fish
5	1,320	Fish	35	2,030	Fish
6	1,400	Fish	36	2,160	Fish
7	1,420	Inverte	37	2,260	Fish
8	1,420	Fish	38	2,310	Fish
9	1,450	Inverte	39	2,310	Fish
10	1,450	Other	40	2,360	Fish
11	1,470	Inverte	41	2,360	Fish
12	1,470	Fish	42	2,390	Fish
13	1,500	Inverte	43	2,410	Fish
14	1,520	Inverte	44	2,440	Fish
15	1,550	Inverte	45	2,460	Fish
16	1,600	Inverte	46	2,560	Other
17	1,630	Inverte	47	2,670	Fish
18	1,650	Other	48	2,720	Inverte
19	1,650	Inverte	49	2,790	Fish
20	1,650	Fish	50	2,840	Fish
21	1,650	Fish	51	3,250	Other
22	1,680	Fish	52	3,280	Other
23	1,700	Inverte	53	3,330	Fish
24	1,730	Other	54	3,560	Fish
25	1,780	Inverte	55	3,580	Fish
26	1,780	Inverte	56	3,660	Fish
27	1,780	Other	57	3,680	Other
28	1,800	Inverte	58	3,710	Fish
29	1,800	Fish	59	3,890	Fish
30	1,850	Fish			

Таблица 4

Characteristic Height

№	Male Height	Female Height	№	Male Height	Female Height
1	67	69	27	66	68
2	70	70	28	71	66
3	70	67	29	65	68
4	65	66	30	67	75
5	70	69	31	71	68
6	64	61	32	74	68
7	68	69	33	67	67
8	61	72	34	65	69
9	67	73	35	70	66
10	70	62	36	69	69
11	66	65	37	67	73
12	57	63	38	67	67
13	68	70	39	69	68
14	63	71	40	66	66
15	69	74	41	74	71
16	63	66	42	69	70
17	74	70	43	69	66
18	71	62	44	63	64
19	72	67	45	70	69
20	68	73	46	74	68
21	73	70	47	70	58
22	63	74	48	69	73
23	67	68	49	65	
24	69	65	50	63	
25	66	61	51	66	
26	69	70	52	69	

Таблица 5

Characteristics

№	Рост, дюйм	Вес, фут	Возраст, год	Wellness1	Wellness2	Пол
1	69	261	32	61,1	59,0	male
2	66	154	78	38,5	50,8	male
3	67	198	32	39,0	65,2	female
4	70	144	35	62,2	75,9	male
5	65	185	50	47,9	71,7	male
6	70	165	58	66,0	47,9	male
7	63	212	33	19,6	44,6	male
8	72	187	61	44,2	59,6	female
9	73	161	45	58,5	70,0	female
10	69	133	66	64,2	74,9	male
11	74	202	29	60,1	44,5	male
12	66	179	39	42,6	50,3	male
13	70	180	45	60,4	54,9	female
14	71	174	43	44,2	75,9	female
15	69	213	39	44,4	44,7	male
16	65	175	57	34,9	59,4	male
17	61	215	20	45,1	54,7	male
18	62	190	36	36,4	65,3	female
19	67	193	32	45,4	49,5	female
20	73	168	62	58,6	53,7	female
21	72	188	32	54,0	64,6	male
22	74	223	56	37,1	37,7	female
23	68	186	59	47,2	43,8	female
24	65	189	24	51,4	64,7	female
25	71	205	20	24,6	69,4	male
26	74	199	68	57,5	48,6	male
27	67	190	66	34,2	62,1	male
28	66	183	58	67,1	47,0	female
29	68	282	35	40,4	60,4	female
30	71	145	67	64,6	51,6	male
31	73	204	49	63,9	87,4	male
32	69	132	32	22,8	64,8	male
33	67	80	32	43,9	74,9	female
34	66	184	48	40,9	66,4	male

35	68	127	49	41,1	61,5	male
36	70	153	67	41,8	70,3	male
37	67	122	71	60,5	46,7	male
38	65	279	21	31,1	76,4	male
39	68	167	66	48,3	43,8	female
40	70	158	23	58,9	57,3	male
41	71	152	77	53,7	35,5	female
42	70	152	45	32,8	77,2	female
43	66	200	67	36,1	59,0	female
44	64	217	25	31,9	64,8	male
45	67	144	41	41,7	72,7	male
46	69	179	34	63,2	55,6	female
47	67	171	79	53,0	37,6	female
48	66	208	32	60,2	78,1	female
49	70	186	25	46,4	53,8	male
50	70	176	67	35,8	60,2	female
51	69	145	72	57,7	79,4	female
52	69	193	32	46,1	72,2	female
53	61	230	32	57,0	35,0	female
54	70	254	32	48,9	58,2	female
55	69	195	47	46,7	62,3	female
56	73	157	43	43,6	60,7	female
57	68	206	25	40,1	63,8	female
58	70	226	62	33,8	55,3	male
59	66	198	71	39,5	48,6	male
60	57	217	45	59,2	42,1	male
61	68	177	32	40,6	72,7	male
62	63	149	40	62,8	53,8	male
63	69	188	32	65,1	76,1	male
64	63	190	46	51,1	79,0	male
65	74	196	32	56,4	58,6	male
66	71	145	32	50,1	58,8	male
67	69	156	32	50,9	85,9	female
68	68	233	27	65,9	63,1	male
69	73	189	43	59,6	64,2	female
70	63	224	57	79,6	70,9	male
71	67	150	35	52,4	55,9	male

72	58	210	33	58,3	81,0	female
73	64	144	31	34,2	44,7	female
74	69	199	57	33,4	62,6	male
75	66	193	32	53,9	60,0	male
76	68	220	32	50,1	53,7	female
77	65	201	75	50,8	66,2	male
78	68	113	32	52,5	67,7	female
79	75	169	35	28,6	75,3	female
80	68	158	44	42,0	68,9	female
81	67	177	54	44,5	62,9	male
82	66	158	32	45,1	48,4	female
83	70	191	46	38,8	80,2	male
84	70	141	35	60,6	51,5	female
85	67	198	32	58,2	49,4	male
86	67	240	21	40,4	61,0	male
87	69	165	65	32,1	66,9	male
88	70	87	40	74,0	65,2	female
89	66	173	52	55,5	84,6	female
90	74	242	32	35,4	72,6	female
91	69	186	36	56,0	65,1	male
92	65	198	32	62,3	54,6	female
93	62	217	32	64,4	67,8	female
94	74	153	53	49,0	54,5	male
95	63	210	71	54,9	45,6	female
96	69	200	32	56,5	51,1	male
97	69	232	25	63,2	63,1	female
98	63	167	75	59,0	49,5	male
99	61	253	44	64,3	64,7	female
100	66	215	32	52,8	53,5	female

Таблица 6

Turtles

№	Length	Width	Height	Gender	№	Length	Width	Height	Gender
1	98	81	38	1	25	93	74	37	2
2	103	84	38	1	26	94	78	35	2
3	103	86	42	1	27	96	80	35	2
4	105	86	42	1	28	101	84	39	2
5	109	88	44	1	29	102	85	38	2
6	123	92	50	1	30	103	81	37	2
7	123	95	46	1	31	104	83	39	2
8	133	99	51	1	32	106	83	39	2
9	133	102	51	1	33	107	82	38	2
10	133	102	51	1	34	112	89	40	2
11	134	100	48	1	35	113	88	40	2
12	136	102	49	1	36	114	86	40	2
13	138	98	51	1	37	116	90	43	2
14	138	99	51	1	38	117	90	41	2
15	141	105	53	1	39	117	91	41	2
16	147	108	57	1	40	119	93	41	2
17	149	107	55	1	41	120	89	40	2
18	153	107	56	1	42	120	93	44	2
19	155	115	63	1	43	121	95	42	2
20	155	117	60	1	44	125	93	45	2
21	158	115	62	1	45	127	96	45	2
22	159	118	63	1	46	128	95	45	2
23	162	124	61	1	47	131	95	46	2
24	177	132	67	1	48	135	106	47	2

Таблица 7

Анализы крови пациентов, прооперированных по методике и А (46 чел.), и В (100 чел.) в одной из больниц г. Екатеринбурга

№	Общий белок		Билирубин		Глюкоза		Мочевина		Лейкоциты	
	A	B	A	B	A	B	A	B	A	B
1	65	58	9,9	4,9	5,1	4	3,4	4,7	5,7	6,5
2	66	74	7,4	6,3	4,54	4,8	1,54	2,8	6,52	8
3	62	61	10,2	9,9	5,3	5	2,8	4,8	9,6	10
4	60	65	9,4	6,7	4,7	4,1	1,4	6,3	10,9	4,1
5	78	70	13,5	7,7	4,4	5,2	1,4	2,2	10,1	7,1
6	71	60	7,6	10	6,82	6	2,19	9,8	7,43	16,5
7	69	74	9,3	6,5	4,3	4,9	1,5	2,5	7,84	15,4
8	63	72	10	5,9	7,6	5,4	2,2	7,2	5,3	17
9	70	66	6,5	4,8	5,8	3,8	4,5	3,8	4,52	13,3
10	76	60	10,6	6	6	4,1	3,6	16,6	13,3	10,5
11	69	74	6,6	4,8	4,7	4,2	2,8	3,4	4,8	11
12	59	67	10	4,9	5,7	6,4	3,6	4,2	8	10,1
13	47	69	12,9	5,5	4,6	3,7	3,5	2,9	8	6,8
14	67	62	4,3	4,6	6,7	4,1	3,4	2,6	5,2	7,8
15	59	85	47,6	7,8	4,7	5	2,1	7,2	11,7	8,1
16	46	64	12,8	4	3,5	6,3	4	1,3	6,81	7,3
17	55	53	13,3	5,4	4,5	4,5	4,1	5,6	8,92	7,1
18	72	68	11,8	5,3	5,8	4	2,5	3,2	7,39	7,2
19	61	60	9,6	5,2	4,8	5,5	1,7	2,6	6,7	5
20	61	68	13,6	12	6	5,6	2,2	8,8	9,82	8,7
21	66	60	13	10	4,5	5,2	1,7	2,4	7,2	8,6
22	48	66	8,9	6	4,5	5,2	2,2	2,2	7,3	7,6
23	73	60	13,9	8,8	5,1	4,1	5,7	4,8	8,2	7,7
24	52	72	12	7	6	4,6	4,7	3,5	6,4	7,2
25	42,5	79	13,6	5,2	4,7	4,9	2,6	2,7	13,3	7,8
26	50	83	23,5	7	6,4	6,4	2,5	3,4	7,7	7,7
27	70	68	18	4	7,7	4,7	2,4	3,6	10,3	7,7
28	56	64	8,2	7,2	4,3	4,1	0,9	5,4	9,53	10,8
29	66	65	17,3	6,3	6,6	5,1	2,1	2	12	9
30	50	64	7	5,5	4,5	4	1,5	4	9,2	7,5

31	57	45	15,2	5	4,9	4	1,9	6,2	7,7	16,1
32	54	76	10,7	6,5	4,2	4,2	1,1	3	13,7	10,1
33	60	80	10,2	6	4,9	4,3	1,2	3,4	8	7,6
34	60	60	7,2	6,9	5,9	4,2	1,7	6,5	6,4	14,5
35	74	70	19,3	4,7	5,3	5	1,8	2,81	8,2	6,7
36	51	72	174	5,5	5,4	5,2	0,6	3	13,6	5,7
37	65	75	14,9	6,4	6,1	4	3,2	4,4	13,2	6,9
38	62	71	18,6	5	5,3	3,5	3,1	4,2	9,8	6,5
39	60	54	10,6	5,6	4,8	5,5	3,1	3,7	7,9	18,5
40	60	54	5,6	5,4	4,8	4,7	1,5	4,2	10,2	6,5
41	61	62	28	7,7	5	5,3	3,2	3,8	6	17
42	81	80	9,8	6,3	8	7	2,5	5,5	13,4	7,1
43	68	78	11,9	7	5,3	5,8	2,7	5	8,98	9
44	76	68	7,9	7,3	4,6	9,2	2,4	6,5	5,1	7,9
45	66	64	8,7	6	4,6	5,1	2	2,7	10,1	15,4
46	62	69	9,1	5	6,3	5	3,8	2,2	15,2	5,6
47		60		7,4		5		3,5		8
48		80		6		5,5		3,6		7,3
49		75		4,5		3,75		3		4,9
50		62		13		5		3,4		10
51		71		8,8		4,1		4,3		15,8
52		63		6		4		5,2		6
53		63		6		4		6,3		16,8
54		70		6		4,6		1,9		12,5
55		78		5		3,7		4,7		6,8
56		47		12		4,3		8		7,9
57		53		7		4,2		4		7,8
58		66		6,7		3,9		3,8		7,5
59		75		5		4,9		3,7		5,5
60		73		5,6		4,5		5		14,5
61		60		6		4,5		3,8		6,3
62		56		6,7		5		3,3		4,7
63		65		13,7		7		6,2		11,3
64		72		5,2		4,5		4,2		8,2
65		66		5		4,6		3,4		7,4
66		69		7,7		6,6		4,6		6,8
67		62		5,2		5		3		6,3

68		55		6,4		5,5		4,8		7,6
69		60		7,8		4,6		1,8		4
70		57		6		4,4		5,3		6,8
71		66		5		5,5		1,8		5,8
72		78		5		6		3		12,4
73		73		4		3,5		4,1		10,5
74		67		7		4,7		13,9		5,8
75		60		10,6		5,5		4		5,9
76		64		5		4		5,8		9,3
77		65		5,2		5,2		6		9,3
78		52		10		4,8		4,26		7,8
79		79		10		4,6		9,9		9,6
80		69		7		4		2,2		6,1
81		64		10,8		3,8		4		14,2
82		68		9,5		5,4		1,8		9,43
83		62		7,5		5,78		4,88		10,9
84		84		6,4		5,4		3,9		4,6
85		64		7,3		7,1		2,1		9,61
86		60		9,8		5,5		1,8		7,2
87		68		13,7		4,9		4,69		5,7
88		77		7,2		5,3		5		9,4
89		68		8,3		4,9		3,9		8,3
90		69		23		4,3		5		7,1
91		54		13		4,2		5,2		16,5
92		71		6		8,6		2,62		6,7
93		61		5		4,4		1,5		8,9
94		53		10,2		7,2		0,7		9,4
95		63		13,5		5,2		2,7		7,2
96		60		10		5,8		2,6		10
97		73		6,8		5,6		3		8,5
98		60		10,4		4,9		5,4		9,7
99		63		6,5		5		2,3		7,9
100		53		8,8		13,4		4,4		9,5

Учебное издание

Вадим Яковлевич Крохалев
Сергей Анатольевич Скопинов
Валерий Алексеевич Телешев

СТАТИСТИКА

Учебное пособие

ISBN 978-5-89895-860-2

Редактор Е. Бортникова
Корректор Л. Моисеева
Дизайн, верстка И. Дзигунова

Оригинал-макет подготовлен:
Издательство УГМУ
г. Екатеринбург, ул. Репина, 3, каб. 310
Тел.: (343) 214-85-65
E-mail: pressa@usma.ru

Подписано в печать 18.12.2018. Формат 60×84/16
Бумага офсетная. Печать цифровая. Усл. печ. л. 6,6
Тираж 100 экз. Заказ № 184

Отпечатано в типографии «Юника»
620014, г. Екатеринбург, ул. Московская, 29
Тел.: +7 (343) 371-16-12