

ФОРМЫ И МЕТОДЫ ПОДГОТОВКИ АНАЛИТИЧЕСКОЙ ИНФОРМАЦИИ

НАИМЕНОВАНИЕ ЗАНЯТИЙ	КОЛИЧЕСТВО ЧАСОВ
ЛЕКЦИИ	18
ПРАКТИКА	54
КОНСУЛЬТАЦИИ	6
ЭКЗАМЕН	26
ВСЕГО АУДИТОРНЫХ ЗАНЯТИЙ	104
СРС	42
ВСЕГО ЧАСОВ	144

Цели и задачи дисциплины

Целью изучения дисциплины «Формы и методы подготовки аналитической информации»

**является овладение знаниями и
практическими навыками в области
информационно-аналитического обеспечения
безопасности бизнеса. А также получение
систематизированных знаний о теории и
практике применения информационных
технологий по сбору и обработке
информационно-аналитической информации,
обеспечивающей безопасность бизнеса.
Задачами дисциплины является изучение
основных требований к информационно-
аналитической системе службы безопасности,
положений, которые охватывают возможности
применение новых информационных
технологий информационно-аналитического**

- ПК-23. способностью соблюдать в профессиональной деятельности требования правовых актов в области защиты государственной тайны и информационной безопасности, обеспечивать соблюдение режима секретности
- ПК-29. способностью анализировать показатели финансовой и хозяйственной деятельности государственных органов, организаций и учреждений различных форм собственности
- ПК-31. способностью осуществлять сбор, анализ, систематизацию, оценку и интерпретацию данных, необходимых для решения профессиональных задач
- ПК-34. способностью на основе статистических данных исследовать социально-экономические процессы в целях прогнозирования возможных угроз экономической безопасности
- ПК-36. способностью анализировать и интерпретировать финансовую, бухгалтерскую и иную информацию, содержащуюся в учетно-отчетной документации, использовать полученные сведения для принятия решений по предупреждению, локализации и нейтрализации угроз экономической безопасности
- ПК-38. способностью анализировать состояние и перспективы развития внешнеэкономических связей и их влияние на экономическую безопасность
- ПК-52. способностью проводить специальные исследования в целях определения потенциальных и реальных угроз экономической безопасности организации.

В результате освоения дисциплины студент должен уметь

применять информационные, аналитические и коммуникативные технологии для решения управленческих задач; применять базовые системы электронных государственных ресурсов в сети Интернет. Формировать отчетность о проделанной информационно-аналитической работе, уметь использовать современные информационные технологии, включая информационно-аналитические системы, профессиональные базы данных и др.

В результате освоения дисциплины студент должен владеть

навыками подготовки и принятия управленческих решений с использованием информационно-коммуникативных технологий; работы со стандартными базами данных и программным обеспечением; проводить мониторинг информационной среды. Оценивать качество информации, методы формирования и анализа информации, достоинства и недостатки различных информационных технологий и систем.

Понятия, **Термины** и определения

Термины и определения	Понятия и определения
<p>СТРУКТУРА И ЗАДАЧИ АНАЛИТИКИ: общие и частные принципы; аналитические технологии и их роль в решении управленческих задач</p>	
Аналитика	<p>Комплекс научных дисциплин, призванных обобщить разнообразные «подходы» (методы и инструментарии) к анализу информации, выработанные в различных отраслях научных исследований</p>




СУЩНОСТЬ АНАЛИТИКИ

В зависимости от специфики исторического и культурного контекста своих исследований, ученые и практики вкладывали в свое понимание феномена «аналитической деятельности». Тогда термин – «АНАЛИТИКА»????.

Что же понимать под термином «аналитика»?

СУЩНОСТЬ АНАЛИТИКИ

<p>«Аналитика» это</p>	<ul style="list-style-type: none"> ◆ методологическая основа процесса обработки информации 	
<p>«Аналитика» это</p>	<ul style="list-style-type: none"> ◆ сущностное знание о процессах реального мира; ◆ методология (метод, способ) познания – получение нового знания, как строго научного, (например, принятая обществом как доказанная «истина»), так и интуитивные методы («Я ТАК ДУМАЮ!!!») 	

– Значок «*Nota Bene*» («достоинно внимания»). Значком выделяются фрагменты текста, по мнению авторов, заслуживающие **особого** внимания. Это не означает, что все прочие фрагменты текста внимания не заслуживают.



СУЩНОСТЬ АНАЛИТИКИ (продолжение)



**«Аналитика»
это**

❖ форма мышления и мироощущения, опирающаяся на научные знания

❖ средство преобразования интуитивных представлений в логический, рациональный план мышления



**«Аналитика»
это**

❖ совокупность методов, с помощью которых можно выявлять **«скрытые смыслы»** в текстах и реальных социально-политических и экономических процессах

Все так думают

СУЩНОСТЬ АНАЛИТИКИ

**«Аналитика»
это**

❖ ядро научно-исследовательской работы (своего рода интеллектуальное оружие, мощный пласт интеллектуальной культуры);

Не все так думают, а только те, кто «добывает знания»

**«Аналитика»
это**

❖ *тот инструмент, с помощью которого «враги» разрушили Советский Союз, и т. д.*

ХА-ХА

Столь широкая палитра мнений и представлений научных сотрудников и практических работников показывает, какое различное содержание может вкладываться в один и тот же термин. Часть этих определений относится не к сущности,

**а к функциям «аналитики»
и формам ее проявления.**

Таким образом, под понятием **«аналитика»** фигурирует целый ряд образований, которые следует различать и по-разному именовать.

ТОГДА «АНАЛИТИКА» объединяет три компонента.

- *Методы методология информационной и аналитической работы.*
- *Организационное обеспечение этой работы.*
- *Создание технологического и методологического обеспечения. Разработка и создание инструментальных средств – инструментариев.*

Сущность аналитики (ещё раз)

ПРОЦЕССЫ, ОПРЕДЕЛЯЮЩИЕ СУЩНОСТЬ АНАЛИТИКИ (СВОЕГО РОДА ФАКТОРЫ)

<p>«Процесс анализа» целей Аналитика» это</p>	<p>❖ ядро научно-исследовательской работы (своего рода интеллектуальное оружие, мощный пласт интеллектуальной культуры)</p>	<p>Не все так думают, а только те, кто «добывает знания»</p>
<p>«Аналитика » это</p>	<p>❖ тот инструмент, с помощью которого «враги» разрушили Советский Союз, и т. д.</p>	<p>ХА-ХА</p>

Схема аналитического процесса



– **Значок «Ценная мысль».** По степени ценности сведения, отмеченные этим значком котируются выше, нежели те, которые помечены значком **«Nota Bene»**. Предполагается, что фрагмент текста, помеченный этим значком, имеет либо высокую практическую значимость, либо по уровню близок к научному обобщению.

Таким образом, **аналитика** – это, прежде всего, **основа интеллектуальной, логической и мыслительной деятельности, направленной на решение практических задач.** В ее основе лежит не столько принцип констатации фактов, сколько принцип «опережения событий», что позволяет организации или индивиду прогнозировать будущее состояние объекта анализа.

Аналитика играет интегрирующую роль в реконструкции и представления прошлого, вскрытии настоящего и **прогнозировании будущего.**

Аналитика выступает в качестве способа организации познавательной деятельности, нацеленной на поиск и вскрытие тех закономерностей и движущих сил, которые на момент начала исследований неизвестны.

Риск обусловлен тем, что никакой изобретенный метод истины заведомо не гарантирует и является лишь еще одной ступенью на пути к познанию феномена. В этом отношении даже ошибочное решение, если оно получило правильную оценку – тоже предмет аналитического осмысления, поскольку потенциально способно вооружить новыми знаниями.

В науке есть два основных пути исследования сущностей, процессов и явлений.

• Априорное формулирование «гипотезы» путем построения теории (стройной цепи логических рассуждений).


• Выявление структуры и закономерностей. Применяя при этом накопленный опыт, знания и умение.



Содержательная сторона аналитики

Виды анализа – анализ как сфера аналитики

Наименование	Наименование
Графический анализ	Структурный анализ
Анализ временных рядов	Ресурсный анализ
Анализ вариаций	Причинно-следственный анализ
Дискриминантный анализ	Кластерный анализ
Сравнительный анализ	Ретроспективный анализ
Метод главных компонент	Семантический анализ
Игровое моделирование	Контент-анализ, логический и лингвистический анализ
Теоретический анализ	Анализ показателей эффективности факторов: дисперсионный анализ; Корреляционный анализ



– **Значок «Definitio»**. Отмечает фрагменты текста, содержащие важные определения. Предполагает научную строгость и эмоциональную нейтральность формулировки вводимых понятий. Как правило, после определения в книге следуют довольно подробные пояснения и обоснования.

Главное – вычленить те стратегически важные **компоненты аналитической деятельности**, которые способствовали успеху в различных отраслях человеческой деятельности и на основе теоретического обобщения предложить достаточно универсальные рекомендации для ее совершенствования.

Аналитические технологии.

Принимаемые решения, основываются на **информациях о предмете управления**. Поэтому от **качественных характеристик этой информации**, таких как **адекватность, полнота, достоверность, своевременность, непротиворечивость и т.п.**, зависит **эффективность его работы**. Поэтому **«Информационные системы»** должны **предоставлять новые изделия и услуги, основанные на информации, которые обеспечат бизнесу конкурентное преимущество на рынке.**

Аналитические технологии – это **методики, которые на основе каких-либо моделей, алгоритмов, математических теорем позволяют по известным данным оценить значения неизвестных характеристик и параметров.**

Например, теорема Пифагора,
которая позволяет по длинам
сторон (***a*** и ***b***) прямоугольника
определить длину его диагонали
(***C***). Технология основана на
известной формуле **$c^2=a^2+b^2$** .

Некоторые комментарии

Другим примером аналитической технологии являются способы, с помощью которых обрабатывает информацию человеческий мозг. Даже мозг ребенка может решать задачи, неподвластные современным компьютерам, такие как распознавание знакомых лиц в толпе или эффективное управление несколькими десятками мышц при игре в футбол. Уникальность мозга состоит в том, что он способен обучаться решению новых задач – игре в шахматы, вождению автомобиля и т.д. Тем не менее, мозг плохо приспособлен к обработке больших объемов числовой информации – человек не сможет найти даже квадратный корень из, большого числа в уме, не используя калькулятора или алгоритма вычисления в столбик. На практике же часто встречаются задачи о числах, гораздо более сложные, чем извлечение корня.

Таким образом, человеку для решения этих задач необходимы дополнительные методики и инструменты.

Аналитические технологии нужны в первую очередь людям, принимающим важные решения – руководителям, аналитикам, экспертам, консультантам. Доход компании в большой степени определяется качеством этих решений – точностью прогнозов, оптимальностью выбранных стратегий.

Прогнозирование:

- курсов валют;
- цен на сырье;
- спроса и предложений;
- дохода компании;
- уровня безработицы;
- числа страховых случаев;

Оптимизация:

- расписаний;
- маршрутов;
- плана закупок;
- плана инвестиций;
- стратегии развития.

Как правило, для реальных задач бизнеса и производства не существует четких алгоритмов решения. Раньше руководители и эксперты решали такие задачи только на основе личного опыта. С помощью аналитических технологий строятся системы, позволяющие существенно повысить эффективность принятых решений.



***** Для справки**

Численность учёных в мире

Конец 18-го века
Середина 19

века

Начало 20 века

Начало 21 века

Около 1 тыс. чел

10 тыс. чел

100 тыс. чел

Свыше 5 млн.

чел

Современная наука включает в себя
около 15 тыс. дисциплин. Более 90%
всех научно-технических достижений
человечества приходится на 20 век

Вот некоторые данные

Соотношение лауреатов Нобелевской премии

Нобелевские премии -- международные премии названные в честь их учредителя шведского инженера НОБЕЛЯ А. Б. Присуждается ежегодно с 1901 года за выдающиеся работы в области физики, химии, медицины и физиологии, экономики. С 1969 года -- за литературные произведения, за деятельность по укреплению мира. Вручается в Стокгольме.



Например, термин АРИФМЕТИКА

Арифметика – наука о числе

(«аритмос» и «арифмос» -- число)

Первое представление о числе возникло из счёта отдельных предметов (камни, деревья, люди и т. п.). Результат счёта:

один, два, три, четыре и т. д.

Эти числа теперь носят название «натуральных чисел» или «целых чисел». Их количество бесконечно

Понятие «натуральные числа» нельзя определять как понятия «БОЛЕЕ ПРОСТЫЕ»



Почему это так?!

ЭВКЛИД (философ) «III век до н. э.» Определял «натуральное число» как **«МНОЖЕСТВО»** состоящее из единиц. (Так трактуют многие современные учебники)

Но термин (слово) **«МНОЖЕСТВО»** или «собрание» или «совокупность» и т. п. совсем не понятие **«ЧИСЛО»** .

Тогда, что же такое число?!

Это неограниченно продолжающийся ряд «целых чисел» – 1, 2, 3, ...и т. д.

Такой ряд называется «натуральным рядом».
Никому не приходило в голову использовать части целых единиц!!!

**Но делить единицу на части
становится необходимостью!**

**Например, необходимо найти длину диагонали
AC квадрата ABCD, сторона которого AB=1 м
(метр). Известно (ещё в глубокой древности),
что площадь квадрата ACEF, построенного на
диагонали, в точности равна удвоенной
площади исходного квадрата, а именно**

**$AC^2=2$, тогда величина
AC – диагональ квадрата ABCD,
должна удовлетворять в
«точности».**

**Между тем, никакое целое число и никакое дробное
число не могут удовлетворять уравнению
 $AC^2=2$ в «точности».**

Поэтому, если нельзя отказаться от геометрической точности измерения длин, то необходимо допущение,

а именно, уравнение $AC^2=2$
имеет точный корень,

$$AC = \sqrt{2}$$

Данное число может иметь только приближённое значение с любой степенью точности (доли единиц). Это и будет число, которое называется **«ИРРАЦИОНАЛЬНЫМ».**

$$AC = \sqrt{2} = 1,41 \text{ или } AC = \sqrt{2} = 1,414213562$$

Тогда **«ИРРАЦИОНАЛЬНОЕ»** число это совокупность (множество) некоторого количества целых положительных или целых отрицательных единиц и такой части единицы, которая не может быть выражена точно целым числом (число может быть и нулём). Совершенно очевидно, что иррациональное число представить нельзя, но можно с любой степенью точности заменить рациональным числом $\frac{m}{n}$. Числа, являющиеся корнями алгебраических уравнений с целыми коэффициентами, называются алгебраическими числами.

На сегодняшний день, в практике используются: **«иррациональные числа» «отрицательные числа», «комплексные числа».**

«Отрицательные числа»

Часто встречаются случаи, когда действие «вычитание» не всегда возможно, нельзя вычесть большее число из меньшего: 5 из 3, 200 из 100 и т. д. Долгое время в практике этого и не было нужно. Индусские математики (VII век до н. э.) нашли образы такого вычисления, используя операции при торговых расчётах. Если купец имел, например 500 денежных единиц, а закупал товара на 300, то у него оставалось 200 денежных единиц: $500-300=200$. А если он имел 300, а закупал товара на 500, то купец оставался в долгу на 200 денежных единиц: $300-500=200$ денежных единиц долга. В этом случае форма записи такой сделки выглядела так

$$300 - 500 = 200$$

с точкой наверху, означающее «200» денежных единиц долга. Такое толкование носило искусственный характер, так как купец никогда не находил сумму долга «вычитанием» $300-500$, а всегда выполнял вычитание $500-300$, кроме того, на этой основе можно лишь с натяжкой объяснить правило сложения и вычитание «чисел с точками наверху», но никогда нельзя было объяснить правило «умножение» или «деления».

«Отрицательные числа» (продолжение)

«НЕВОЗМОЖНОСТЬ» вычитания большего числа из меньшего обуславливает тем, что натуральный ряд чисел бесконечен толь в одну сторону. Но если последовательно вычитать число «1», начиная, например из числа «7», то можно получить числа 6, 5, 4, 3, 2, 1, а далее вычитание даёт уже «отсутствие числа» и дальнейшее действие вычитание невозможно. Потому что, если необходимо сделать вычитание всегда возможным, то надо учитывать следующее:

- «отсутствие числа» считать также «числом» - ноль;
- от этого числа (ноль) считать возможным отнимать ещё одну единицу и т. д.

Тогда возможно получать новые числа «с точками на верху». В настоящее время форма записи выглядит так: «-1», «-2», «-3»...и т. д. Эти числа называются целыми отрицательными числами. И стоящий в переде знак «-» носит название знака количества в отличии от знака вычитания знака действия. **Введение целых отрицательных чисел ведёт за собой и использование дробных отрицательных чисел.**

КОМПЛЕСНЫЕ ЧИСЛА

Алгебраическое уравнение **второй степени** иначе называется **«квадратным»**. Уравнение имеет **общий вид $ax^2+bx+c=0$** , a , b , c , – числа или буквенные выражения, содержащие известные величины (коэффициенты), причём коэффициент **«a»** не может быть равен нулю, иначе уравнение будет не квадратным (уравнение первой степени). При делении обе части уравнения на коэффициент **«a»** получается **уравнение вида $x^2+px+g=0$** $P = \frac{b}{a}; g = \frac{c}{a}$ **де**
Уравнение вида $x^2+px+g=0$ называется приведённым,
уравнение вида $ax^2+bx+c=0$ – не приведённым.

КОМПЛЕКСНЫЕ ЧИСЛА

Решение не приведённого уравнения $ax^2+bx+c=0$ имеет вид

$$x = \sqrt{m}$$

если m – положительное число, то корень квадратный из этого числа может иметь два значения: одно положительное, другое – отрицательное (абсолютные значения этих чисел одинаковые).

Например, $x^2 = 9$, удовлетворяет значениям

$x=+3$ и $x=-3$, то есть x имеет два значения «+3» и «-3» и перед радикалом часто ставятся знаки «+» и «-»

$$x = \pm\sqrt{m}; x = \pm\sqrt{9}$$

Числа \sqrt{m} и m могут быть

иррациональными числами. Например, при решении уравнения вида $x^2 = \pi$.

$$x^2 = \pi$$

Геометрически это означает: найти сторону квадрата равного по площади кругу с радиусом 1.

Его корень есть $x = \pm\sqrt{\pi}$

КОМПЛЕКСНЫЕ ЧИСЛА

(испоминание)

Если t – отрицательное число, то уравнение $x^2=t$ (например, $x^2=-9$) не может иметь ни положительного, ни отрицательного корня – уравнение не имеет решения оно не существует. В практике такие уравнения используются только для уравнений 3-й степени. Квадратные корни из отрицательного числа называют «мнимыми числами». Сумма действительного и мнимого числа называется «комплексным числом». Например, $2 + \sqrt{-3}$

Напоминание по выборкам

Выборки и выборочные распределения

Целью статистических выводов является вывод о некоторой совокупности, используя **выборку** из неё. Т. е. **выборка** основана на том, что необходимо предположить при использовании случайных величин

Если вся наблюдаемая (исследуемая) **совокупность** состоит из **N -элементов**, а для обследования берётся **выборка из n -элементов**, то каждая из

$$N!$$
$$(N - n)! * n!$$

возможных **выборок** может быть изменена с равной вероятностью. Такая процедура называется **взятием случайной выборки**.

На практике «получения случайных выборок» встречаются трудности, и при этом могут быть полезны **таблицы случайных чисел**.

*** Форма записи факториала.

«Факториалом» в математике называют произведение всех натуральных чисел, включая указанное число.

Обозначается факториал восклицательным знаком после числа, например:

$$4! = 1 * 2 * 3 * 4 = 24$$

В общем виде формулу для нахождения факториала можно записать так: $n! = 1 * 2 * 3 * 4 * \dots * (n - 2) \cdot (n - 1) \cdot n$



Заполните!



Факториал определён только для натуральных чисел и нуля.

Факториал нуля и единицы это «1»:

$$0! = 1; 1! = 1.$$

Термин «факториал» ввел в 1800 году французский математик

Аргобаст Луи Франсуа Антуан.

Обозначение $n!$ придумал чуть позже немецкий математик **Кристиан Крамп в 1808 году.**

******* **Выборочное среднее и выборочная дисперсия и будут являться «статистиками».**
Эти величины соответственно и будут **характеризовать** положение **центра и рассеивание выборки (т. е. дисперсию)**

Часто дисперсию (раздробленность данных) называют «ошибкой» (что не всегда правильно)

Выборочные распределения

Часто оказывается возможным найти **распределение вероятностей данной статистики, если известно распределение для совокупности, из которой была взята выборка.**

Распределение вероятностей **статистики** называется **выборочным распределением.**

Одним из наиболее важных выборочных распределений является **нормальное распределение.**

Если **Y** – нормальная случайная величина, то её **плотность распределения вероятностей** имеет вид

$$f(y) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\left(\frac{1}{2}\right) \cdot [(y-\mu)/\sigma]^2}, \text{ где } (-\infty < y < \infty)$$

или

$$f(y) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-[(y-\mu)/2\sigma]^2} - \text{функция плотности вероятности}$$

где $(-\infty < y < \infty)$ – среднее значение и дисперсия

$\sigma^2 > 0.$

*** Кривая нормального распределения
(кривая функции плотности вероятности)

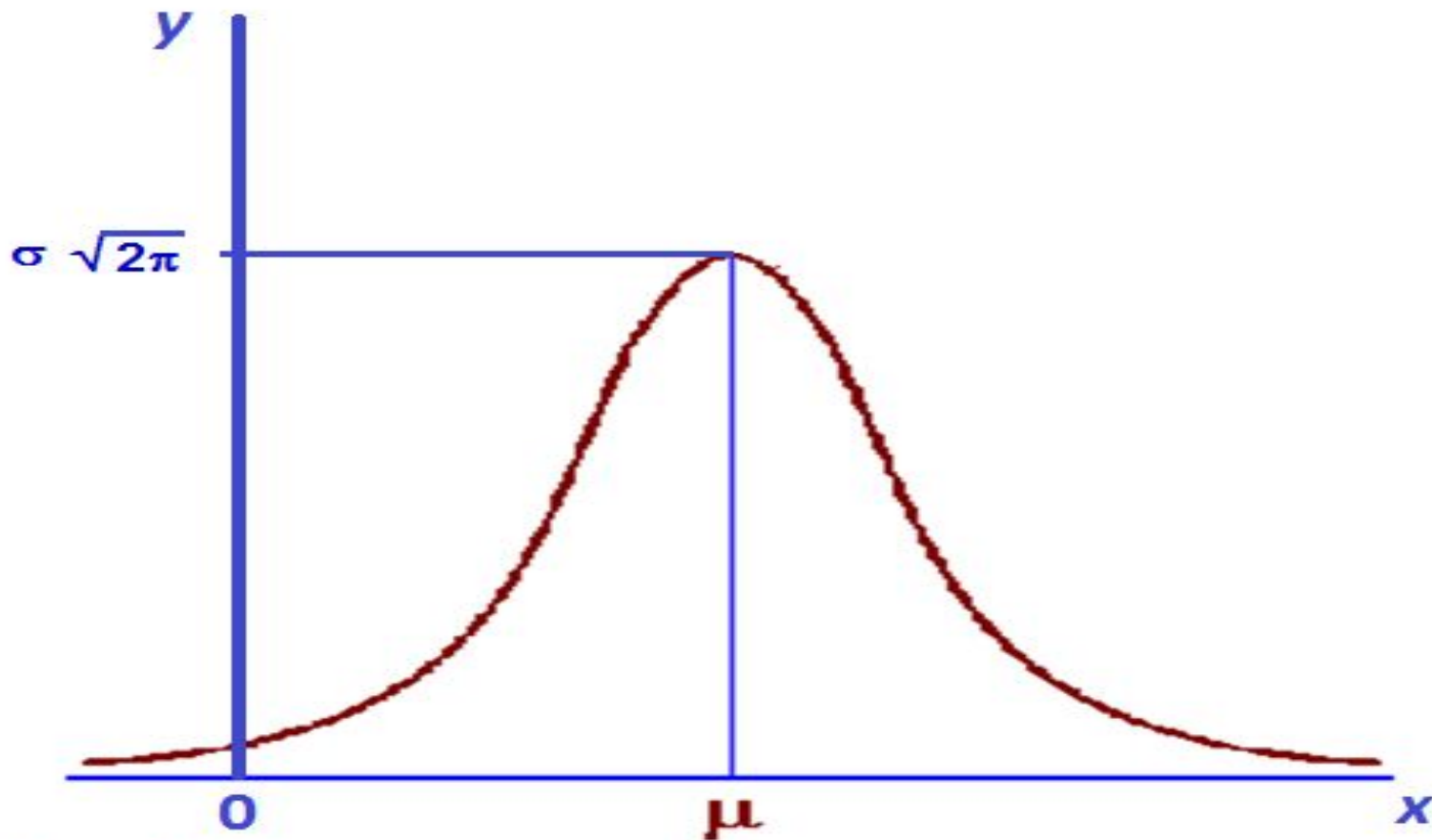


Рис.1. Кривая нормального распределения

μ
 σ^2
Напоминание μ – математическое ожидание (средняя величина); σ^2 – средне-квадратичное отклонение (дисперсия – раздробленность)



Функция Лапласа
(нормальное распределение называют
Лапласовским распределением)

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

– функция ошибок

Функцию Лапласа называют
«функцией ошибок»

с обозначением $erf(x)$
(ограниченное применение).

Другой вид функции

$$\bar{\Phi}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

*Носит название
«Нормированная
функция Лапласа»*

Наблюдения

В выборке при статистической оценке определяется любая функция от множества результатов наблюдений, не содержащих неизвестных параметров.

Например, $Y_1, Y_2, Y_3, Y_3 \dots Y_n$, представляют выборку некоторых наблюдений. Тогда **выборочное среднее** будет определено как (например, средняя арифметическая величина)

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

а выборочная дисперсия

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

или выборочное стандартное (среднеквадратичное) отклонение

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$$

Оба этих выражения связаны между собой следующими соотношениями

$$\bar{\Phi}(x) = \frac{1}{2} \Phi\left(\frac{x}{\sqrt{2}}\right) \quad \text{или} \quad \bar{\Phi}\left(\frac{x}{\sqrt{2}}\right) = \frac{1}{2} \Phi(x)$$

Для того чтобы определить попадание некоторых случайных величин подчинённых закону случайного распределения на участок числовой оси от **a** до **b** – участок **(a, b)** используется функция ЛАПЛАСА в следующей интерпретации

$$P(a < x < b) = \frac{1}{2} \left[\left(\Phi \frac{b-\mu}{\sigma\sqrt{2}} \right) - \left(\Phi \frac{a-\mu}{\sigma\sqrt{2}} \right) \right]$$

На практике часто используется
сокращённое обозначение



$$y \approx N(\mu, \sigma^2)$$

т. е. переменная Y распределена по нормальному
закону со средним μ и дисперсией σ^2

Нормальное распределение играет
центральную роль
в статистической теории.



Частным случаем **нормального
распределения** является
**стандартизованное нормальное
распределение**,
а именно при $\mu=0$ и $\sigma^2 = 1$.

Тогда совершенно очевидно,
что если выполняется условие,

$$y \approx N(\mu, \sigma^2)$$

то случайная переменная (или
стандартизованное нормальное
распределение y)

$$Z = \frac{y - \mu}{\sigma}$$

подчиняется стандартизованному
нормальному распределению, а именно

$$Z \approx N(0, 1)$$

Напоминание. *Частным случаем нормального
распределения является стандартизованное нормальное
распределение $\sigma^2 = 1$

а именно при $\mu=0$ и \bullet



Стандартизованное нормальное распределение y

$$Z = \frac{y - \mu}{\sigma}$$

называется **нормированным**,
а кумулятивная функция распределения —
«интегральной»
(или просто функцией распределения).
(Таблица кумулятивной функции
стандартизованного нормального
распределения приведена в табл.2).

$$\Phi(Z) = \int_{-\infty}^Z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

Таблица 2

Кумулятивная функция стандартизированного нормального распределения

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	Z
0,0	0,500 00	0,503 99	0,507 99	0,511 97	0,515 95	0,519 94	0,523 92	0,527 90	0,513 88	0,535 86	0,0
0,1	0,539 83	0,549 79	0,547 76	0,551 72	0,555 67	0,559 62	0,563 56	0,567 49	0,571 42	0,575 34	0,1
0,2	0,579 26	0,583 17	0,587 06	0,590 95	0,594 83	0,598 71	0,602 57	0,606 42	0,610 26	0,614 09	0,2
0,3	0,617 91	0,621 72	0,625 51	0,629 30	0,633 07	0,636 83	0,640 58	0,644 31	0,648 03	0,651 73	0,3
0,4	0,655 42	0,659 10	0,662 76	0,666 40	0,670 03	0,673 64	0,677 24	0,680 82	0,684 38	0,687 93	0,4
0,5	0,691 46	0,694 97	0,698 47	0,701 94	0,705 403	0,708 84	0,712 26	0,715 66	0,719 04	0,722 40	0,5
0,6	0,725 75	0,729 07	0,732 37	0,735 65	0,738 913	0,742 15	0,745 37	0,748 57	0,751 75	0,754 90	0,6
0,7	0,758 03	0,761 15	0,764 24	0,767 30	0,770 35	0,773 37	0,776 37	0,779 35	0,782 30	0,785 23	0,7
0,8	0,788 14	0,791 35	0,793 89	0,796 73	0,799 54	0,802 34	0,805 10	0,807 85	0,810 57	0,813 27	0,8
0,9	0,788 14	0,791 035	0,821 21	0,823 81	0,826 39	0,828 94	0,831 47	0,833 97	0,836 46	0,838 91	0,9
1,0	0,841 34	0,843 75	0,846 13	0,848 49	0,850 83	0,853 14	0,855 43	0,857 69	0,859 93	0,862 14	1,0
1,1	0,864 33	0,866 50	0,868 64	0,870 76	0,872 85	0,874 93	0,876 97	0,879 00	0,881 00	0,882 97	1,1
1,2	0,884 93	0,886 86	0,888 77	0,890 65	0,892 51	0,894 35	0,896 16	0,897 96	0,899 73	0,901 47	1,2
1,3	0,903 20	0,904 90	0,906 58	0,908 24	0,909 88	0,911 49	0,913 08	0,914 65	0,916 21	0,917 73	1,3
1,4	0,919 24	0,920 73	0,922 19	0,923 64	0,925 06	0,926 47	0,927 85	0,929 22	0,930 56	0,931 89	1,4
1,5	0,933 19	0,934 48	0,935 74	0,936 99	0,938 22	0,939 43	0,940 62	0,941 79	0,942 95	0,944 08	1,5
1,6	0,945 20	0,946 30	0,947 38	0,948 45	0,949 50	0,950 53	0,951 54	0,952 54	0,953 525	0,954 48	1,6
1,7	0,955 43	0,956 37	0,957 28	0,958 18	0,959 07	0,959 94	0,960 80	0,961 64	0,962 465	0,963 27	1,7
1,8	0,964 07	0,964 85	0,965 62	0,966 37	0,967 11	0,967 84	0,968 56	0,969 26	0,969 95	0,970 62	1,8
1,9	0,971 28	0,971 93	0,972 57	0,973 20	0,973 81	0,974 41	0,975 00	0,975 58	0,976 15	0,976 70	1,9

$$\Phi(Z) = \int_{-\infty}^Z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	Z
2,0	0,977 25	0,977 78	0,978 31	0,978 82	0,979 32	0,979 82	0,980 30	0,980 77	0,981 24	0,981 690	2,0
2,1	0,982 14	0,982 57	0,983 00	0,983 41	0,983 82	0,984 22	0,984 61	0,985 00	0,985 37	0,985 74	2,1
2,2	0,986 10	0,986 45	0,986 79	0,987 13	0,987 45	0,987 78	0,988 09	0,988 40	0,988 70	0,988 90	2,2
2,3	0,989 28	0,989 56	0,989 83	0,990 10	0,990 36	0,990 61	0,990 86	0,991 11	0,991 34	0,991 58	2,3
2,4	0,991 80	0,992 02	0,992 24	0,992 45	0,992 66	0,992 86	0,993 05	0,993 24	0,993 43	0,993 61	2,4
2,5	0,993 79	0,993 96	0,994 13	0,994 30	0,994 46	0,994 61	0,994 77	0,994 92	0,995 06	0,995 20	2,5
2,6	0,995 34	0,995 47	0,995 60	0,995 73	0,995 85	0,995 98	0,996 09	0,996 21	0,996 32	0,996 43	2,6
2,7	0,996 53	0,996 64	0,996 74	0,996 83	0,996 93	0,997 02	0,997 11	0,997 20	0,997 28	0,997 36	2,7
2,8	0,997 44	0,997 52	0,997 60	0,997 67	0,997 74	0,997 81	0,997 88	0,997 95	0,997 01	0,998 07	2,8
2,9	0,998 13	0,998 19	0,998 25	0,998 31	0,998 36	0,998 41	0,998 46	0,998 51	0,998 56	0,998 61	2,9
3,0	0,998 65	0,998 69	0,998 74	0,998 78	0,998 82	0,998 86	0,998 89	0,998 93	0,998 97	0,999 00	3,0
3,1	0,999 03	0,999 06	0,999 10	0,999 13	0,999 16	0,999 18	0,999 21	0,999 24	0,999 26	0,999 24	3,1
3,2	0,999 31	0,999 34	0,999 36	0,999 39	0,999 40	0,999 42	0,999 44	0,999 46	0,999 48	0,999 50	3,2
3,3	0,999 52	0,999 53	0,999 55	0,999 57	0,999 58	0,999 60	0,999 61	0,999 62	0,999 64	0,999 65	3,3
3,4	0,999 66	0,999 68	0,999 69	0,999 70	0,999 71	0,999 72	0,999 73	0,999 74	0,999 75	0,999 76	3,4
3,5	0,999 77	0,999 78	0,999 79	0,999 80	0,999 81	0,999 81	0,999 82	0,999 83	0,999 83	0,999 84	3,5
3,6	0,999 84	0,999 85	0,999 85	0,999 86	0,999 86	0,999 87	0,999 87	0,999 88	0,999 88	0,999 89	3,6
3,7	0,999 89	0,999 90	0,999 90	0,999 90	0,999 91	0,999 92	0,999 92	0,999 92	0,999 92	0,999 92	3,7
3,8	0,999 93	0,999 93	0,999 93	0,999 94	0,999 94	0,999 94	0,999 95	0,999 95	0,999 95	0,999 95	3,8
3,9	0,999 95	0,999 95	0,999 96	0,999 96	0,999 96	0,999 96	0,999 96	0,999 96	0,999 97	0,999 97	3,9

Количественные и качественные данные

Данные понятия можно представить как оценку данных. Например, заключения экспертов по оценке экономических показателей, или данные экспертизы криминалистических исследований и т. п.

Вот несколько ситуаций.

Вопрос.1. Ваше мнение о качестве выпускаемой продукции предприятием.

Ответы.

- 1. Отличная (очень хорошая);**
- 2. Хорошая (в основном реализуется);**
- 3. Удовлетворительная (реализации продукции неполная);**
- 4. Плохая (в основном скапливается на складе предприятия);**
- 5. Совершенно плохая (продукцию никто не покупает – всё на склад).**

Вопрос.2. Оценка поручена.

Ответы.

- 1. Индивидуальному эксперту (ИЭ);**
- 2. Комиссии (КМ).**

Вопрос 3. Возраст экспертов.

Ответы.

- 1. От 25 до 35 лет – молодые специалисты (МС);**
- 2. От 36 до 43 лет – опытные;**
- 3. Смешанный: от 25 до «45 – опыт в сочетании с молодостью» (ОиМ);**
- 4. Только зрелые опытные специалисты (ОС) – более 45 лет.**

Количественные и качественные данные

Данные понятия можно представить как оценку данных. Например, заключения экспертов по оценке экономических показателей, или данные экспертизы криминалистических исследований и т. п.

Вот несколько ситуаций.

Вопрос.1. Ваше мнение о качестве выпускаемой продукции предприятием.

Ответы.

- 1. Отличная (очень хорошая);**
- 2. Хорошая (в основном реализуется);**
- 3. Удовлетворительная (реализации продукции неполная);**
- 4. Плохая (в основном скапливается на складе предприятия);**
- 5. Совершенно плохая (продукцию никто не покупает – всё на склад).**

Вопрос.2. Оценка поручена.

Ответы.

- 1. Индивидуальному эксперту (ИЭ);**
- 2. Комиссии (КМ).**

Вопрос 3. Возраст экспертов.

Ответы.

- 1. От 25 до 35 лет – молодые специалисты (МС);**
- 2. От 36 до 43 лет – опытные;**
- 3. Смешанный: от 25 до «45 – опыт в сочетании с молодостью» (ОиМ);**
- 4. Только зрелые опытные специалисты (ОС) – более 45 лет.**

Вопрос 4. Сколько образцов продукции необходимо для соответствующей экспертизы?

Ответы.

- 1. Достаточно 2-х экземпляров;**
- 2. Необходимо от 3-х до 6-ти экземпляров;**
- 3. Необходимо от 3-х до 10 экземпляров;**
- 4. Достаточно 1-го экземпляра.**

Результаты оценки приведены в табл.1. Здесь, совершенно очевидно, что результаты можно подразделить на две группы:

«качественные» и «количественные».

К качественным данным отнесены ответы на вопросы 1 и 2.

К количественным – ответы на 3 и 4 вопросы.

Так выглядят результаты оценки

Эксперты или члены комиссии	Ответы на вопрос 1	Ответы на вопрос 2	Ответы на вопрос 3	Ответы на вопрос 4
«А»	Хорошая	(ИЭ)	25 до 35 – (МС)	1
«Б»	Удовлетворительная	(ИЭ)	36...43 – опытные	5
«В»	Плохая	(КМ)	36...43 – опытные	7
«Г»	Хорошая	(ИЭ)	(ОС) – более 45 лет	4
«Д»	Совершенно плохая	(КМ)	25...45 – (ОиМ)	3
«Е»	Хорошая	(КМ)	25 до 35 – (МС)	1
«Ж»	Отличная	(КМ)	25 до 35 – (МС)	2
«З»	Удовлетворительная	(ИЭ)	36...43 – опытные	1
«И»	Удовлетворительная	(КМ)	36...43 – опытные	3
...				

Понятно, что **вопрос 1** экспертам относится к **качественным данным**. Однако на практике, например при оценке по бальной системе, такие данные часто рассматриваются как **количественные**. То есть, могут возникнуть случаи, когда это может выглядеть так.

Ответы.

Отличная («очень хорошая») – соответствует баллу «5»;

Хорошая («в основном реализуется») – «4»;

Удовлетворительная («реализации продукции неполная») – «3»;

Плохая («в основном скапливается на складе предприятия») – «2»;

Совершенно плохая («продукцию никто не покапает – всё на склад») – «1».

"5"

"4"

"3"

"2"

"1"

Или так

Отличная («очень хорошая») –
соответствует баллу «2»;

Хорошая («в основном реализуется») – «1»;

Удовлетворительная («реализации
продукции неполная») – «0»;

Плохая («в основном скапливается на
складе предприятия») – «-1»;

Совершенно плохая («продукцию никто не
покапает – всё на склад») – «-2».

Ясно, что существует теоретические и практические (реальные) представления о действительности. Поэтому, одни и те же данные возможно рассматривать как количественные, так и качественные, всё зависит от ситуации, то есть где они используются.

Вот пример, посмотрите на таблицу 2 и попытайтесь определить, к каким категориям данных относятся «графы»:

- 1. Характеристика денежной массы;**
- 2. Денежная масса это;**
- 3. Комфортная температура хранения денег (бумажных купюр), °C;**
- 4. Оптимальная величина денег в «наминале», % от общего капитала**

Таблица 2

Мнение специалистов	Характеристика денежной массы	Денежная масса это	Комфортная температура хранения денег, °С	Оптимальная величина денег в «наминале», % от общего капитала
«А»	Показатель массы денег	Количественный показатель	15	14,5
«Б»	Скорость оборота денег	Качественный показатель	25	12,5
«В»	Запас денежной массы на 1 руб. ВВП	И качественный и количественный показатель	35	16,9
«Г»	Купюрное строение денег	Нет определения	20	18,8
...				

Ответы.

«Характеристика денежной массы» и «денежная масса» это – относятся **к качественным данным**.

«Комфортная температура хранения денег, °С»;
«Оптимальная величина денег в «наминале», % от общего капитала» – относятся **к количественным данным**.

Количественные данные Ряды распределения и гистограммы.

Выбор варианта «Результатов экспертной оценки», табл.3. «Отклонение результата от его средней величины – μ ».

Таблица 3
**Отклонение результатов оценки
от средней величины**

Результат №	Отклонение от средней величины, %	Результат №	Отклонение от средней величины, %	Результат №	Отклонение от средней величины, %
1	7,9	5	6,9	9	5,7
2	7,4	6	7,2	10	8,9
3	6,9	7	7,6	11	6,7
4	7,3	8	8,0	12	11,5

Первое, что необходимо предпринять, это разбить все «отклонения» на группы по «интервалам отклонений» и принадлежность оценочных данных по «экспертам». У каждого эксперта свой интервал «отклонений» и в каждом «результате» только один «отклонений». А именно, в каждом «результате» свой интервал отклонений. Каждый результат, у каждого «эксперта» только один результат с конкретной величиной «отклонений» – это и есть распределение по группам (табл.4).

Таблица 4

Распределение «интервалов отклонения» по группам

Эксперт	Интервал отклонение от средней величины, %	Распределение результатов по экспертам	Среднее значение интервала	Количество результатов (относительная частота)
1	5,7...6,9	1	6,3	0,083
2	7,0...7,5	3	7,25	0,25
3	7,6...8,0	4	7,8	0,33
4	8,1...8,4	3	8,25	0,25
5	8,5...11,5	1	10,0	0,083
	Всего	12		1,00

Результаты принадлежат «экспертам» в соответствии с «отклонениями» результатов оценки.

И каждому «эксперту» принадлежит разное количество «отклонений». Число «результатов» каждого «эксперта» называется «частотой».

Теперь можно определить, кому больше всего принадлежат «результаты».

У 3-го эксперта (табл.4) больше всего «результатов» – «4».

А относительная частота (ОЧ) принадлежность количества результатов у 3-го эксперта в относительных единицах (или в %) составит 0,33 (или 33,3%).

Относительная частота равна доле распределения от всей рассматриваемой совокупности

$$ОЧ = \frac{\text{Часть совокупности}}{\text{Вся совокупность}}$$

Относительная частота результатов по среднему значению результатов оценки величин 7,6...8,0 (средняя величина 7,8%, эксперт 3) равна 0,333 или 33,3%.

$$ОЧ = \frac{\text{Часть совокупности}}{\text{Вся совокупность}} = \frac{4 * 1}{12} = 0,333$$

или

$$ОЧ = \frac{4 * 100}{12} = 33,333\%$$

В графической интерпретации это выглядит так (рис.1). По горизонтали (по оси X) отложены средние значения «интервала отклонений», по вертикали (по оси Y – распределение результатов по экспертам) – частота (гистограмма 1) или относительная частота (гистограмма 2). Ширина столбца равна величине «интервала отклонений» (средина интервала обозначена средней величиной «отклонений»).



Рис.1. Гистограмма распределения результатов

Средние величины:

средняя арифметическая, средняя геометрическая и средняя гармоническая величина. Например, результат представления экспертных оценок 18 экспертов. Все эксперты распределены по трём предприятиям по 6 человек. Результаты оценки исследований представлены в табл. 5.

Таблица 5

Результат оценки

Предприятие «А»		Лаборатория «Б»		Лаборатория «В»	
Эксперт	Данные экспертизы, кол-во	Эксперт	Данные экспертизы, кол-во	Эксперт	Данные экспертизы, кол-во
Эксперт «А-1»	86	Эксперт «Б-1»	89	Эксперт «В-1»	225
Эксперт «А-2»	74	Эксперт «Б-2»	71	Эксперт «В-2»	47
Эксперт «А-3»	54	Эксперт «Б-3»	67	Эксперт «В-3»	57
Эксперт «А-4»	111	Эксперт «Б-4»	58	Эксперт «В-4»	47
Эксперт «А-5»	53	Эксперт «Б-5»	110	Эксперт «В-5»	94
Эксперт «А-6»	90	Эксперт «Б-6»	73	Эксперт «В-6»	0
Средний результат	78	Средний результат	78	Средний результат	94

Средние величины:

средняя арифметическая, средняя геометрическая и средняя гармоническая величина. Например, результат представления экспертных оценок 18 экспертов. Все эксперты распределены по трём предприятиям по 6 человек. Результаты оценки исследований представлены в табл. 5.

Таблица 5

Результат оценки

Предприятие «А»		Предприятие «Б»		Предприятие «В»	
Эксперт	Данные экспертизы, кол-во	Эксперт	Данные экспертизы, кол-во	Эксперт	Данные экспертизы, кол-во
Эксперт «А-1»	86	Эксперт «Б-1»	89	Эксперт «В-1»	225
Эксперт «А-2»	74	Эксперт «Б-2»	71	Эксперт «В-2»	47
Эксперт «А-3»	54	Эксперт «Б-3»	67	Эксперт «В-3»	57
Эксперт «А-4»	111	Эксперт «Б-4»	58	Эксперт «В-4»	47
Эксперт «А-5»	53	Эксперт «Б-5»	110	Эксперт «В-5»	94
Эксперт «А-6»	90	Эксперт «Б-6»	73	Эксперт «В-6»	0
Средний результат	78	Средний результат	78	Средний результат	94

Средний арифметический результат это тот, который приходится на одного эксперта (человека) предприятия, т. е. количество экспертиз, приходящихся на одно предприятие поделённое на число экспертов.

Предприятие «А»

$$A = \frac{86 + 74 + 54 + 111 + 53 + 90}{6} = 78$$

Предприятие «Б»

$$B = \frac{89 + 71 + 67 + 58 + 110 + 73}{6} = 78$$

Предприятие «В»

$$B = \frac{225 + 47 + 57 + 47 + 94}{5} = 94$$

Расчёт средних геометрической и гармонической величин производится по следующим выражениям

Средняя геометрическая величина

$$\sqrt[n]{x_1 * x_2 * x_2 * \dots x_n}$$

Средняя гармоническая величина

$$\frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}}$$

где $x_1, x_2, x_3, \dots x_n$ – величины переменных





Медиана

В тех случаях, когда имеются слишком большие или слишком малые значения наблюдений необходимо определять не среднюю величину, а медиану.

Медиана – значение, которое приходится на середину ряда, если разложить результаты в порядке возрастания или убывания. Например, результат представления экспертных оценок (см. табл. 5.)

Предприятие «А»	–	53	54	<u>74</u>	<u>86</u>	90...111
Предприятие «Б»	–	58	67	<u>71</u>	<u>73</u>	89...110
Предприятие «В»	–	47	47	<u>57</u>	<u>94</u>	225



Медиана предприятия «А»:
если значения середины ряда равны
74 86, тогда средняя величина,
т. е. медиана равна

$$\frac{74 + 86}{2} = 80,0$$

Медиана предприятия «Б»:
если значения середины ряда равны
71 73, тогда средняя величина,
т. е. медиана равна

$$\frac{71 + 73}{2} = 72,0$$

Медиана предприятия «В»:
если значения середины ряда равны
57 77, тогда средняя величина,
т. е. медиана равна

$$\frac{57 + 94}{2} = 75,50$$



Медиана (продолжение)

Если ряд состоит **из нечётного** числа значений, то медианой будет величина, находящаяся точно по середине:

47 47 57 94 225 – величина
57 – медиана

Если ряд состоит **из чётного** числа значений, то медианой будет среднее значение между **3-й и 4-й** (вариант лаборатории «А») величинами:

54 64 74 79 86...111 – величина
74 79 – медиана, а именно

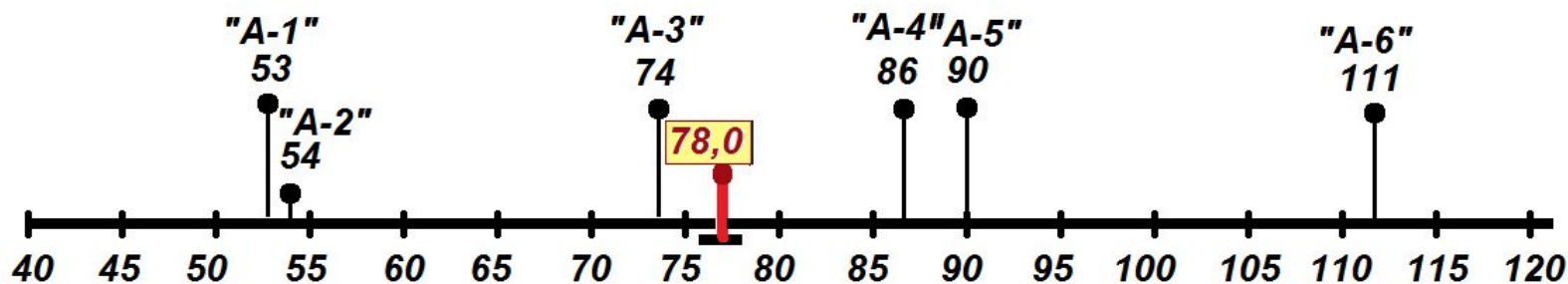
$$\frac{74 + 79}{2} = 76,5$$

Стандартное отклонение

Если рассмотреть результаты оценки предприятий «А» и «Б» (табл.5), то очевидно, что средние величины имеют одинаковые значения – «78» (рис.2).

Лаборатория "А"

Средняя величина "76,5"



Лаборатория "Б"

Средняя величина "72"

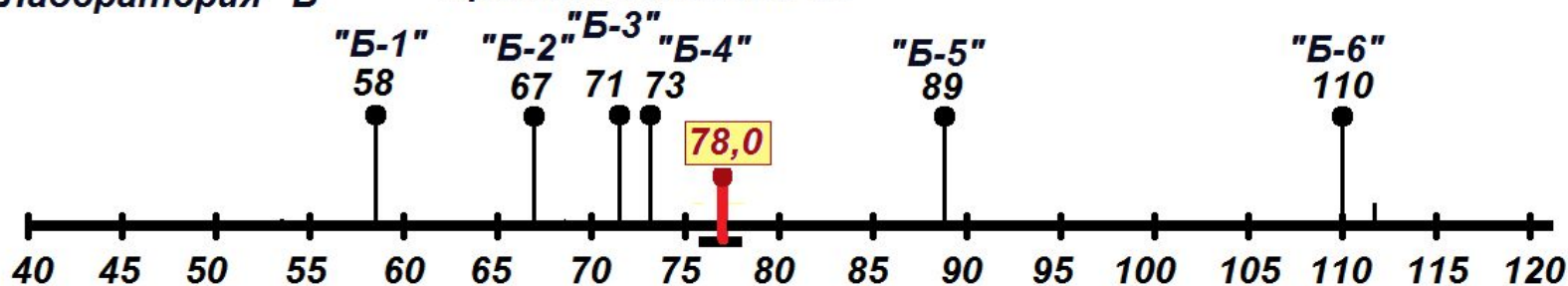
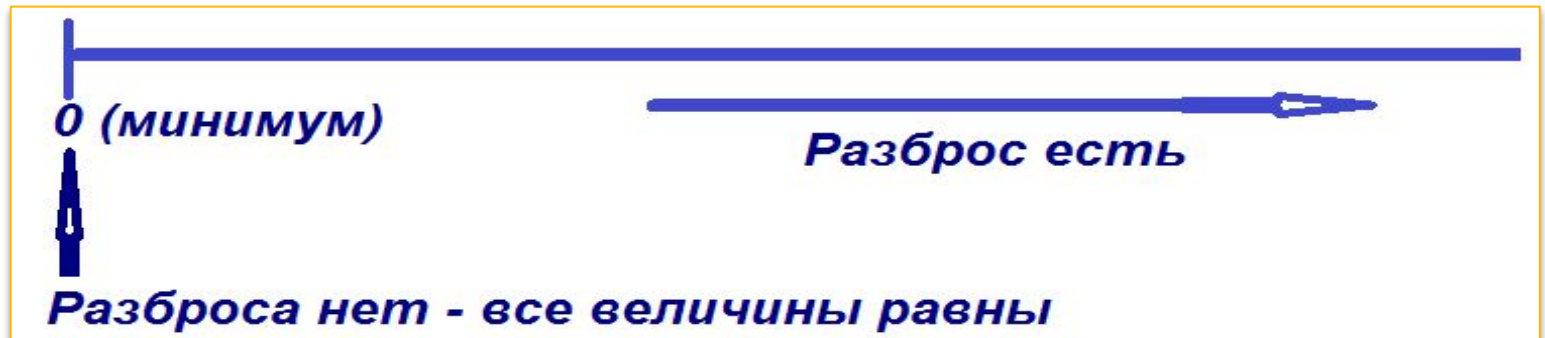


Рис.2. Шкала результатов экспертных оценок предприятий «А» и «Б»: средние величины «А» и «Б» одинаковы – «78», но ситуации, представленные по шкале сильно отличаются

Стандартное отклонение (продолжение)

Следовательно, стандартное отклонение (СО) это показатель, характеризующий **«отдельное отличие»** от средней величины. И становится очевидным, что величина отклонения не может быть **меньше «нуля»**



И стандартное отклонение возможно рассчитать по формуле

$$CO = \sqrt{\frac{\sum_i^n (i - \text{я величин} - \text{среднего значения})^2}{n - \text{значений (общее количество)}}$$

Стандартное отклонение результатов оценки предприятия «А»

$$CO_{\text{«А»}} = \sqrt{\frac{(86-78)^2 + (74-78)^2 + (54-78)^2 + (111-78)^2 + (53-78)^2 + (90-78)^2}{6}} = 20,47$$

Стандартное отклонение результатов оценки предприятия «Б»

$$CO_{\text{«Б»}} = \sqrt{\frac{(89-78)^2 + (71-78)^2 + (67-78)^2 + (58-78)^2 + (110-78)^2 + (73-78)^2}{6}} = 17,03$$

Стандартное отклонение результатов оценки предприятия «В»

$$CO_{\text{«В»}} = \sqrt{\frac{(225-94)^2 + (47-94)^2 + (57-94)^2 + (47-94)^2 + (94-94)^2}{5}} = 79,73$$

Совершенно очевидно, что больший разброс результатов – отдельное отклонение от средней величины, будет на предприятии «А» против показателей предприятия «Б» (20,47 и 17,03).

Формулу для расчёта стандартного отклонения можно записать **и в таком виде**

$$CO = \sqrt{\frac{\sum_i^n (i\text{-я величин} - \text{среднего значения})^2}{n\text{-значений (общее количество)} - 1}}$$

где от общего количества рассматриваемых величин **вычитается «1»**.

Первая формула

$$CO = \sqrt{\frac{\sum_i^n (i\text{-я величин} - \text{среднего значения})^2}{n\text{-значений (общее количество)}}$$

применяется при вычислении стандартного отклонения **генеральной совокупности**,

вторая формула

$$CO = \sqrt{\frac{\sum_i^n (i\text{-я величин} - \text{среднего значения})^2}{n\text{-значений (общее количество)} - 1}}$$

при определении стандартного отклонения **выборочной совокупности**



От снх 31 10 15

Генеральная совокупность – все рассматриваемые результаты (объекты, события и т. п.), выборочная совокупность это группа результатов отобранных из генеральной совокупности. То есть, если имеется возможность получить данные о всех объектах совокупности (случай, результат, события и т. п.).

Например, как результаты представленные в табл. 5. В практике обычно такие данные собрать не представляется возможным (данные генеральной совокупности), поэтому почти всегда применяется вторая формула

$$CO = \sqrt{\frac{\sum_i^n (i\text{-я величин} - \text{среднего значения})^2}{n\text{-значений (общее количество)} - 1}}$$

Вероятность. Вычисление вероятности

Функция распределения плотности вероятности.

При обработке результатов любого типа информации очень часто применяется термин **«вероятность»** чего-либо. «Данная информация маловероятна» или «вероятна». Вероятность события меньше «0,05» (или «0,002,» «»0,08» ...) Что необходимо знать, чтобы вычислить вероятность события, результатов оценки информации и т. д...? **Например результат оценки экспертами экономического состояния предприятий металлургической промышленности. Удалось оценить 575 предприятий со средним «положительным» результатом (стабильного функционирования) – 53. При этом стандартное отклонение равно 10. Результат оценки в виде гистограммы будет выглядеть так, рис. 3.**

«Результаты оценки»

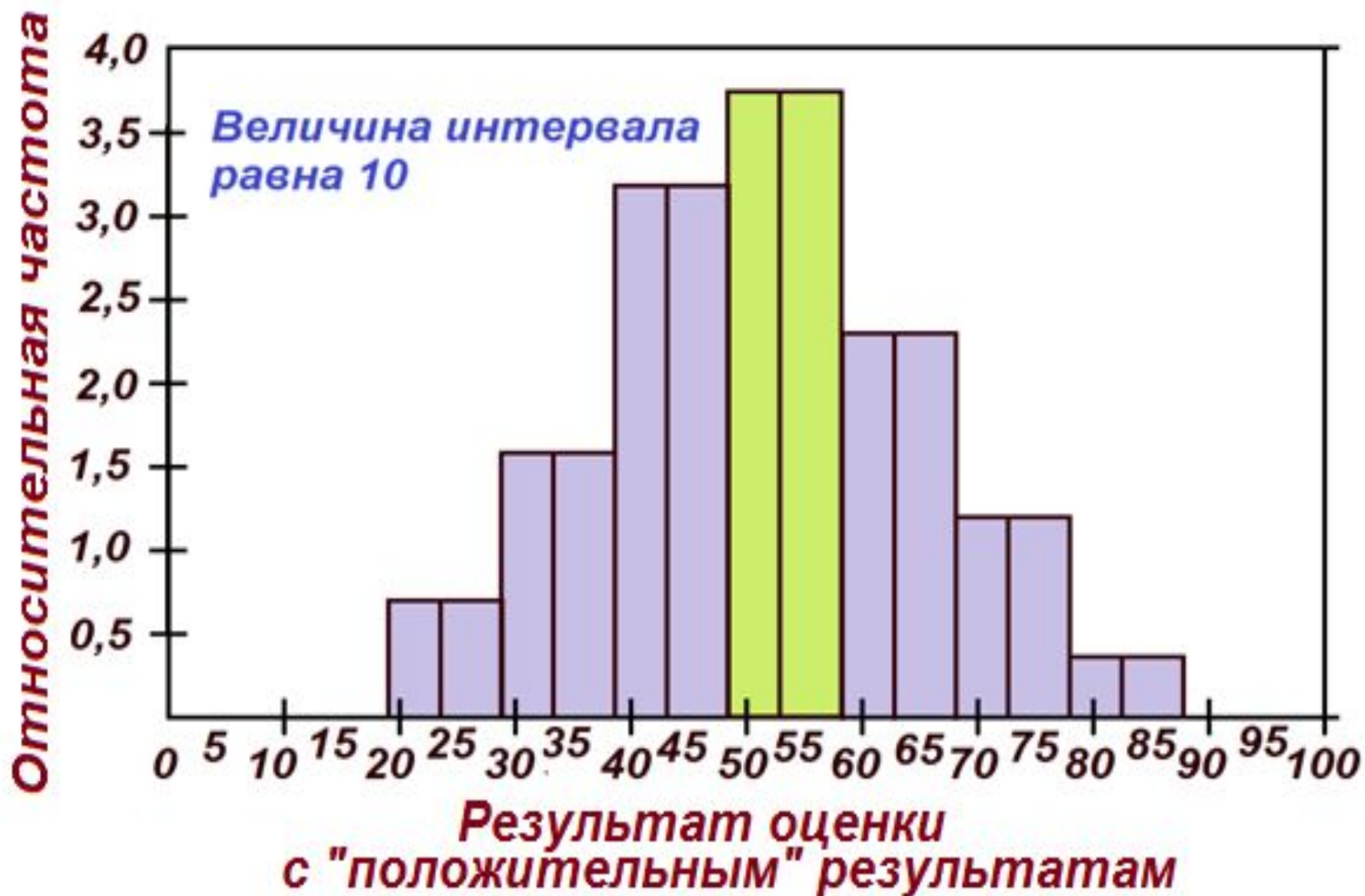
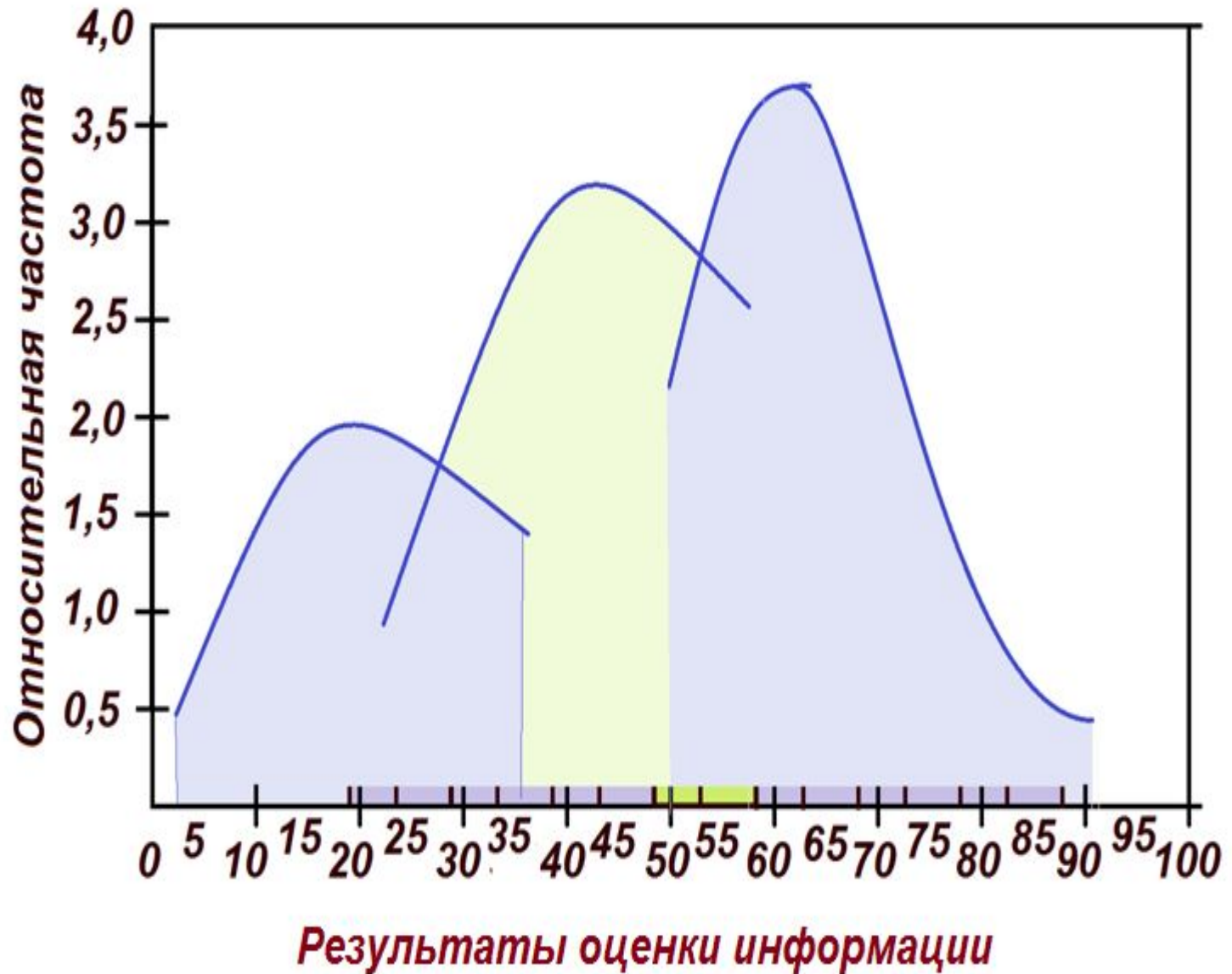


Рис. 3. Гистограмма «результатов оценки»

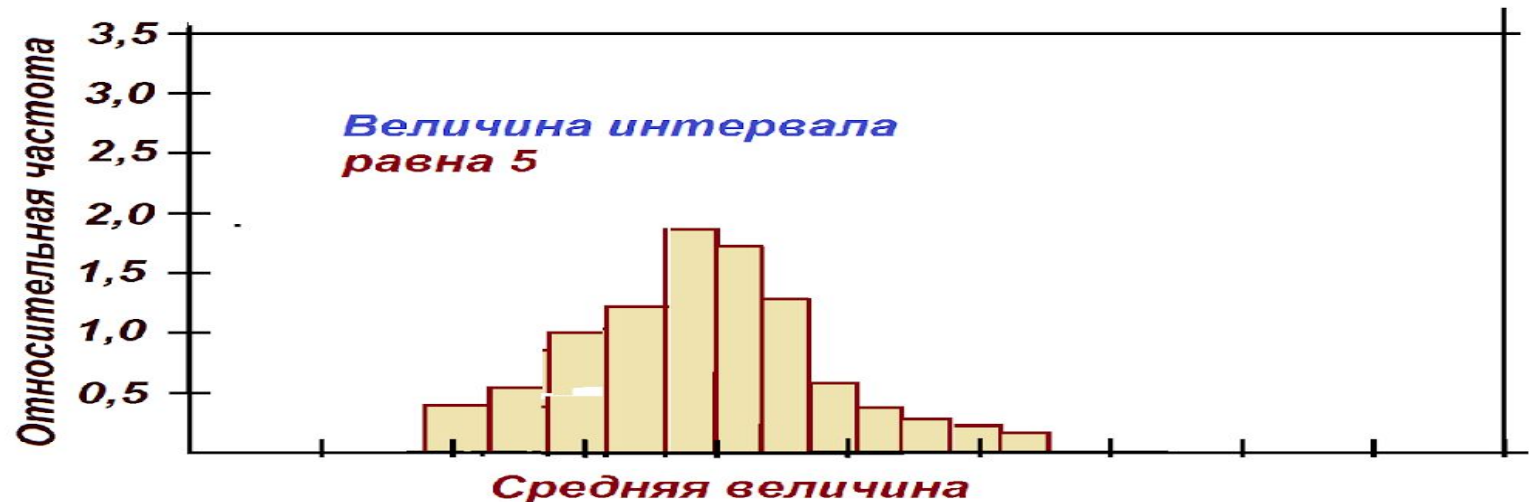
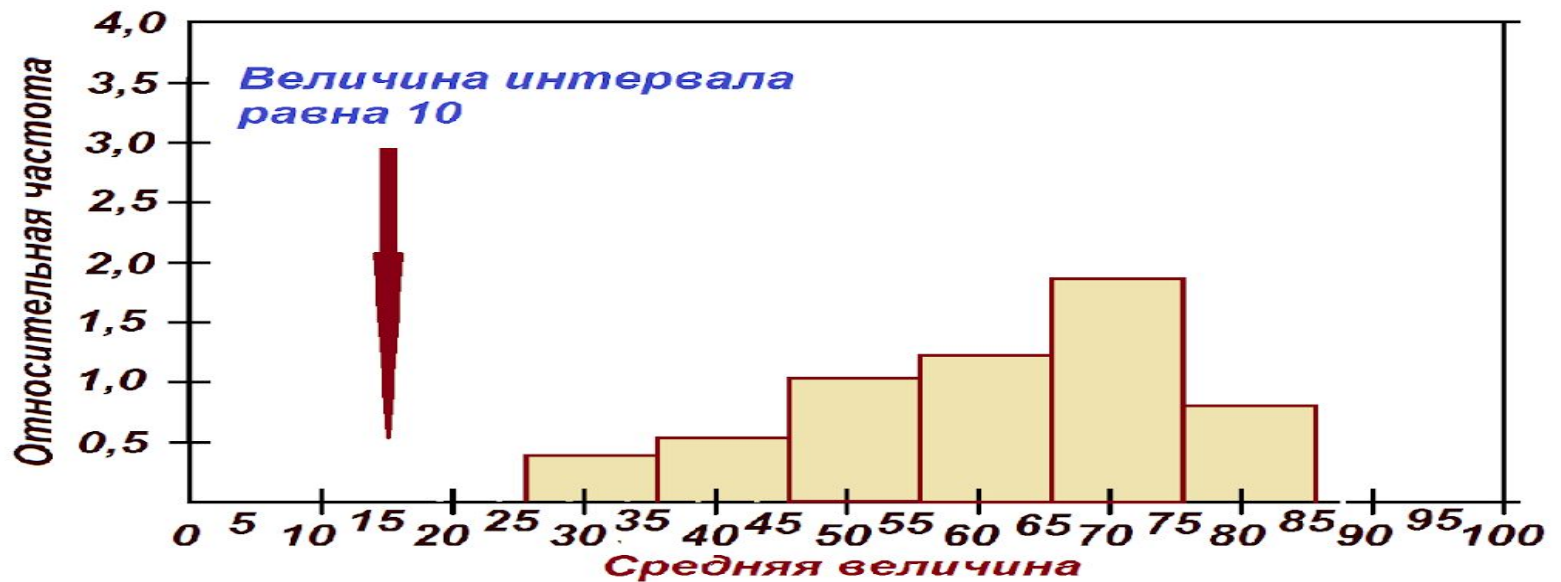
Если величина интервала на гистограмме стремится к нулю, то получается кривая распределения, а функция описываемая этой кривой называется функцией распределения вероятностей (рис.4. и рис.5). Существует множество форм кривых распределения, например, такие формы кривых как на **рис. 5 и рис. 6.**

Виды гистограмм

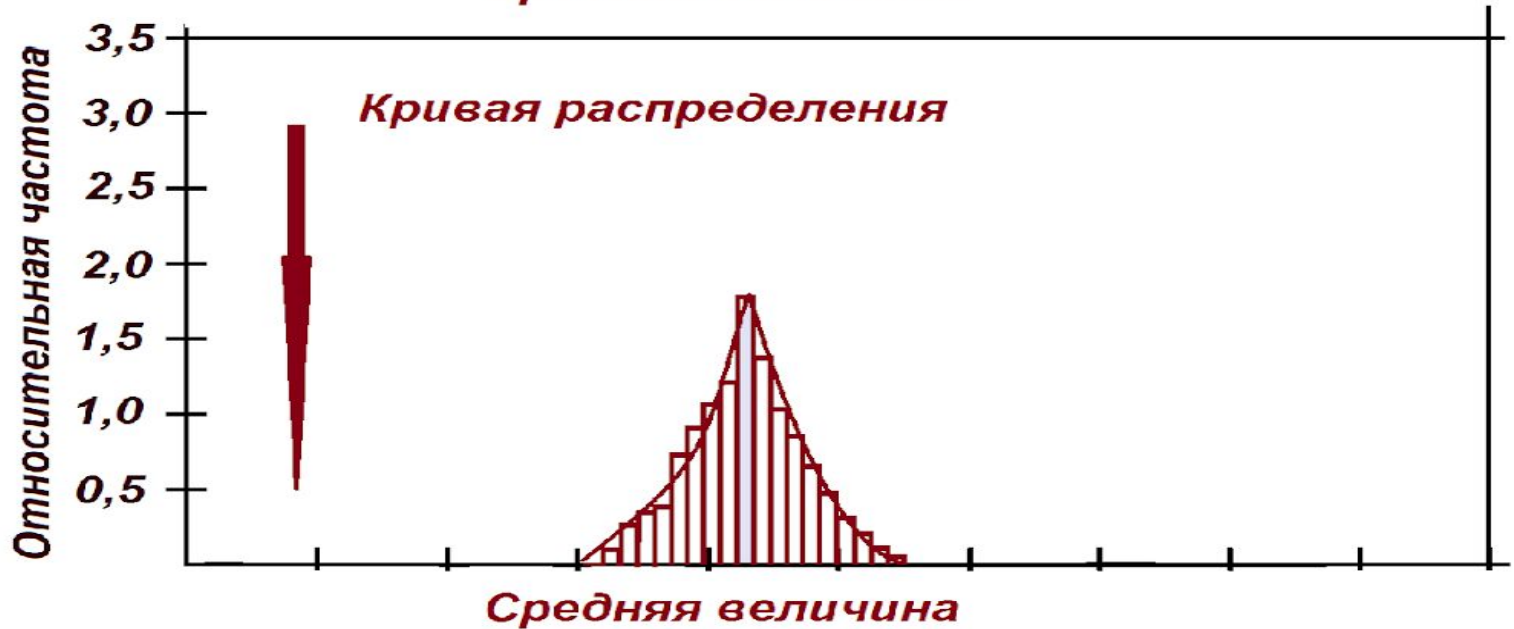
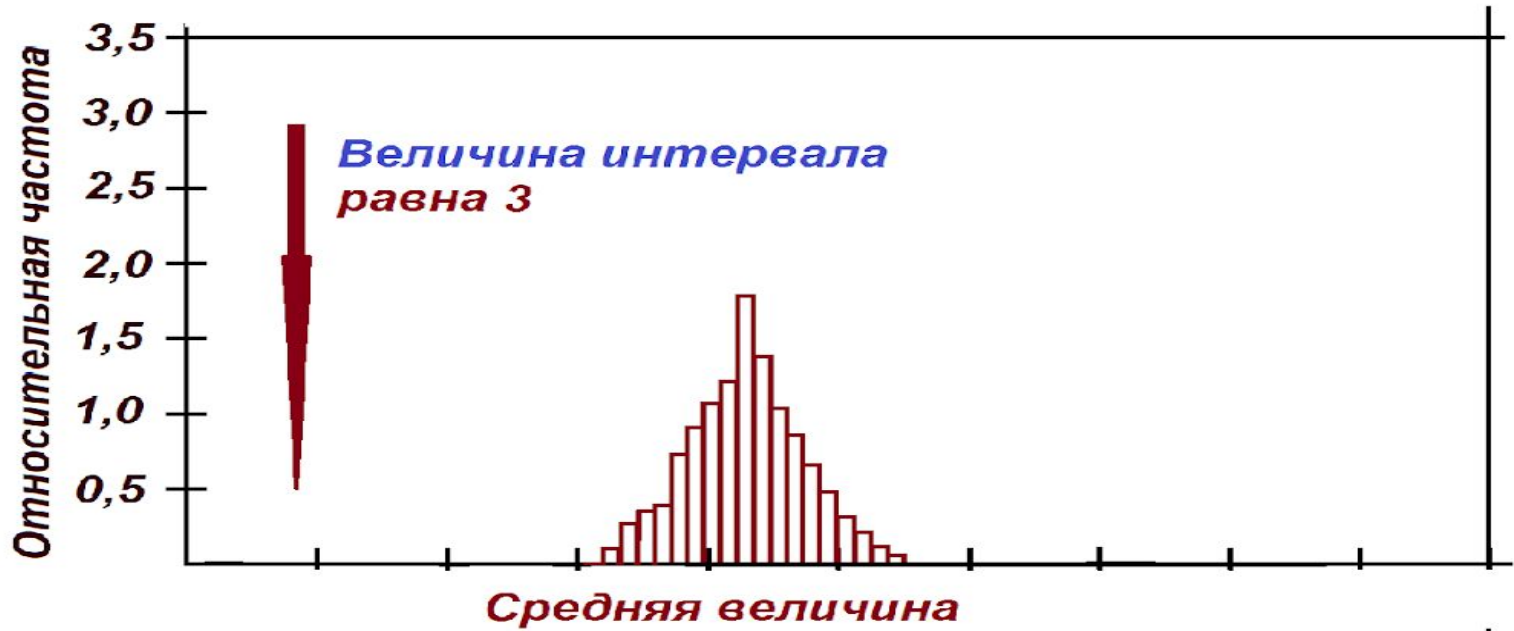


Виды гистограмм (продолжение)

Гистограммы с различными величинами интервала



Виды гистограмм (продолжение)



***** Напоминание** Хи-квадрат распределение
(χ^2 – распределение)

Если функция распределения вероятности выражается формулой

$$f(x) = \begin{cases} \frac{1}{2^{\frac{1+k}{2}} \cdot \int_0^{\infty} x^{\frac{k}{2}-1} \cdot e^{-x} dx} x^{\frac{k}{2}-1} \cdot e^{-x} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0 \end{cases}$$

то говорят, что величина x имеет распределение Хи-квадрат

(- распределение с числом степеней свободы k). **Например кривые распределения (рис. 15) для случаев, когда число степеней свободы равны 2, 10 и 20.**

***** Напоминание** Хи-квадрат распределение
(χ^2 – распределение) (продолжение)

Кривые – распределения

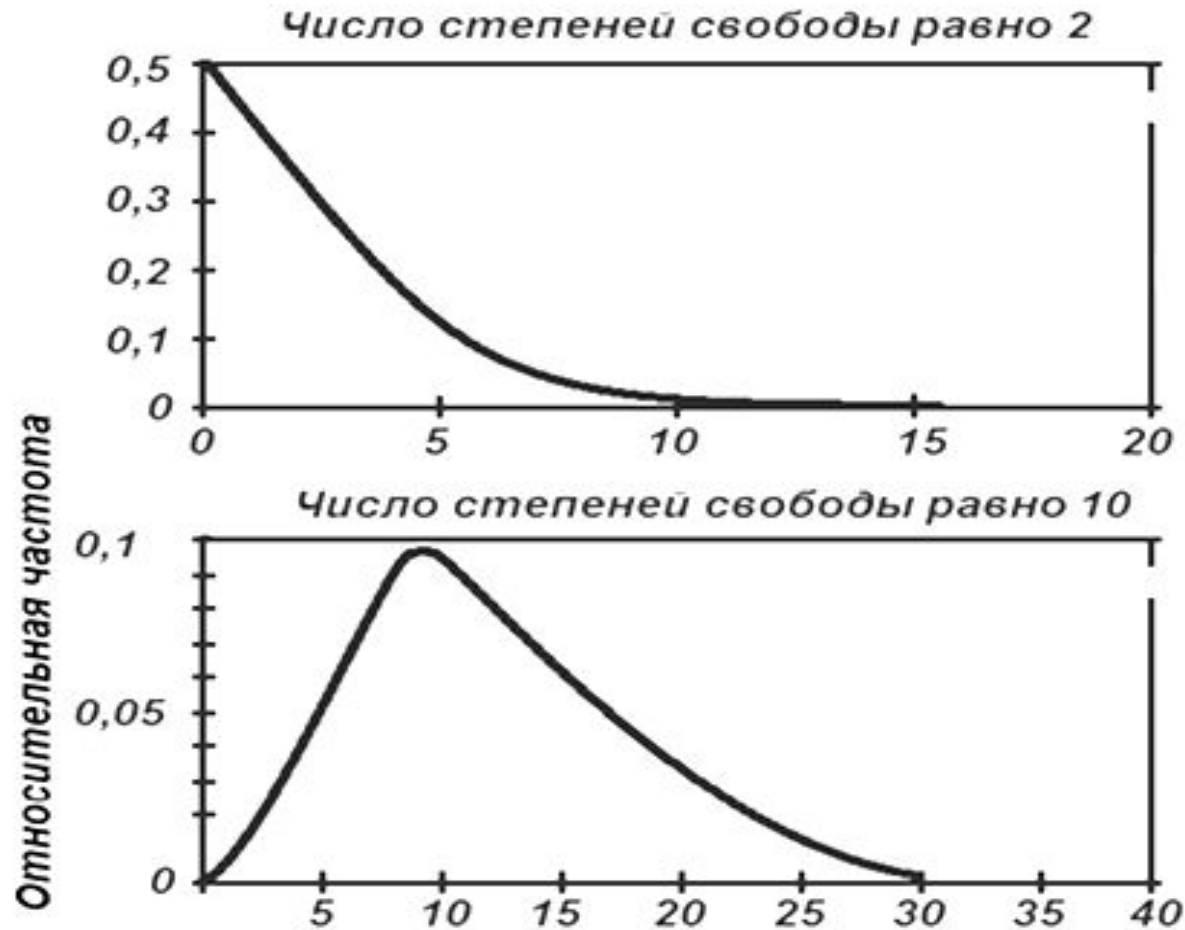


Рис. 15. Кривые распределения χ^2 с числом степеней свободы k , 2 и 10

***** Напоминание** **Хи-квадрат распределение**
(– распределение) (продолжение)

Кривые – распределения

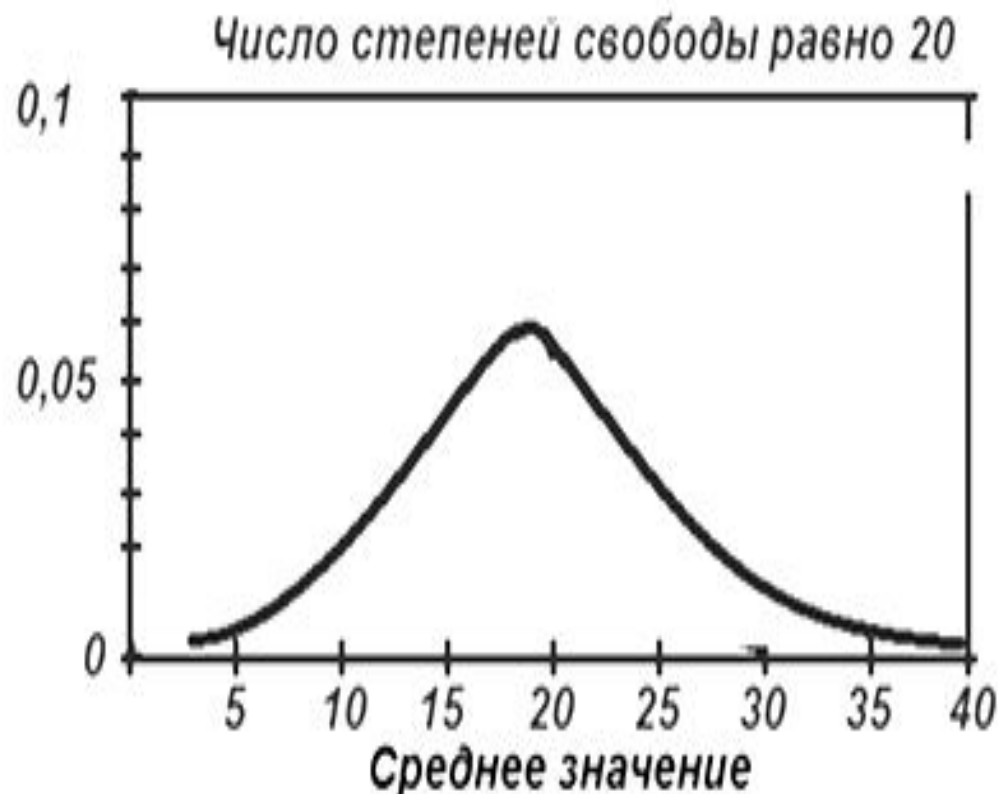


Рис. 15. Кривые распределения χ^2 с числом степеней свободы k , 20

Кривые χ^2 распределения

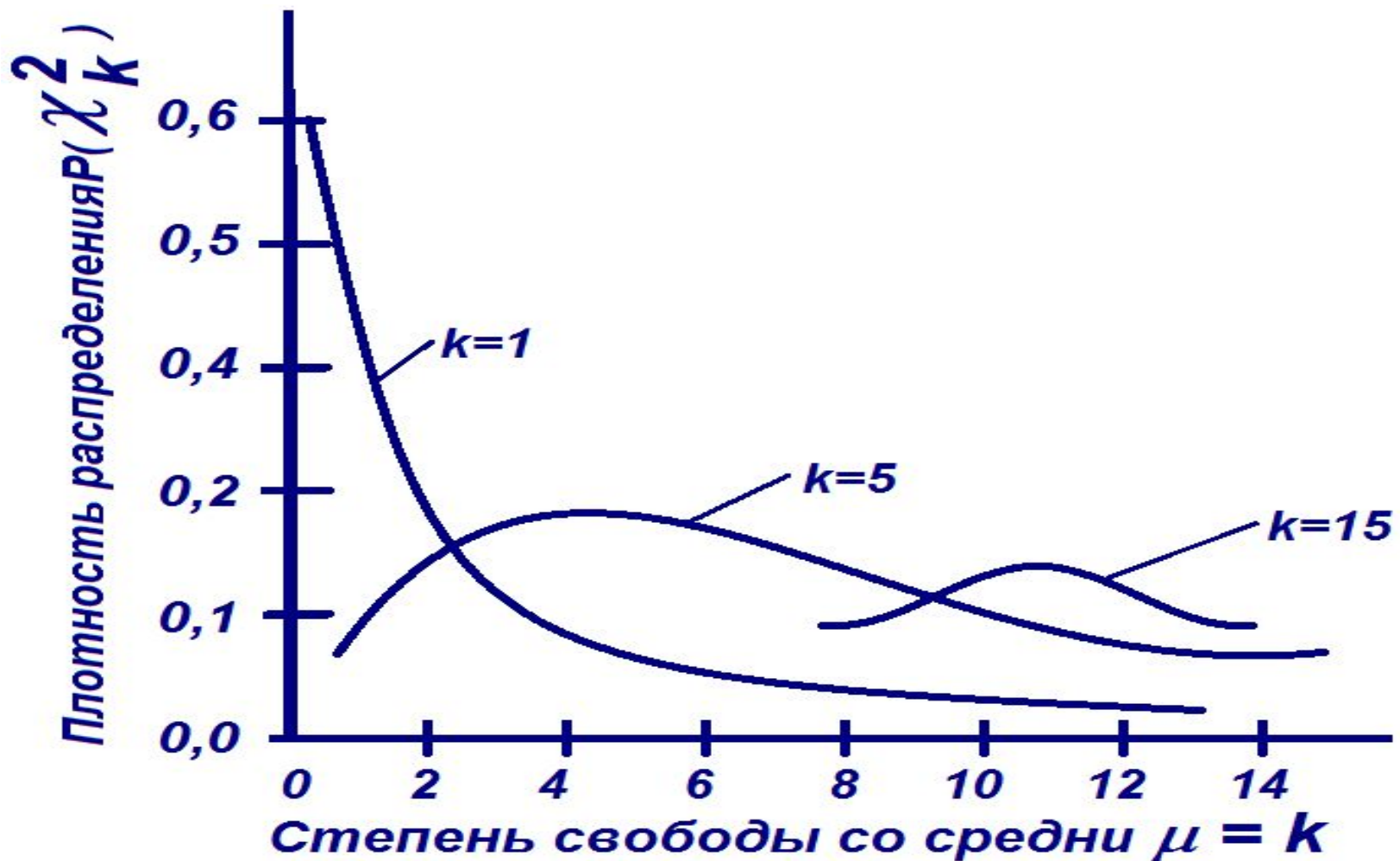


Рис. 16. Кривые χ^2 распределения со средней величиной $k=1, 5, 15$

*****Степень свободы. Характеристика формы графика χ^2 – распределения**

Поэтому, если изменить число степеней свободы форма графика тоже изменится. Степень свободы есть характеристика размера выборки. Чем больше выборка, тем больше степень свободы

Значения вероятности

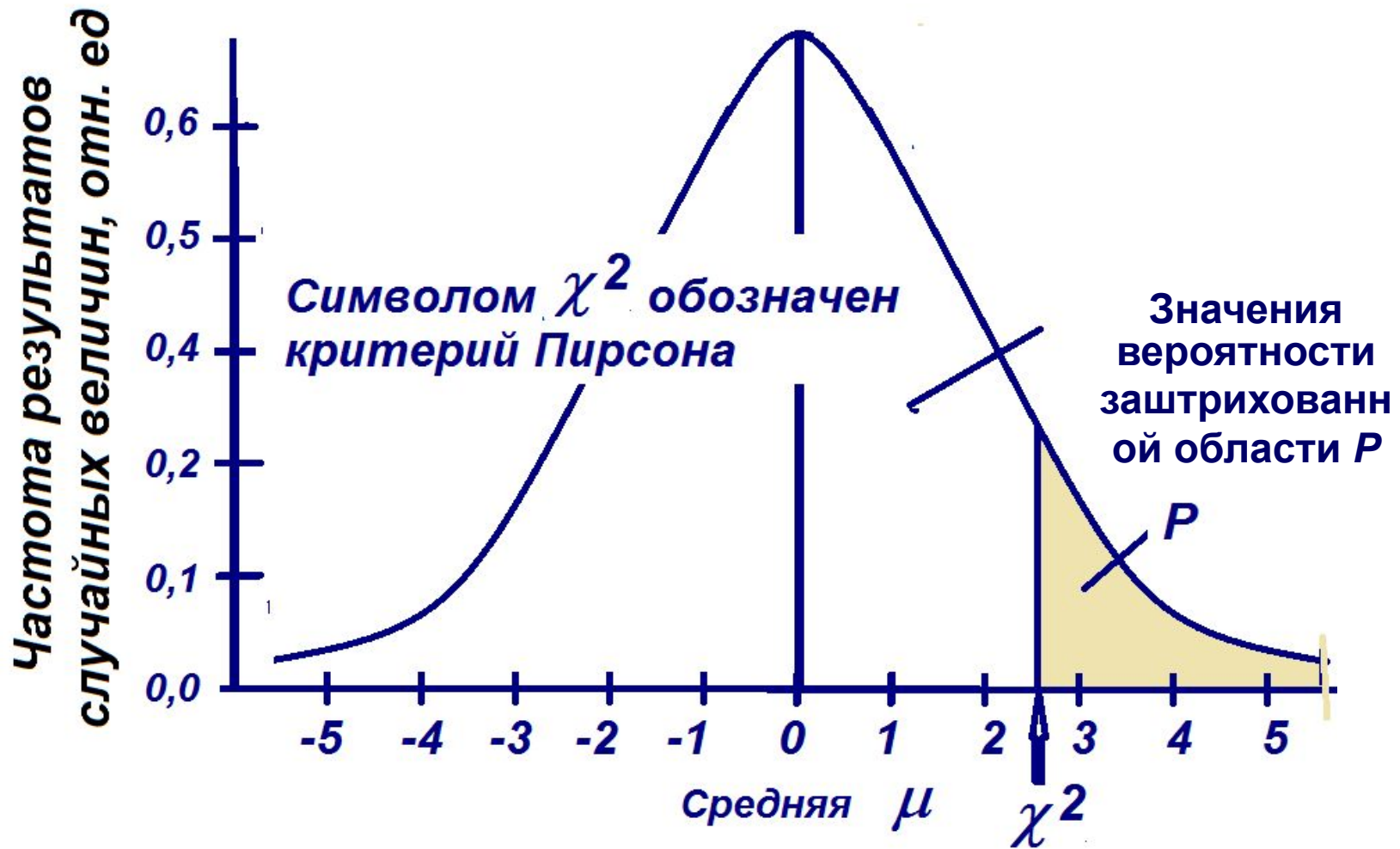


Рис. 17. Значения вероятности χ^2 – распределения

***** Так же как существует таблица стандартного нормального распределения, есть таблица и**

χ^2 – распределения.

Это таблица, в которой указываются значения χ^2

(см. ось x на графике, рис. 15 и 16, табл. 9) т. е. соответствует значению вероятности заштрихованной области P (площадь и доля – рис. 17)

Это сходно с таблицей стандартного нормального распределения. табл. 8)

Таблица - χ^2 распределения – табл.9

χ^2 – распределение

Таблица 9

степень свободы / P	0,995	0,99	0,975	0,95	0,05	0,025	0,01	0,005
1	0,000039	0,0002	0,0010	0,0039	3,8415	5,0239	6,6349	7,8794
2	0,0100	0,0201	0,0506	0,1026	5,015	7,3778	0,2104	10,5065
3	0,0717	0,1148	0,2158	0,3518	7,8147	9,3484	11,3449	12,8381
4	0,2070	0,2971	0,4844	0,7107	9,4877	11,1433	13,2767	14,8602
5	0,4118	0,5543	0,8312	1,1455	11,0705	12,8325	15,0863	16,7496
6	0,6757	0,8721	1,2373	1,6354	12,5916	14,4494	16,8119	18,5475
7	0,9893	1,2390	1,6899	2,1673	14,0671	16,0128	18,4753	20,2777
8	1,3444	1,6465	2,1797	2,7326	15,5073	17,5345	20,0902	21,9549
9	1,7349	2,0879	2,7004	3,3251	16,9190	19,0228	21,6660	23,5893
10	2,1558	2,5582	3,2470	3,9403	18,3070	20,4832	23,2093	25,1881
...

Если число степеней свободы равно «1», а величина « $P=0,05$ », то значение, находящееся на пересечении «1» и «0,05» (см. табл.9) будет равно **«3,8415»**.

Таблица стандартного нормального распределения позволяет по значению координаты (в пределах заштрихованной области рис. 18) найти соответствующую вероятность. По вертикали (рис. 18) определяется соответствующая координата на оси



Рис. 18. Значения вероятности χ^2 – распределения с оценкой вероятности

От сих

***** Распределение Стьюдента.
Распределение Фишера или
F-распределение**

Распределение Стьюдента. При оценке результатов экспертных заключений так же используется и такая формула распределения вероятностей

$$f(x) = \frac{\int_0^{\infty} x^{\frac{k+1}{2}-1} e^{-x} dx}{\sqrt{n\pi} \int_0^{\infty} x^{\frac{k}{2}-1} e^{-x} dx} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

где ***k*** – число степеней свободы. Если плотность вероятностей выражена данной формулой, то это означает, что величина имеет **распределение Стьюдента** с числом степеней свободы ***k***

(рис. 19).

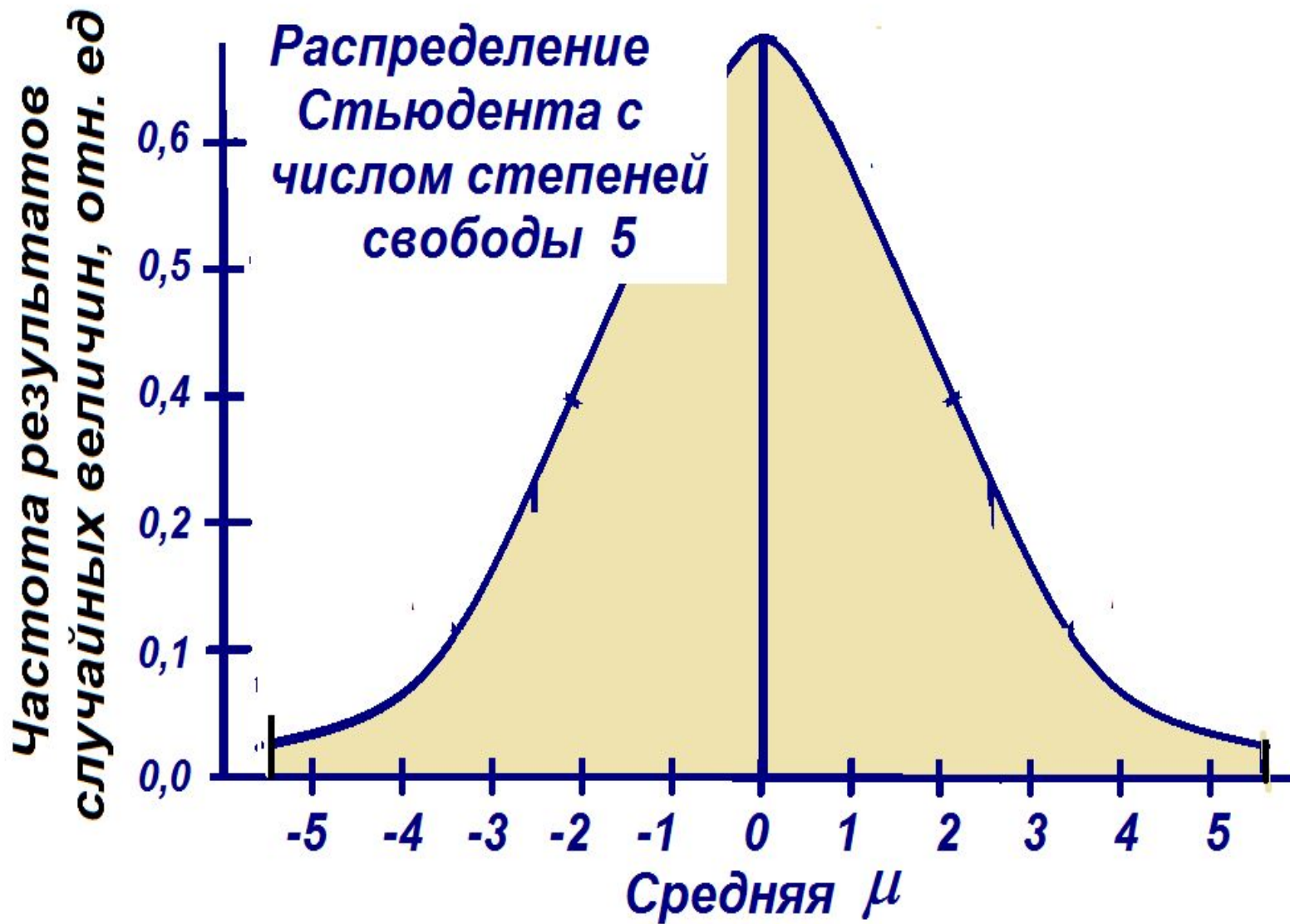


Рис. 19. Распределение Стьюдента

*** *Распределение Фишера.*

Не менее часто, при обработке данных, используется и такая формула распределения вероятностей

$$f(x) = \begin{cases} \frac{\left(\int_0^\infty x^{\frac{k+m}{2}-1} e^{-x} dx\right) * k^{\frac{k}{2}} * m^{\frac{m}{2}}}{\left(\int_0^\infty x^{\frac{k}{2}-1} e^{-x} dx\right) * \left(\int_0^\infty x^{\frac{m}{2}-1} e^{-x} dx\right)} * \frac{x^{\frac{k}{2}-1}}{(kx+m)^{\frac{k+m}{2}}} & \text{при } x > 0, \\ 0 & \text{при } x \leq 0 \end{cases}$$

где k и m – число степеней свободы величины x

Если плотность вероятностей выражена этой формулой, то это означает, что величина x имеет *распределение Стьюдента* с числом степеней свободы k и m (рис. 20)

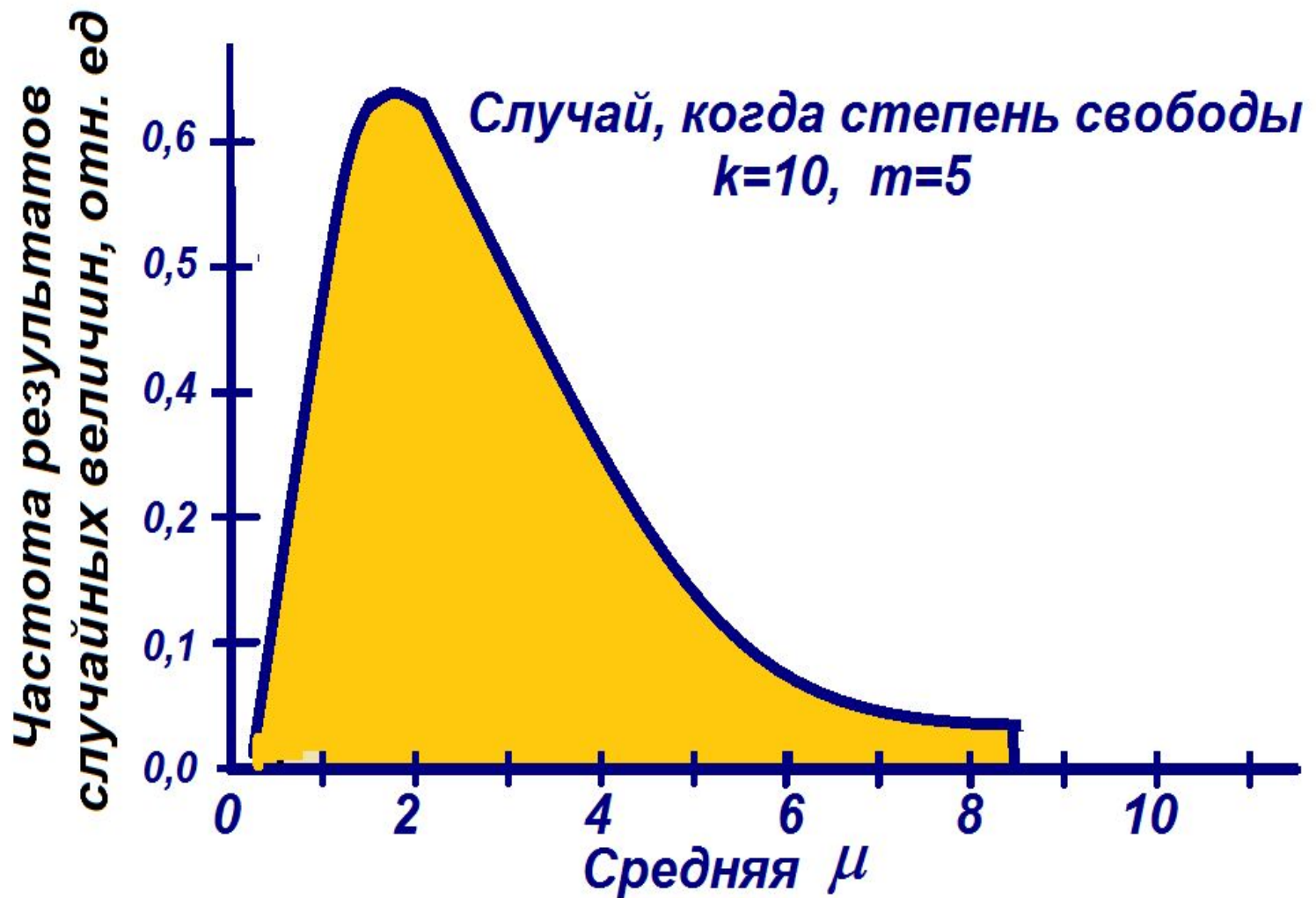


Рис. 20. Распределение Фишера

с χ^2 Пример случайной величины
– «хи-квадрат» распределением.

Если $y_1, y_2, y_3, \dots, y_n$ – случайная выборка из
распределения $N(\mu, \sigma^2)$, то

$$\frac{SS}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \bar{y})^2 \cong \chi_{n-1}^2$$

т.е. величина (отношение выборочной
дисперсии к квадрату дисперсии
отклонения) подчиняется
распределению

χ^2

– «хи-квадрат» с $n-1$ степенями свободы

Величина

$$SS = \sqrt{S^2} = \sqrt{(y_i - \bar{y})^2}$$

в числителе выражения

$$\frac{SS}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \bar{y})^2 \cong \chi_{n-1}^2$$

называется скорректированная сумма квадратов или сумма квадратов отклонений от среднего

Анализ уравнения

$$s = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$$

свидетельствует о том, что выборочную дисперсию возможно записать в виде

$$S^2 = \frac{1}{n-1} SS$$

Если наблюдения в выборке являются

$$NID(\mu, \sigma^2)$$

(распределённые по нормальному закону с нулевым средним и единичной дисперсией, что обозначается аббревиатурой

$$NID(0, 1)$$

то величина

$$S^2$$

распределение

рассматривается как

$$\left[\frac{\sigma^2}{n-1} - 1 \right] \chi_{n-1}^2$$

Таким образом, при условии, что исходная совокупность *соответствует нормальному закону*, распределение выборочной дисперсии

Таким образом, при условии, что исходная совокупность *соответствует нормальному закону*, распределение выборочной дисперсии

$$s^2 = \frac{1}{n-1} SS$$

отличается от распределения хи-квадрат – χ^2

лишь постоянным множителем

Если

$$z \text{ и } \chi_k^2$$

независимые случайные переменные
со стандартизованным нормальным

$$z = \frac{y - \mu}{\sigma}$$

и

$$\chi^2$$

«хи-квадрат распределением
соответственно,

то случайная величина

$$t_k = \frac{z}{\sqrt{\chi_k^2 / k}}$$

подчиняется

$$t_k$$

– распределению
(распределению Стьюдента)

с k степенями свободы и обозначением t_k

*** Плотность вероятности

t_k

имеет вид

$$f(t) = \frac{\Gamma[(k + 1/2)]}{\sqrt{k\pi} * \Gamma(k/2)} * \frac{1}{[(t^2/k) + 1]^{(k+1)/2}}, \quad -\infty < t < \infty,$$

причём *среднее и дисперсия*
соответственно

t_k

$$\mu = 0 \text{ и } \sigma^2 = \frac{k}{k-2} \text{ при } k > 2$$

(здесь Γ – гамма функция Эйлера*

(* Гамма функция – есть

$$\int_0^{\infty} e^{-x} \cdot x^{p-1} dx$$

4. ПРОВЕРКА ГИПОТЕЗ

Гипотеза (от греческого hypothesis – основание, предположение, догадка)

– это утверждение о значениях параметров распределения вероятностей (гипотеза в общем случае – суждение, относящееся к распределению случайной величины). Например, если предположить, **что средний выход продукта** какого-либо производства **составляет 94,5%**. Такое утверждение можно представить в формализованном виде как

$$H_0 : \mu = 94,5$$

$$H_1 : \mu \neq 94,5$$

нулевая гипотеза

альтернативная гипотеза

H_0

Утверждение

$$H_0 : \mu = 94,5$$

называется *нулевой гипотезой*,

 H_1

Утверждение

$$H_1 : \mu \neq 94,5$$

называется *альтернативной гипотезой*

Поскольку

 H_1

определяет значения

 μ

которые *либо больше, либо меньше 94,5* т.е. отношение

$$94,5 < \mu > 94,5$$

**есть двусторонняя
альтернатива**

Значение среднего

$$\mu = 94,5$$

утверждение

H_0

задаваемого нулевой
гипотезой

$$H_0: \mu = 94,5$$

определяется одним из трёх

- *среднее может быть известно* из результатов ранее проводившихся наблюдений;
- *среднее может быть известно* из теории исследуемого процесса (по полученной модели);
- *среднее может быть известно* из заданных условий

Проверка гипотезы состоит в следующем. Рассматривается случайная выборка наблюдений Y_i , по которой находится значение некоторой **«статистики»**, и принимается решение, отклонить или принять *нулевую гипотезу* –

$$H_0: \mu = 94,5.$$

Для этого необходимо знать распределение **«СТАТИСТИКИ»**, используемой для проверки, в предположении

«ИСТИННОСТИ»
нулевой гипотезы

$$H_0: \mu = 94,5.$$

а так же **«множество знаний статистики»**, которые привели бы к отклонению гипотезы

Такое *множество знаний статистики* называется **критической областью**, или **областью отклонения гипотезы**

При оценке гипотез встречаются погрешности («ошибки») *двух родов*:

- если нулевая гипотеза H_0 **отклоняется**, когда она **истина**, то совершается погрешность («ошибка») **1-го рода**;

- если нулевая гипотеза H_0 **не отклоняется**, когда она **ложна**, то совершается погрешность («ошибка») **2-го рода**;

Вероятностям этих погрешностей («ошибкам») присвоены специальные обозначения:

Если обозначить $\alpha = P$ – т. е. допустить («ошибку») погрешность 1-го рода, т. е. отклонить, гипотезу H_0 , когда она – *истина*;

Если обозначить $\beta = P$ – т. е. допустить («ошибку») погрешность 2-го рода, т. е. не отклонить, гипотезу H_0 , когда она – *ложна*;

Кроме того, часто применяется такое понятие, как **«мощность критерия»**, которое определяется как: **МОЩНОСТЬ = $1 - \beta = P$** (или $1 - \alpha$) отклонить, когда гипотеза H_0 ложна



4. ПРОВЕРКА ГИПОТЕЗ



При проверке гипотез в общем случае задаётся величина **α -вероятность** – погрешность («ошибка») **1-го рода**, которая называется **«УРОВНЕМ ЗНАЧИМОСТИ» критерия** и выбирается (задаётся) процедура проверки, обеспечивающую малую (приемлемую) величину («ошибки») погрешности 2-го рода, т. е. **β -вероятность**

Проверка гипотез относительно средних

Вот несколько часто встречающихся задач на проверку гипотез.

- Сравнение средних при известной дисперсии.**
- Сравнение средних при неизвестной дисперсии.**
- Сравнение дисперсий**

Случай 1

В начале, когда y_i есть нормальная случайная переменная с неизвестным средним μ , и известной дисперсией σ^2 . Необходимо проверить гипотезы: при $\alpha = 0,05$

$$H_0 : \mu = \mu_0;$$

$$H_1 : \mu \neq \mu_0;$$

где μ_0 – заданная средняя величина

Для проверки нулевой гипотезы

$$H_0: \mu = \mu_0$$

необходимо на основе выборки из **n -наблюдений y_i** найти численное значение **относительной (процентной) точки «статистики»**

$$Z_0 = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}$$

лежащей в основе критерия оценки.

Нулевая гипотеза отклоняется, если

$$H_0: \mu = \mu_0$$

$$|Z_0| > Z_{\alpha/2}$$

где $Z_{\alpha/2}$ верхняя $\alpha/2$ относительная (процентная) точка стандартизованного нормального распределения

Такая процедура проверки обоснуется следующим образом:
 1. В соответствии с **«центральной предельной теоремой»** выборочное среднее есть

$$\bar{y} = N(\mu, \sigma^2/n)$$

Поэтому, если H_0 - **«истина»**, то величина Z_0 из соотношения

$$Z_0 = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$$

подчиняется закону

$$N(0, 1)$$

и можно ожидать, что $(1-\alpha)$ – **доверительная вероятность** или $100*(1-\alpha)$ – **доверительных процентов** значений Z_0 попадут в интервал между

$$-Z_{\alpha/2} \text{ и } Z_{\alpha/2}$$

2. Появление выборки, для которой Z_0 лежит вне этого интервала, т. е. интервала ($-Z_0$ и $Z_{\alpha/2}$), было бы условием истинности нулевой гипотезы $H_0: \mu = \mu_0$, чем-то необычным и дало бы основания для отклонения нулевой гипотезы H_0 .

Необходимо отметить, что $\alpha = P$ – допустить «ошибку» (погрешность) 1-го рода здесь используется как критерий оценки

При решении ряда задач может оказаться желательным **отклонять нулевую гипотезу H_0** , только при условии, что истинное значение среднего μ **превосходит μ_0** , т. е. **может быть принята гипотеза H_1** :

$$H_1: \mu > \mu_0$$

В этом случае формулируется **односторонняя альтернатива, т. е.**

$$H_1: \mu > \mu_0$$

Тогда нулевая гипотеза $H_0: \mu = \mu_0$ **отклоняется при доверительных процентах (или $(1-\alpha)$ – доверительной вероятности) $Z_0 > Z_\alpha$**

***Если же необходимо отклонить нулевую гипотезу $H_0: \mu = \mu_0$

только при $\mu < \mu_0$, то в качестве альтернативной гипотезы принимается $H_1: \mu > \mu_0$, а нулевая гипотеза $H_0: \mu = \mu_0$ отклоняется, если выполняется соотношение «доверительных процентов» значений $Z_0 < -Z_\alpha$

Процедура проверки рассмотренных гипотез приведена в табл. 19.

Таблица 19

Проверка гипотез относительно средних при известной дисперсии

Оцениваемые гипотезы	Статистика для проверки (нормальное распределение)	Критерии отклонения
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$	$Z_0 = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}$	$ Z_0 > Z_{\alpha/2}$
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$		$Z_0 < -Z_{\alpha}$
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$		$Z_0 > Z_{\alpha}$

***4. ПРОВЕРКА ГИПОТЕЗ

Оцениваемые гипотезы	Статистика для проверки (нормальное распределение)	Критерии отклонения
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 \neq \gamma \end{cases}$	$Z_0 = \frac{\bar{y}_1 - \bar{y}_2 - \gamma}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$ Z_0 > Z_{\alpha/2}$
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 < \gamma \end{cases}$		$Z_0 < -Z_{\alpha}$
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 > \gamma \end{cases}$		$Z_0 > Z_{\alpha}$

Пример 1. Оценка гипотез на основе стандартизованного нормального распределения

Предприятие, которое реализует **ВОЛОКНО**, интересуется, превосходит ли средняя цена 1м^2 за партию в $200\ \$$. Известно: цена 1м^2 волокна составляет $200\ \$$; стандартное отклонение от цены составляет $10\ \$$ за 1м^2 . Выборка цены волокна составила 4 ; выборочное среднее цены равно $214\ \$$ за 1м^2 .

Необходимо проверить гипотезу

$$\begin{cases} H_0: \mu = 200\$; \\ H_1: \mu > 200\$ \end{cases} \text{ при } \alpha = 0,05$$

Решение

1. Числовое значение («информации») статистики, используемое для проверки нулевой гипотезы

$$H_0: \mu = 200\$$$

равно

$$Z_0 = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{214 - 200}{10/4} = 2,800$$

Если по условию задаётся величина погрешности (вероятности) 1-го рода – «ошибка 1-го рода» при $\alpha=0,05$

то в соответствии с табл. 2, (кумулятивная функция стандартизированного нормального распределения)

$$Z_{\alpha=0,05} = 0,9974 \pm (0,81)$$

Т. е. числовое значение статистики (табл. 2), которое можно представить как

$$Z_0 = 2,80 + 0,00 = 2,800$$

Десятые и сотые доли величины Z необходимо представить так, чтобы по горизонтали (табл. 2)

$$Z = 0,000,$$

$$\text{а по вертикали: } Z = 2,800.$$

На пересечении горизонтальных и вертикальных значений величины Z в соответствующей ячейки

стоит **число 0,99744.**

Это и будет площадь заштрихованного участка под кривой в относительных единицах **за минусом 0,5** – соответствующей площади под кривой. С величинами средней μ (все заданные средние значения по условию задачи, положительные величины)

Таким образом, площадь под кривой стандартного нормального распределения (см. рис.29) будет равна

$$0,99744 - 0,5 = 0,49744 = Z_{\alpha=0,05}$$



Это и будет площадь заштрихованного участка под кривой в относительных единицах за минусом 0,5 с величинами средней μ (все заданные средние значения по условию задачи, положительные величины)

Рис. 29. Гистограмма с оценкой площади (0,49744)

2. Таким образом, нулевая гипотеза

$$H_0: \mu = 200 \$$$

«отклоняется»,

так как представленный критерий отклонения оценки гипотезы,

$$\begin{cases} H_0: \mu = 200 \$; \\ H_1: \mu > 200 \$ \end{cases}$$

удовлетворяет условию отклонения статистики (см. табл.19):

$$Z_0 > Z_\alpha$$

$$Z_0 = 2,800 > Z_{\alpha=0,05} = 0,49744$$

И вывод заключается в том, что средняя цена по партии превосходит 200 \$ за 1 м² волокна (см. табл. Ex1)

Пример 2

Оценка гипотез на основе стандартизованного нормального распределения

Предприятие, которое реализует волокно, интересуется, превосходит ли средняя цена 1м^2 за партию **в 204 \$**. Известно: цена 1м^2 волокна составляет **197 \$**;
стандартное отклонение от цены составляет **25 \$ за 1м^2** . Выборка цены волокна составила **4**; выборочное среднее цены равно **204 \$ за 1м^2** . Необходимо проверить гипотезу

$$\begin{cases} H_0: \mu = 197\$; \\ H_1: \mu < 197\$ \end{cases}$$

при

$$\alpha = 0,05$$

Решение (пример 2)

1. Числовое значение статистики Z_0 , используемое для проверки нулевой гипотезы $H_0: \mu = 197\$$ равно

$$Z_0 = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{204 - 197}{25/4} = 0,560$$

Если по условию задаётся величина («ошибки») погрешности (вероятности)

1-го рода – $\alpha = 0,05$, то в соответствии с табл. 2, числовое значение Z_0 равно **0,21226*****

$$Z_{\alpha=0,05} = 0,71226 - 0,5 = 0,21226$$

***На пересечении горизонтальных и вертикальных значений ($Z = 0,060$ – по горизонтали; $Z = 0,5$ – по вертикали) **стоит число 0,71226**

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490

$$Z_0 = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{204 - 197}{25/4} = 0,560$$

Комментарии

Числовое значение статистики (см. табл.2) можно представить как $Z_0 = 0,5 + 0,06 = 0,560$.

Тогда $Z_{\alpha=0.05} = 0,21226$ (см рис. 30)

$$Z_{\alpha=0,05} = 0,71226 - 0,5 = 0,21226$$

Это и будет площадь заштрихованного участка под кривой (см. рис.30) в относительных единицах **за минусом 0,5** – соответствующей площади под кривой с отрицательными величинами средней μ (все заданные средние значения по условию задачи, положительные величины). Таким образом, площадь под кривой стандартного нормального распределения (см. рис. 30) будет **равна 0,21266**

4. ПРОВЕРКА ГИПОТЕЗ

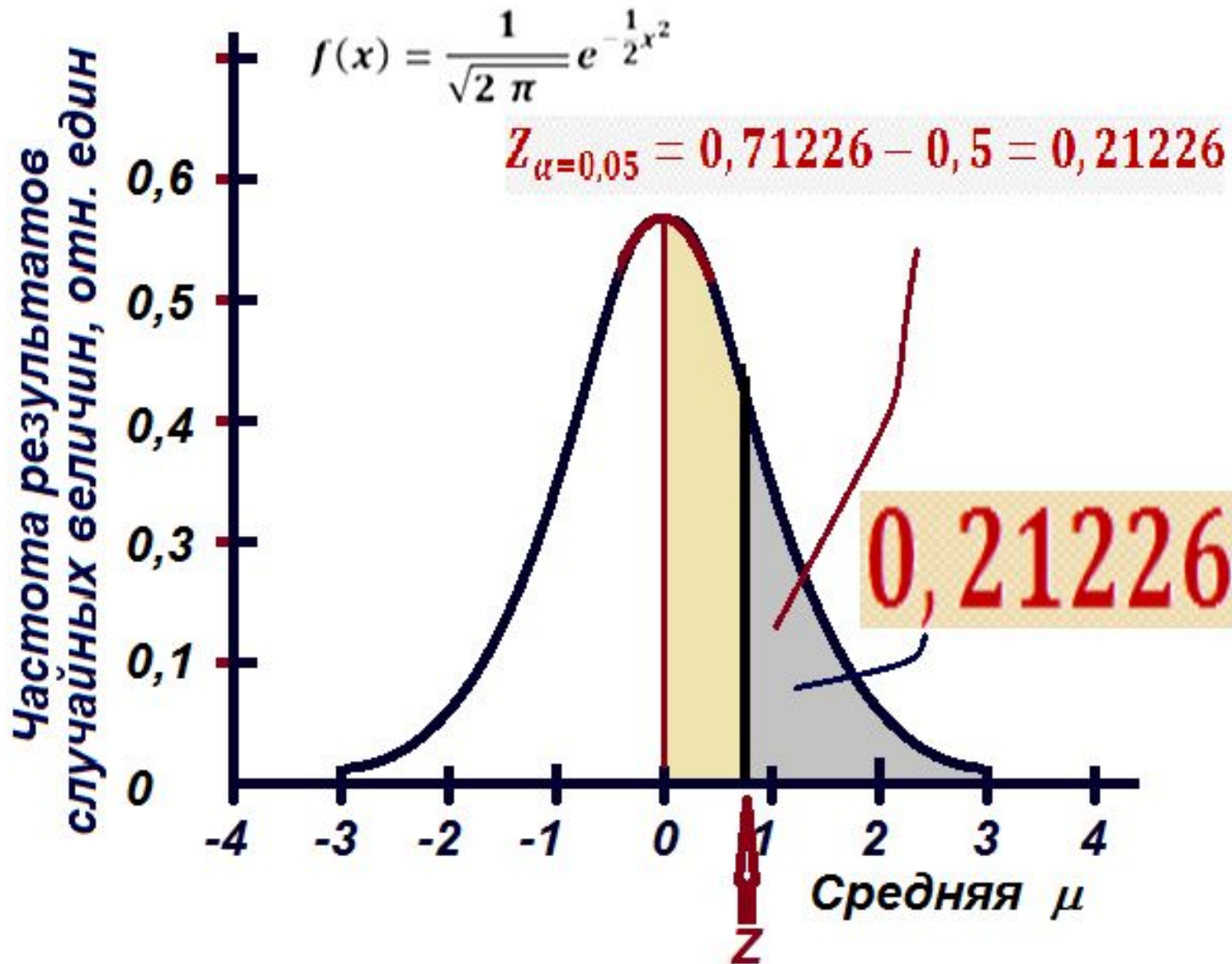


Рис.30. Гистограмма с оценкой площади $Z_{\alpha=0,05}=0,21226$

2. Таким образом, альтернативная гипотеза

$$H_1: \mu < 197$$

«не отклоняется»,

так как представленный «критерий отклонения» оценки гипотезы

$$\begin{cases} H_0: \mu = 197\$; \\ H_1: \mu < 197\$ \end{cases}$$

не удовлетворяет условию отклонения статистики (см. табл.19):

$$Z_0 < -Z_\alpha$$

$$Z_0 = 0,560 > Z_{\alpha=0,05} = 0,21226$$

И вывод заключается в том, что средняя цена по партии волокна **не будет превосходить 197 \$ за 1 м².**



***4. ПРОВЕРКА ГИПОТЕЗ (продолжение по t-критерию)

4. ПРОВЕРКА ГИПОТЕЗ(продолжение)

Если дисперсия распределения совокупности неизвестна, то приходится делать «дополнительные предложения о нормальности распределения». При этом имеется ввиду то, что небольшие отклонения от «нормальности» не приводят к существенному (значимому) искажению результата.

При проверке гипотез

$$\begin{cases} H_0: & \mu = \mu_0; \\ H_1: & \mu \neq \mu_0 \end{cases}$$

в случае неизвестной дисперсии для оценки σ^2 используется выборочная дисперсия S^2

Широта (диапазон) распределения вероятностей или рассеивание случайной величины может характеризоваться дисперсией (раздробленностью), которая определяется как

$$\sigma^2 = \begin{cases} \int_{-\infty}^{\infty} (y - \mu)^2 f(y) dy, & y - \text{непрерывна,} \\ \sum_{\text{по всему } y} (y - \mu)^2 p(y), & y - \text{дискретна} \end{cases}$$

Величина

$$SS = \sqrt{S^2} = \sqrt{(y_i - \bar{y})^2}$$

**называется скорректированная
сумма квадратов или сумма
квадратов отклонений от среднего**

*****Анализ уравнения**

$$s = \sqrt{S^2} = \sqrt{\frac{1}{1-n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

**свидетельствует о том, что
выборочную дисперсию возможно
записать в виде**

$$S^2 = \frac{1}{n-1} SS$$

Т. е. используется выборочная дисперсия

$$S^2$$

используется выборочная дисперсия

$$S^2$$

Заменяя в выражении относительной (процентной) точки статистики

$$Z_0 = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}$$

«дисперсию»

$$\sigma^2$$

на «выборочную дисперсию» S^2 ,

получается статистика для проверки гипотезы

$$t_0 = \frac{\bar{y} - \mu_0}{S / \sqrt{n}}$$

Тогда, нулевая гипотеза $H_0: \mu = \mu_0$
отклоняется, если выполняются
условия

$$|t_0| > t_{\alpha/2; n-1}$$

где $t_{\alpha/2; n-1}$ – верхняя – $\alpha/2$
относительная (процентная) точка t -
распределения
с $n - 1$ степенями свободы

Условия проверки рассмотренных
гипотез приведены в табл. 20.

Проверка гипотез относительно средних нормально распределённых совокупностей при неизвестной дисперсии

Таблица 20

***4. ПРОВЕРКА ГИПОТЕЗ (продолжение по t-критерию)

Оцениваемые гипотезы	Статистика для проверки (нормальное распределение)	Критерии отклонения
$\begin{cases} H_0: \mu = \mu_0 \\ H_0: \mu \neq \mu_0 \end{cases}$	$t_0 = \frac{\bar{y} - \mu_0}{S/\sqrt{n}}$	$ t_0 > t_{\alpha/2;n-1}$
$\begin{cases} H_0: \mu = \mu_0 \\ H_0: \mu < \mu_0 \end{cases}$		$t_0 < -t_{\alpha;n-1}$
$\begin{cases} H_0: \mu = \mu_0 \\ H_0: \mu > \mu_0 \end{cases}$		$t_0 > t_{\alpha;n-1}$

Проверка гипотез относительно средних нормально распределённых совокупностей при неизвестной дисперсии

Таблица 20 (продолжение)

4. ПРОВЕРКА ГИПОТЕЗ (продолжение)

Оцениваемые гипотезы	Статистика для проверки (нормальное распределение)	Критерии отклонения
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 \neq \gamma \end{cases}$	<p>при $\sigma_1^2 = \sigma_2^2$</p>	$ t_0 > t_{\alpha/2;v}$
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 < \gamma \end{cases}$	$t_{0(s_p)} = \frac{\bar{y}_1 - \bar{y}_2 - \gamma}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$t_0 < -t_{\alpha;v}$
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 > \gamma \end{cases}$	$v(s_p) = n_1 + n_2 - 2,$ <p>где</p>	$t_0 > t_{\alpha;v}$
	$S_p^2 = \frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2}$	

Проверка гипотез относительно средних нормально распределённых совокупностей при неизвестной дисперсии

Таблица 20 (продолжение)

4. ПРОВЕРКА ГИПОТЕЗ (продолжение)

Оцениваемые гипотезы	Статистика для проверки (нормальное распределение)	Критерии отклонения
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 \neq \gamma \end{cases}$	<p>при $\sigma_1^2 \neq \sigma_2^2$</p> $t_{0(s)} = \frac{\bar{y}_1 - \bar{y}_2 - \gamma}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$ t_0 > t_{\alpha/2;v}$
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 < \gamma \end{cases}$	$v(s) = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1+1} + \frac{(S_2^2/n_2)^2}{n_2+1}} - 2$	$t_0 < -t_{\alpha;v}$
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 > \gamma \end{cases}$	$v(s) = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1+1} + \frac{(S_2^2/n_2)^2}{n_2+1}} - 2$	$t_0 > t_{\alpha;v}$

Например, есть две совокупности, распределённые нормально с неизвестными средними μ_1 μ_2 и неизвестными дисперсиями

$$\sigma_1^2 \text{ и } \sigma_2^2$$

В этом случае процедура проверки гипотез

$$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 \neq \gamma \end{cases}$$

для таких совокупностей будет зависеть от выполнения условия

(равенства или неравенства дисперсий)

$$\sigma_1^2 = \sigma_2^2 \text{ или } \sigma_1^2 \neq \sigma_2^2$$

Случай 1, когда выполнено условие

$$\sigma_1^2 = \sigma_2^2$$

Для проверки нулевой гипотезы

$$H_0: \mu_1 - \mu_2 = \gamma$$

необходимо взять два случайных информационных объёма n_1 и n_2 из первого и второго соответственно.

Затем произвести расчёт дисперсии

$$S_p^2 = \frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2}$$

и числовое значение статистики

$$t_{0(S_p)} = \frac{\bar{y}_1 - \bar{y}_2 - \gamma}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

**при числе степеней свободы
t-распределения**

$$V(S_p) = n_1 + n_2 - 2$$

**где S_1^2 и S_2^2 – критерии оценки
выборочной дисперсии**

Нулевая гипотеза

$$H_0: \mu_1 - \mu_2 = \gamma$$

из совокупности

$$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 \neq \gamma \end{cases}$$

«отклоняется», если будет выполнено условие по «критерию отклонения»

$$|t_0| > t_{\alpha/2; \nu} \quad (\text{см. табл.20})$$

Если нет оснований предполагать, что дисперсии одинаковы, т. е. $\sigma_1^2 \neq \sigma_2^2$, то в этом случае статистика для проверки гипотезы рассчитывается по формуле (см. табл. 20):

$$t_{0(s)} = \frac{\bar{y}_1 - \bar{y}_2 - \gamma}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

при числе степеней свободы t -распределения

$$v(s) = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 + 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 + 1}} - 2$$

Такую процедуру проверки называют объединённым t -критерием, так как оба объёма информации объединены для получения оценки общей дисперсии

Другой случай. Пусть y_1 и y_2 есть две нормальные случайная переменная с неизвестными средними μ_1 и μ_2 , отличающиеся друг от друга на постоянную величину γ и известными дисперсиями σ_1^2 и σ_2^2

Необходимо проверить гипотезу, которая в формализованной интерпретации имеет вид

$$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 \neq \gamma \end{cases}$$

В этом случае для проверки гипотезы

$$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 \neq \gamma \end{cases}$$

определяется случайная выборка из n_1 -наблюдений из первой совокупности и n_2 -наблюдений из второй совокупности, после чего находится относительное (процентное) численное значение «статистики»

$$Z_0 = \frac{\bar{y}_1 - \bar{y}_2 - \gamma}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Причём σ_i^2 равна

$$\sigma_i^2 = \frac{1}{n_i} \sum_{i=1}^n (\bar{y} - y_i)^2$$

Гипотеза

$$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 \neq \gamma \end{cases}$$

**отклоняется», если
выполняется условие**

$$|Z_0| > Z_{\alpha/2}$$

При одной односторонней «альтернативе»

$$H_1: \mu_1 - \mu_2 > \gamma$$

нулевая гипотеза $H_0: \mu_1 - \mu_2 = \gamma$

отклоняется, если $Z_0 > Z_\alpha$

При другой односторонней «альтернативе»

$$H_1: \mu_1 - \mu_2 < \gamma$$

нулевая гипотеза $H_0: \mu_1 - \mu_2 = \gamma$

отклоняется, если $Z_0 < -Z_\alpha$

Процедура проверки рассмотренных гипотез приведена в табл. 19 (окончание)

***4. ПРОВЕРКА ГИПОТЕЗ (продолжение по t-критерию)

Оцениваемые гипотезы	Статистика для проверки (нормальное распределение)	Критерии отклонения
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 \neq \gamma \end{cases}$	$Z_0 = \frac{\bar{y}_1 - \bar{y}_2 - \gamma}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$ Z_0 > Z_{\alpha/2}$
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 < \gamma \end{cases}$		$Z_0 < -Z_{\alpha}$
$\begin{cases} H_0: \mu_1 - \mu_2 = \gamma \\ H_1: \mu_1 - \mu_2 > \gamma \end{cases}$		$Z_0 > Z_{\alpha}$

Случай 1 при t -критерии, когда выполнено условие

$$\sigma_1^2 = \sigma_2^2.$$

Для проверки нулевой гипотезы

$$H_0: \mu_1 - \mu_2 = \gamma$$

необходимо взять два случайных информационных объёма n_1 и n_2 из первого и второго соответственно.

Затем производится расчёт дисперсии

$$S_p^2 = \frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2}$$

и числовое значение статистики

$$t_{0(S_p)} = \frac{\bar{y}_1 - \bar{y}_2 - \gamma}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

при числе степеней свободы t -распределения

Затем производится расчёт дисперсии

$$S_p^2 = \frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2}$$

и числовое значение статистики

$$t_{0(S_p)} = \frac{\bar{y}_1 - \bar{y}_2 - \gamma}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

**при числе степеней свободы
t-распределения**

$$V(S_p) = n_1 + n_2 - 2,$$

где S_1^2 S_2^2 - критерии оценки выборочной дисперсии

Пример. Оценка равенства гипотезы дисперсии совокупности, распределённой нормально

В отличие от процедур сравнения средних процедура сравнения дисперсии чувствительны к допущению о нормальности

Удалось получить две информации о стоимости товара конкурентов. Одна информация по 8 источникам, другая - по 4.

В формализованном виде это можно представить так:

*$n_1=8$ и $n_2=4$
(см. табл. ¹Исходных данных)*

табл. @@@. Исходные данные)

*** Оценка равенства гипотезы дисперсии совокупности, распределённой нормально

Величина	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀	Y ₁₁	Y ₁₂	Y ₁₃	Y ₁₄	Y ₁₅	Y ₁₆
Y _{i-1}	275	270	280	275	285	265	275	275	0	0	0	0	0	0	0	0
Y _{i-2}	250	260	255	265	0	0	0	0	0	0	0	0	0	0	0	0
кол-во выборки n _{Yi-1}	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
кол-во выборки n _{Yi-2}	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0

Заданные критерии	
γ	α
7,000	0,08

Расчётные величины по исходным данным							
Кол-во выборки n (количество информации)	Y _(ср)	σ_1^2	σ_2^2	S ₁ ²	S ₂ ²	μ_1	μ_2
8	275,00	31,250		35,714		н/н	н/н
4	257,50		31,250		41,667		
	Расчётные формулы						

$$\sigma_i^2 = \frac{1}{n_i} \sum_{i=1}^{n_i} (\bar{y}_{n_i} - y_{i(n_i)})^2 \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{y} - y_i)^2$$

См. расчётную таблицу





