**Введение в теорию статистического вывода**

Недавно прочитал замечательную книгу [Хитосе Кумэ. Статистические методы повышения качества](http://baguzin.ru/wp/?p=15984). Книга в первую очередь адресована практикам, причем не только менеджерам, но и рабочим. В книге раскрыты методы сбора и обработки данных для управления качеством (см. также [Семь основных инструментов контроля качества](http://baguzin.ru/wp/?p=1332)). В книге содержится одна глава, посвященная статистическому выводу. С одной стороны, она несколько выбивается из общей канвы книги, с другой стороны, в ней содержится краткий обзор по теме, причем изложенный весьма доходчиво (более подробно см. [Левин. Статистика для менеджеров с использованием Microsoft Excel](http://baguzin.ru/wp/?p=5285)). Собственно, перед вами эта глава с небольшими сокращениями и моими комментариями.

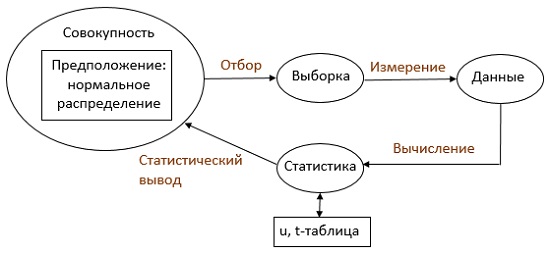


Рис. 1. Место статистического вывода в наблюдениях и вычислениях

### Статистика, как функция от результатов наблюдений

Когда мы хотим оценить среднее генеральной совокупности, мы берем множество наблюдений из этой совокупности и вычисляем их среднее. Вычисленное по выборке значение, такое, как выборочное среднее, называется статистикой. Иначе говоря, *статистика* – это функция от результатов наблюдений.

Надо отличать статистику от параметра генеральной совокупности. Для этого мы вместо термина «среднее» часто используем термины выборочное среднее и генеральное среднее. *Параметр совокупности* – это некоторое постоянное значение, которое, однако, фактически неизвестно. С другой стороны, мы всегда можем вычислить статистику по выборке, но эта статистика будет варьировать от выборки к выборке. Хотя мы хотим знать параметр совокупности, нам доступна только выборка, получаемая из этой совокупности. Следовательно, мы вынуждены оценивать параметр совокупности, основываясь на статистике (рис. 1).

### Распределение статистики

Для статистического вывода надо знать распределения различных статистик. Проведем эксперимент. Воспользовавшись компьютером, генерируем множество выборок из 5 случайных чисел, подчиняющихся нормальному распределению со средним µ = 50 и дисперсией σ2 = 22, или в краткой нотации – N(50, 22). Заметим, что µ и σ2 – параметры генеральной совокупности. В Excel для получения одного целого случайного числа, соответствующего распределению N(50, 22) воспользуемся формулой =ОКРУГЛ(НОРМ.ОБР(СЛЧИС();50;2);0). На рис. 2 представлено 10 случайных выборок. Также можно увидеть следующие статистики (заметим, что все они относятся к конкретным выборкам, поэтому и варьируются от строки к строке): среднее (х̅), размах (R), дисперсию (s2), среднеквадратичное отклонение (s), и два стандартных средних (u, t). Слово «стандартное» означает, что они приведены к µ = 0, σ2 = 1. Подробнее о различиях в вычислении u и t:

где n – размер выборки.

Т.е., при вычислении *u* мы основываемся на параметре генеральной совокупности – σ, а при вычислении *t* – статистике выборки *s*.

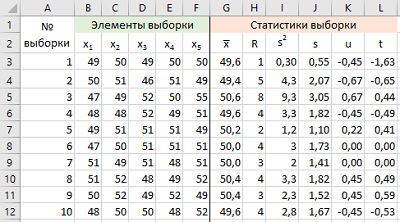


Рис. 2. Результаты выборочного эксперимента в Excel

Для первой выборки (строка 2) формулы в Excel:

х̅ = СРЗНАЧ(B2:F2)

R = МАКС(B2:F2)-МИН(B2:F2)

s2 =ДИСП.В(B2:F2)

s =СТАНДОТКЛОН.В(B2:F2)

(Подробнее о последней функции см. [СТАНДОТКЛОН.В и СТАНДОТКЛОН.Г: в чем различие?](http://baguzin.ru/wp/?p=15634))

В приложенном Excel-файле сделано 10 000 выборок. Гистограммы для этих 10 000 наблюдений приведены на рис. 3–7, где хорошо видны контуры распределений. Поскольку формулы в файле содержат функцию СЛЧИС(), все значения пересчитываются при каждом изменении на листе. Поэтому ваши значения будут отличаться, от того, что представлено на рисунках. При построении графиков в Excel я воспользовался новой возможностью, доступной в Excel 2016 – частотной диаграммой; подробнее см. [Новые диаграммы в Excel 2016](http://baguzin.ru/wp/?p=16080).

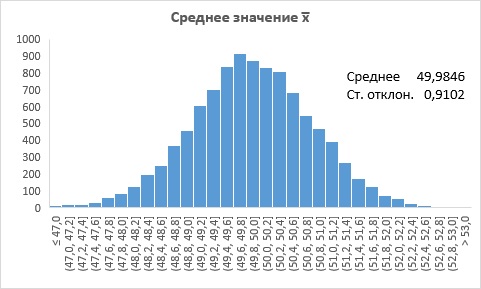


Рис. 3. Распределение среднего х̅

Видно, что:

* распределение х̅ симметрично относительно среднего и имеет нормальную форму;
* среднее значение х̅ = 49,9846 очень близко к генеральному среднему, равному 50,0;
* стандартное отклонение х̅, равное 0,9102, близко к генеральному стандартному отклонению, равному 2/0,8944.

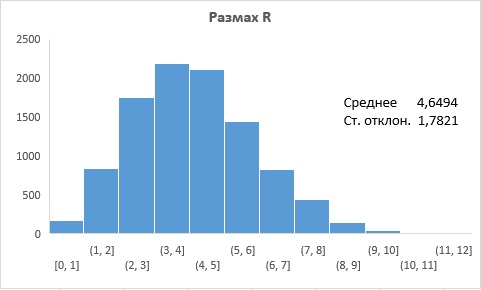


Рис. 4. Распределение размаха R

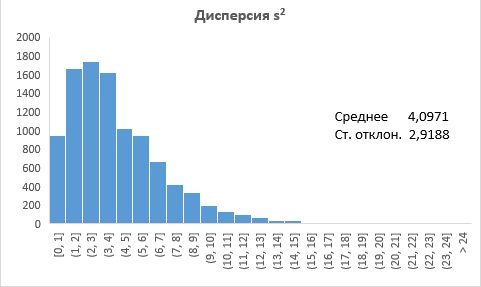


Рис. 5. Распределение дисперсии s2

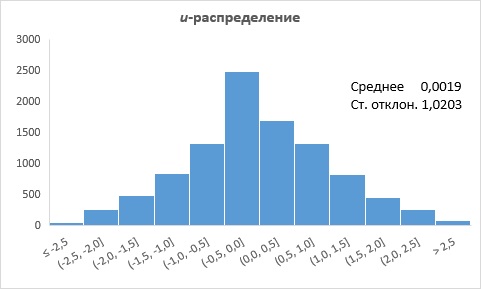


Рис. 6. *u*-распределение



Рис. 7. *t*-распределение

### Распределение х̅

В общем, справедливы следующие теоремы

**Теорема 1.** Пусть х1, х2, …, xn – это *n* наблюдений из генеральной совокупности со средним µ и дисперсией σ2, а х̅ – выборочное среднее. Тогда математическое ожидание, дисперсия и стандартное отклонение х̅ равны, соответственно:

**Теорема 2.** Пусть х1, х2, …, xn – выборка из распределения N(µ, σ2), а х̅ – выборочное среднее. Тогда х̅ распределено, как N(µ, σ2/n).

Важным представляется тот факт, что стандартное отклонение выборочного среднего в раз меньше, чем для совокупности (рис. 8). Точность выборочного среднего как оценки генерального среднего пропорциональна корню квадратному из объема выборки .

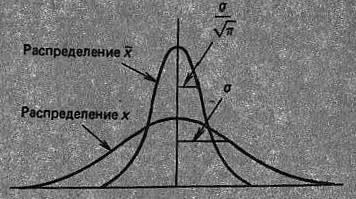


Рис. 8. Распределение случайной величины х и выборочного среднего х̅

**Теорема 3.** (Центральная Предельная Теорема) Распределение выборочного среднего из любой совокупности с конечной дисперсией стремится к нормальному распределению по мере роста объема выборки до бесконечности.

В соответствии с теоремой 2 выборочное среднее из совокупности с нормальным распределением имеет в точности нормальное распределение. А теорема 3 доказывает, что даже если распределение некоторой совокупности и ненормально, выборочное среднее все равно распределено приблизительно нормально. Приближение становится все лучше по мере роста числа *n*, но приемлемым оно становится уже при столь малом значении, как 5. На рис. 9, например, показаны распределения выборочного среднего из *n* наблюдений, взятых из равномерного распределения. Именно благодаря Центральной Предельной Теореме мы можем делать различные статистические выводы, используя выборочное среднее в предположении, что совокупность распределена нормально.

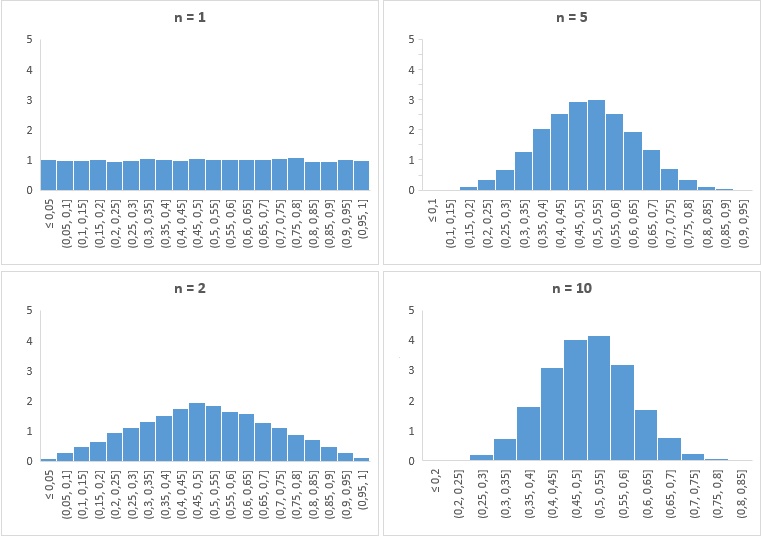


Рис. 9. Пример действия Центральной Предельной Теоремы

### Распределение R

На рис. 4 можно увидеть, что:

* распределение имеет положительную асимметрию;
* среднее, равное 4,65, более чем в 2 раза превосходит генеральное стандартное отклонение σ (напоминаю, что мы строили распределению N(µ, σ2), где µ = 50, σ = 2);
* стандартное отклонение, равное 1,78, составляет около 0,9 от генерального стандартного отклонения σ = 2.

**Теорема 4.** Пусть R – размах выборки (х1, х2, …, xn) из распределения N(µ, σ2). Тогда математическое ожидание и стандартное отклонение размаха R есть:

(6) *E(R) = d2σ*

(7) *D(R) = d3σ*

где d2 и d3 – константы, зависящие от *n* (рис.10)



Рис. 10. Константы

Для случая *n* = 5 имеем d2 = 2,326 и d3 = 0,864. По теореме 4 мы можем оценить σ из выражения

(8) *σ̂ = R/d2*

### Распределение V (s2)

Из рис. 5 видно, что:

* распределение положительно асимметрично;
* среднее, равное 4,09, почти равно генеральной дисперсии σ2 = 22 = 4;
* стандартное отклонение, равное 2,92, составляет около 0,7 от генеральной дисперсии.

**Теорема 5.** Пусть V – дисперсия выборки (х1, х2, …, xn) из распределения N(µ, σ2). Тогда математическое ожидание и стандартное отклонение V:

(9) *E(V) = σ2*

Статистика V используется для оценки генеральной дисперсии σ2. Из выражения (9) следует, что ожидание как раз и равно σ2. Формула для оценки, в которой ожидание оказывается равным оцениваемому параметру генеральной совокупности, называется несмещенным оценивателем. Значит, V – это несмещенный оцениватель для σ2. Конечно, х̅ тоже несмещенный оцениватель для µ (подробнее о смещении см. [СТАНДОТКЛОН.В и СТАНДОТКЛОН.Г: в чем различие?](http://baguzin.ru/wp/?p=15634)).

Между прочим, V получается в результате деления суммы квадратов отклонений *S* на (*n* – 1), что следует из определения дисперсии:

Может показаться странным, что *S* делится именно на (*n* – 1), а не на *n*, и одна из причин этого заключается в свойстве (9). Другая причина состоит в том, что число независимых переменных, используемых при вычислении *S*, равно (*n* – 1). Когда, например, *n* = 1, при любом значении х1 мы имеем *S* = 0. Значит, в выборке с *n* = 1 нет информации о дисперсии. Мы можем получить информацию о вариации только при *n* ≥ 2, а *n* наблюдений содержат информацию о (*n* – 1) переменной. Это число называется степенями свободы. Обозначив их через φ, в данном случае будем иметь:

(12) ф = *n* – 1

### u-распределение

**Теорема 6.** Пусть х̅ – среднее выборки (х1, х2, …, xn) из распределения N(µ, σ2). Тогда статистика

распределена как нормированное (или стандартное) нормальное распределение.

Обозначим двустороннюю α-процентную точку нормированного нормального распределения через u(α). А именно

(13) Pr{|u| ≥ u(а)} = α, где u распределено как N(0, 12).

Это распределение используется при проверках гипотез и оценивании генеральных средних, когда σ известна.

### t-pacnределение

Из рис. 6 и 7 мы видим, что распределение *t* похоже на распределение *u*, только с несколько большей вариацией.

**Теорема 7.** Подставив в выражение (1) вместо σ выборочное стандартное отклонение *s*, получим:

Величина *t* распределена как *t*-распределение со степенями свободы *ф = n – 1.* Поскольку вместо σ подставлена ее оценка, вполне естественно, что *t*-распределение имеет б***о***льшую вариацию, чем нормированное нормальное распределение. Когда степеней свободы, ф = *n* – 1, немного, у этого распределения получаются длинные «хвосты». Когда же *n* очень велико, *s* будет очень близко к σ. Таким образом, мы можем надеяться, что для больших *n* распределение *t* будет мало отличаться от нормированного нормального распределения. Действительно, ведь при степенях свободы, равных бесконечности, *t*-распределение совпадает с нормированным нормальным распределением.

Распределение *t* однозначно определяется степенями свободы φ (рис. 11). Обозначим двустороннюю α-процентную точку *t*-распределения со степенями свободы φ через t(φ, α), т.е.

(14) Pr{|t| ≥ t(φ, а)} = α

где *t* распределено как *t*-распределение с φ степенями свободы (рис. 12). Распределение *t* используется для проверки гипотез и оценивания генеральных средних, когда σ не известна, или для проверки и оценивания разности двух генеральных средних.

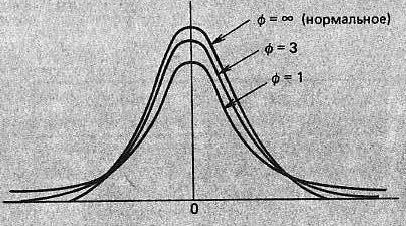


Рис. 11. *t*-распределение

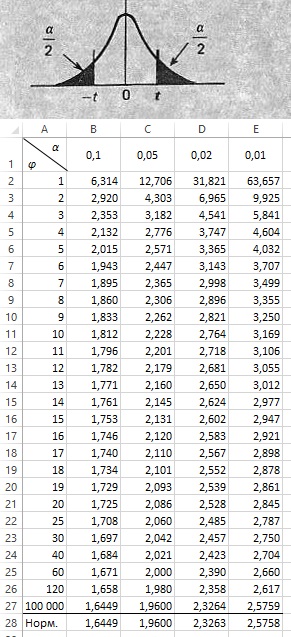


Рис. 12. Процентные точки для *t*-распределения

Эта таблица применяется для получения значений *t* при заданных числах степеней свободы (*φ*) и двусторонней вероятности (*α*). Когда, например, известно, что *φ* = 10, а *α* = 0,05, имеем *t* = 2,228. В принятых нами обозначениях табличное значение есть *t(α, φ)*. Значит, t(10, 0,05) = 2,228. Когда *φ* = ∞, табличные значения совпадают со значениями процентилей нормированного нормального распределения, т.е. t(∞, α) = u(α). Формула в Excel для расчета *t* в ячейке В2 =СТЬЮДЕНТ.ОБР.2Х(B$1;$A2). В строке 28 формула нормального распределения *u* =НОРМ.СТ.ОБР(1-B1/2).

### Проверка гипотез

**Пример 1.** Прочность на разрыв (предел прочности) нержавеющей стали, производимой на одном из заводов, сначала была стабильной со средним 72 кг/мм2 и стандартным отклонением 2,0 кг/мм2. Недавно разладился один агрегат. Он был снова налажен, и для определения эффекта наладки были взяты 10 образцов (рис. 13).



Рис. 13. Прочность на разрыв десяти образцов

Предположим, что стандартное отклонение осталось тем же, что и до наладки. Можем ли мы заключить, что наладка привела к изменению прочности на разрыв стали? Давайте сначала уясним следующие вопросы.

1. До наладки прочность на разрыв стали была распределена как N(µ0, σ02). Здесь µ0 и σ0 известны и равны 72,0 и 2,0 соответственно.
2. После наладки прочность осталась распределенной как N(µ, σ2), где σ предполагается такой же, что и σ0 = 2,0, а µ не известно.
3. Вопрос в том, равны ли µ и µ0 или нет? А именно проблема сводится к сравнению двух генеральных средних µ0 и µ.
4. Десять образцов отбирались случайно из N(µ, σ2). Они несут некоторую информацию о µ.

Возьмем среднее х̅ по 10 образцам:

Это значение отличается от того, что было до наладки, µ0 = 72,0 кг/мм2. Однако из этого мы еще не можем заключить, что после наладки действительно изменилась прочность стали, поскольку выборочное среднее х̅ имеет вариацию и вовсе не всегда равно генеральному среднему. Давайте рассмотрим гипотезу о том, что после наладки прочность стали не изменилась. Если бы эта гипотеза была верна, то х̅ имело бы нормальное распределение со средним µ0 = 72,0 и стандартным отклонением

Следовательно, нормализация случайной величины u0 получается из

и эта величина распределена как N(0, 12). Поскольку из (15) мы имеем х̅ = 75,0, то значение u0:

Вероятность, что |u0| будет столь велико, как 4,74, крайне мала (около 0,000002). Это означает либо что произошло совершенно необычное событие, либо что наша гипотеза не верна. Таким образом, мы должны подозревать, что гипотеза µ = µ0 = 72,0 не верна. Решения на основе выборочных наблюдений о том, таков ли параметр генеральной совокупности или нет (например, решение о равенстве µ = µ0 = 72,0) называются *проверками гипотез.*

Предположив, что гипотеза не верна, мы говорим, что мы ее отбросили. В нашем примере, когда нулевая гипотеза Н0: µ = 72,0 будет верна, значение u0, вероятно, окажется близким к нулю. Значит, когда |u0| превышает некоторый предел, мы отбрасываем Н0. Можно, например, следовать такому предписанию: «Когда |u0| ≥ 1,96, отбросить Н0, а когда |u0| < 1,96, принять Н0». Придерживаясь этого, получим, что вероятность отбросить Н0, когда она на самом деле верна, равна:

(19) Pr(|u0| ≥ 1,96) = 0,05

Такая вероятность называется *уровнем значимости* и ее обычно обозначают α. Этот тип ошибки, при которой гипотеза неверно отбрасывается, называется *ошибкой первого рода*. Величина α обычно выбирается равной 0,05 (5%) или 0,01 (1%). Область значений u0, в которой гипотеза Н0 отбрасывается, называется *областью отбрасывания*, а область, где Н0 принимается, называется *областью принятия* (рис. 14).

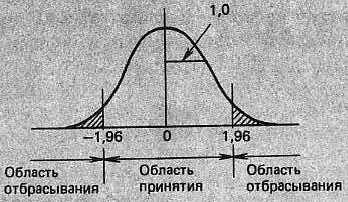


Рис. 14. Двусторонний критерий

Когда гипотеза отбрасывается, можно твердо настаивать, что «то значение параметра, которое фигурирует в гипотезе, неверно». С другой стороны, даже если мы примем гипотезу, все-таки нельзя безоговорочно утверждать, что эта гипотеза действительно верна, поскольку можно было бы принять многие другие гипотезы на основе имеющейся выборки данных, но лишь одна из них окажется действительно верной. Принимая гипотезу, которая на самом деле не верна, мы совершаем ошибку, называемую ошибкой *второго рода;* ее вероятность обозначается β. Для µ ≠ µ0 имеем:

и вероятность ошибки второго рода иллюстрируется на рис. 15. Когда гипотеза не верна, мы хотим отбросить ее с высокой вероятностью. Вероятность отбрасывания неверной гипотезы называется мощностью критерия и обозначается *Р*, где *Р = 1 – β*. Значение *Р*, как видно из рис. 16, меняется в зависимости от разности между µ и µ0.

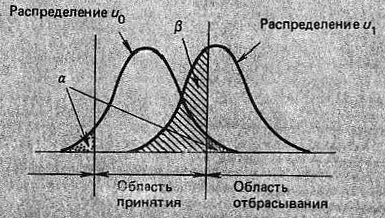


Рис. 15. Ошибка второго рода

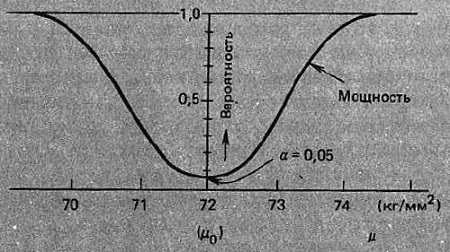


Рис. 16. Кривая мощности для примера 1

Метод проверки гипотезы показан в таблице (рис. 17).

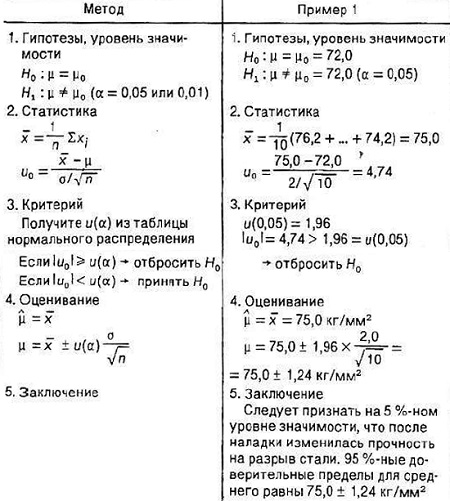


Рис. 17. Проверка гипотезы и оценивание генерального среднего, когда σ известна

**Пример 2.** Можно ли в примере 1 считать, что после наладки оборудования увеличилась прочность стали? Допустим, что все остальные условия сохранились прежними. В примере 1 нас интересовало, привела ли настройка оборудования к изменению прочности или нет. Когда гипотеза Н0: μ = 72,0 была отброшена, мы пришли к решению, что μ ≠ 72,0. В этом примере нам необходимо выяснить, увеличилась ли после настройки прочность или нет. Если нулевая гипотеза будет отброшена, то мы заключим, что μ > 72,0.

Когда гипотеза Н0 отбрасывается, та гипотеза, которую мы принимаем, называется *альтернативной гипотезой* (или альтернативой) и обозначается H1. В примере 1 суть Н1 была следующей:

*(21) Н1: μ ≠ μ0 = 72,0,*

тогда как в примере 2 нулевая гипотеза:

*(22) Н1: μ > μ0 = 72,0,*

При проверке альтернативной гипотезы (21) область отбрасывания приходилась на оба хвоста распределения, как показано на рис. 14. Для (22), напротив, область отбрасывания приходится только на правый хвост, как показано на рис. 18. Первый из этих критериев называется двусторонним критерием, а второй – односторонним критерием.



Рис. 18. Односторонний критерий

Метод проверки гипотезы для примера 2 приведен в таблице (рис. 19).

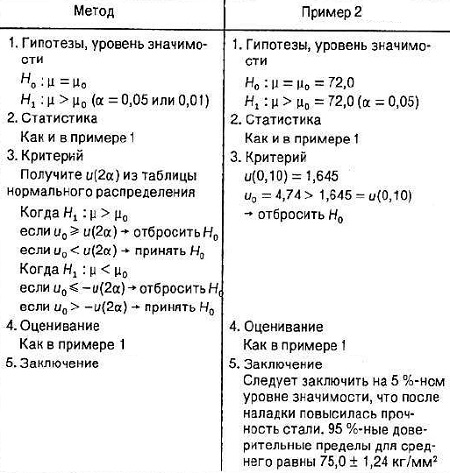


Рис. 19. Проверка гипотезы и оценивание генерального среднего (односторонний критерий)

### Оценивание параметров

**Пример 3.** В примере 1 мы заключили, что средняя прочность изменилась. Пусть так, но каково же теперь новое генеральное среднее? Оценивание – это процесс анализа выборки, направленный на предсказание соответствующего значения параметра генеральной совокупности. *Точечная оценка* – это оценка параметра совокупности, представленная одним числом. Точечная оценка параметра θ обозначается θ̂. Например, генеральное среднее μ в примере 1 оценивается с помощью х̅:

*(23) μ̂ = х̅ = 75,0 кг/мм2*

*Интервальная оценка* – это оценка параметра совокупности, даваемая двумя числами, между которыми, как считается, лежит значение параметра.

Положим, что х1, х2, …, xn извлечены из N(µ, σ2), тогда выборочное среднее х̅ распределено как N(µ, σ2/n). Запишем

тогда *u* будет распределено как N(0, 12). Следовательно, вероятность, что значение *u* окажется между ± u(α) равна (1 – α), где u(α) - двусторонняя α-процентная точка нормированного нормального распределения. Или

Преобразуя выражение 23, получим:

Этот интервал называется 100(1 – α)%-ным доверительным интервалом для μ, а верхняя и нижняя границы этого интервала называются доверительными пределами. Согласно выражению (25) вероятность того, что интервал накроет μ равна (1 – α). Эта вероятность, например (1 – α), называется *доверительным уровнем*. Обычно доверительный уровень выбирается равным 0,95 или 0,99. Выбрав 0,95 из (25) получим 95%-ные доверительные пределы:

Ответ на вопрос примера 3 приведен в этапе 4 таблицы на рис. 17. Итак, мы можем оценить, что новое среднее с вероятностью 95% находится в диапазоне μ = 73,76 … 76,24.

### Проверка гипотез и оценивание генеральных средних, когда σ неизвестна

В разных задачах приходится рассматривать те или иные виды критериев и методов оценивания. Основные принципы и подходы остаются при этом точно такими же, что описаны выше, меняются лишь используемые статистики. Некоторые из наиболее распространенных критериев и оценок будут рассмотрены далее. Для начала мы обсудим критерий и оценку генерального среднего, когда σ неизвестна.

Пусть из совокупности со средним μ и стандартным отклонением σ извлечены *n* образцов. Тогда выборочное распределение

можно представить, как нормированное нормальное распределение. Мы пользовались этим для проверки гипотезы и оценивания генерального среднего, когда σ была известна. Однако иногда σ неизвестна. В таком случае кажется естественным заменить σ на s и применить выражение

вместо выражения для *u*, где *s* – выборочное стандартное отклонение:

Чтобы воспользоваться этой статистикой, нам надо знать выборочное распределение *t*. Оно известно как *t*-распределение с (n – 1) степенями свободы.

Для проверки гипотез

*(28) Н0: μ = μ0 и H1: μ ≠ μ0*

вычислим t0 из

и сравним полученный результат с *t(n – 1, α)*, где *t(n – 1, α) –* двусторонняя α-процентная точка *t*-распределения с (n – 1) степенями свободы. Если |t0| ≥ t(n – 1 , α), гипотеза Н0 отбрасывается, а если |t0| < t(n – 1 , α), Н0 принимается. В случае одностороннего критерия, т.е. когда H1: μ > μ0 или H1: : μ < μ0, значение t0 сравнивается с *t(n – 1, 2α)* или –*t(n – 1, 2α)* (минус *t*) соответственно.

Поскольку значение *t*, определяемое из (2), распределено как *t*-распределение с *φ = n – 1*, получим:

Подставив вместо t в (30) выражение  и выполнив преобразования, находим:

В результате получатся следующие 95%-ные доверительные пределы для μ:

Пример и метод анализа показаны в таблице (рис. 20).

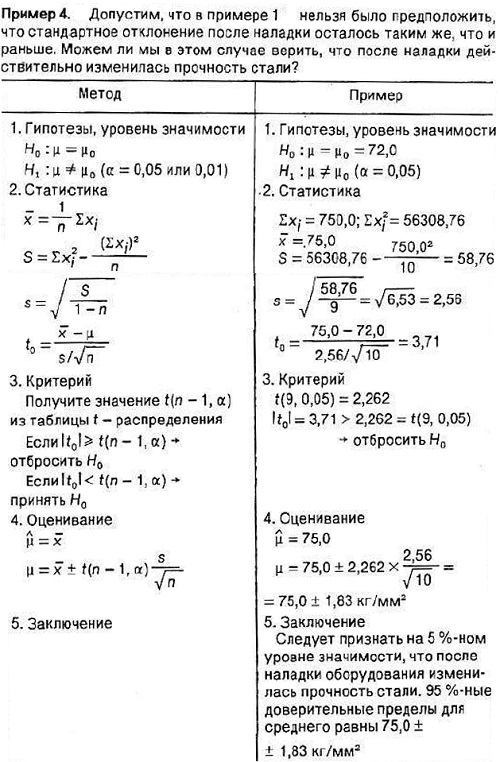


Рис. 20. Проверка и оценивание генерального среднего, когда σ неизвестна

### Проверка гипотез и оценивание различий между двумя генеральными средними

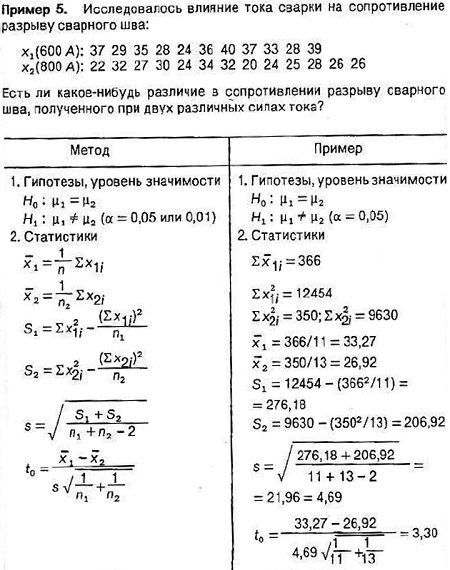
Пусть у нас есть два набора образцов изделий, и мы хотим знать, равны средние двух популяций или нет. Предположим, что х11, х12, …, x1n1 – это n1 образцов из первой совокупности, a х21, х22, …, x2n2 – это n2 образцов из второй совокупности, причем среднее и дисперсия первой совокупности равны μ1 и σ2, а для второй – μ2 и σ2. Для проверки гипотезы о разности между двумя генеральными средними мы испытаем разность (х̅1 – х̅2). Поскольку распределения выборочных средних х̅1 и х̅2 можно рассматривать как N(μ1, σ2/n1) и N(μ2, σ2/n2) соответственно, то и разность (х̅1 – х̅2) тоже будет иметь нормальное распределение N{μ1 – μ2, σ2(1/n1– 1/n2)}. Следовательно, после нормирования

получится распределение N(0, 12). Давайте теперь предположим, что σ неизвестна. Подставив s вместо σ, получим:

где

a *S1* и *S2* – суммы квадратов для каждой выборки. Величина *t* в (34) имеет *t*-распределение с *φ = n1 + n2 – 2*. Если между этими двумя средними нет разницы, то, положив μ1 = μ2 в (34), мы получим:

Пример и метод анализа показаны в таблице (рис. 21).



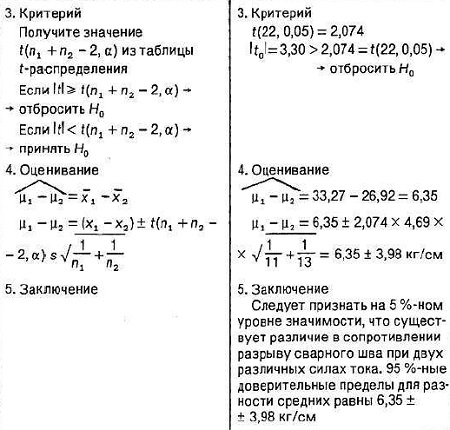


Рис. 21. Проверка гипотезы и оценивание разности двух генеральных средних

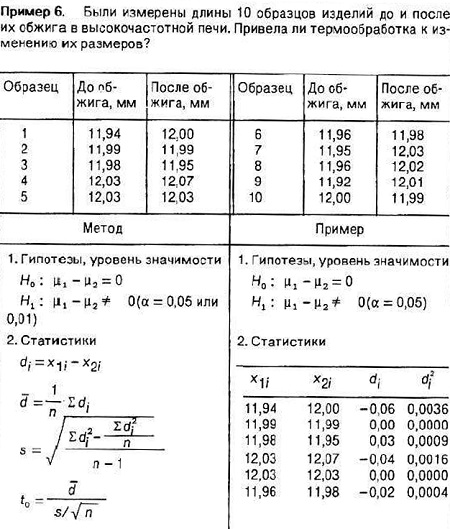
### Проверки гипотез и оценивание для парных наблюдений

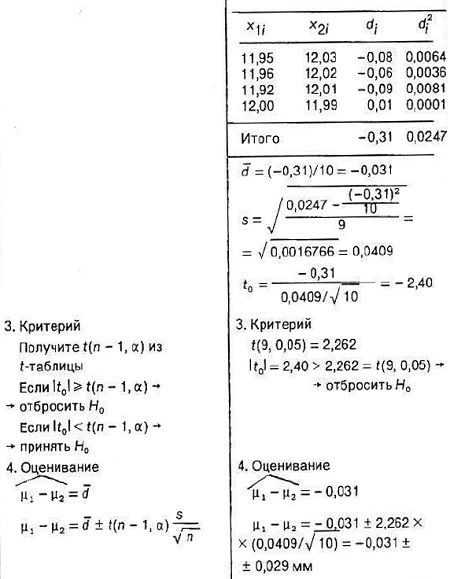
При отборе из двух совокупностей иногда значимое различие в средних оказывается обусловленным дополнительными факторами, тогда как на самом деле различия отсутствуют, а мы пытаемся измерить этот эффект. Например, в эксперименте по проверке, какой из двух типов удобрений (А или В) лучше, на каждой из 10 сортоиспытательных станций выделено по две делянки с пшеницей. На одной из делянок вносится удобрение А, а на другой – В. Если сравнить среднее из 10 наблюдений за делянками с удобрением типа А со средним из 10 наблюдений за делянками с удобрением типа В, то наблюдаемое различие (если оно обнаружится) может быть обусловлено разными типами почв или неодинаковыми метеоусловиями, а вовсе не какими-то различиями между самими удобрениями. Может быть и так, что различия между удобрениями в действительности есть, но они затушевываются другими непонятными факторами. Для преодоления этих трудностей прибегают к такому типу эксперимента, в котором проводятся парные наблюдения. Мы должны гарантировать, что два члена любой пары будут идентичны во всех отношениях, кроме одного, которое мы и пытаемся измерить. Значит, каждая пара делянок должна закладываться на практически одинаковой почве, в одних и тех же метеоусловиях и т.п.

Пусть x1i, будет первым членом *i*-й пары, а x2i – вторым членом. У нас есть, скажем, *n* пар наблюдений

(x11, x21), (x12, x22), …, (x1i, x2i), … , (x1n, x2n)

Если взять разности di = x1i – x2i, то получатся *n* наблюдений *di*. Нас интересует гипотеза о том, что μ1 = μ2. Эта гипотеза утверждает, что между способами обработки делянок нет различия, т.е. нет разницы в парах. Здесь возможны различия между парами, но они могут быть элиминированы при переходе к разностям *di*. Если эта гипотеза верна, то значение *di*, должно следовать распределению со средним, равным нулю. Проверка осуществляется точно так же, как и в примере 4 (*t*-критерий с *φ = n – 1*). Пример и метод анализа показаны в таблице (рис. 22).





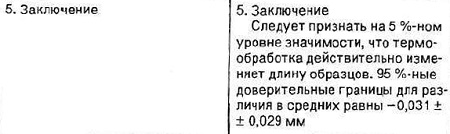


Рис. 22. Проверка гипотезы и оценивание для парных наблюдений

### Критерии значимости для коэффициента корреляции

Если мы хотим узнать силу связи между двумя переменными *х* и *у*, надо вычислить выборочный коэффициент корреляции *r*. Предположим, что *х* и *у* имеют двумерное нормальное распределение с генеральным коэффициентом корреляции ρ. Даже если ρ = 0, выборочный коэффициент корреляции r не всегда будет равен нулю. Проверить, равен ли нулю генеральный коэффициент корреляции ρ или нет, можно с помощью следующей теоремы.

Теорема 8. Пусть (хi, уi), i = 1, 2, …, n, будут n образцами из двумерного нормального распределения с коэффициентом корреляции, равным нулю. Обозначив выборочный коэффициент корреляции через r, получим, что

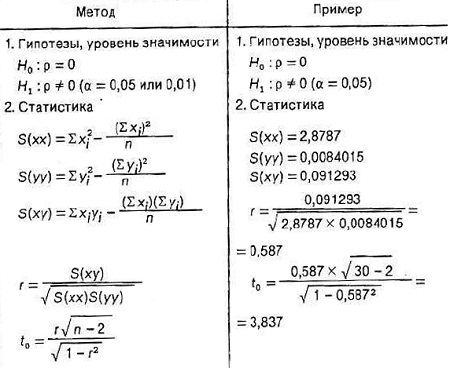
имеет *t*-распределение с *(n – 2)* степенями свободы.

В таблице (рис. 23) приведены 30 пар данных о давлении воздуха и проценте дефектов в пластиковых емкостях. Можем ли мы утверждать, что между этими двумя характеристиками есть корреляция?



Рис. 23. Данные о давлении воздуха и проценте дефектов в пластиковых емкостях

Метод проверки гипотезы, о том, что генеральный коэффициент корреляции равен нулю, приведены в таблице (рис. 24).



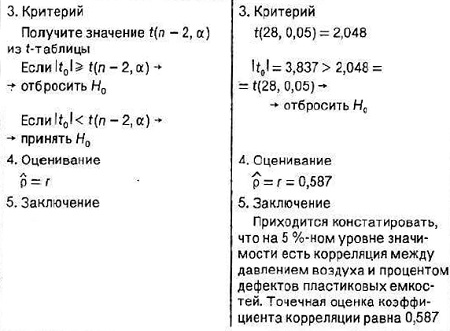


Рис. 24. Критерий для выборочного коэффициента корреляции

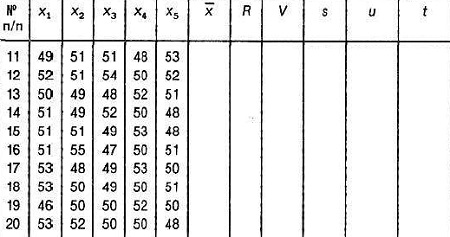
Итак, вы познакомились с основами и методами статистического вывода. Кроме этих методов, есть еще много более сложных. Например, 1) критерии и оценки для генеральной дисперсии; 2) дисперсионный анализ (см. [Однофакторный дисперсионный анализ](http://baguzin.ru/wp/?p=5884)); 3) планирование и анализ экспериментов (см. [Фишер. Статистический вывод](http://baguzin.ru/wp/?p=15657)); 4) многофакторный регрессионный анализ (см. [Двухфакторный дисперсионный анализ](http://baguzin.ru/wp/?p=5919)) и т.д.

### Упражнения

Упражнение 1. Объясните разницу в содержании следующих пар терминов:

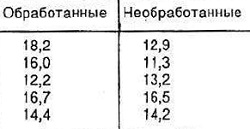
1. генеральная совокупность и выборка;
2. параметр совокупности и статистика;
3. ошибки первого рода и второго рода;
4. уровень значимости и доверительный уровень.

Упражнение 2. Вычислите статистики для 10 выборок:



Подсказка. Если затрудняетесь, вернитесь к рис. 2 в начале заметки.

Упражнение 3. В одном исследовании требовалось установить, влияет или нет определенная обработка поверхности некоторого материала на его сопротивление трению. Были изготовлены опытные образцы материала, затем 5 из них были обработаны, а остальные 5 остались необработанными. Было измерено сопротивление трению 10 образцов. Результаты приведены ниже:



Ответьте на следующие вопросы:

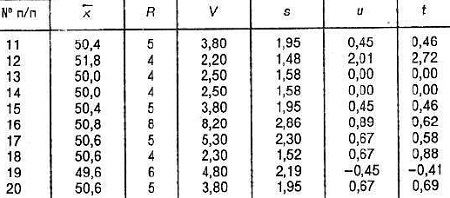
1. Можно ли утверждать, что обработка повышает сопротивление трению? Проверьте вашу гипотезу.
2. Найдите 95 %-ные доверительные пределы для разности средних сопротивлений трению между обработанными и необработанными образцами материала.
3. Улучшите план приведенного выше эксперимента, если это возможно.

### Ответы

Упражнение 1.

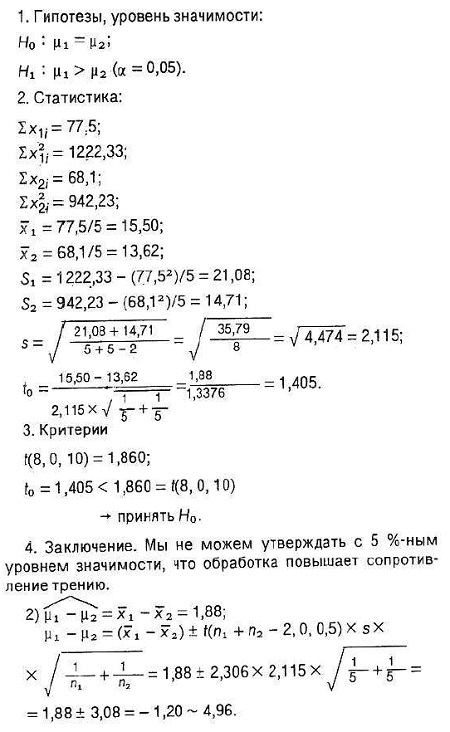
1. Генеральная совокупность – это что-то, над чем мы совершаем действия; выборка составляет от нее малую часть. Опираясь на эту выборку, мы можем оценить характеристики исходной совокупности.
2. Параметром совокупности называется константа, характеризующая данную совокупность, а статистикой – функция от выборочных наблюдений, полученных из совокупности. Опираясь на значение статистики, мы можем делать выводы относительно соответствующего параметра совокупности.
3. При проверке гипотез можно сделать два вида ошибок. Ошибка первого рода – это ошибочное отбрасывание верной гипотезы, а ошибка второго рода – ошибочное принятие ложной гипотезы.
4. Уровень значимости есть вероятность некорректного отбрасывания гипотезы, а доверительный уровень – это вероятность того, что доверительные пределы накроют истинное значение оцениваемого параметра.

Упражнение 2.



Упражнение 3.

1. Проверяем гипотезу, что обработка повышает сопротивление трению.



3) Сравнение надо делать между максимально похожими друг на друга образцами. Для достижения этого было бы лучше всего разрезать каждый из специальных образцов пополам, затем одну половину обработать, а другую нет. Тогда в ходе анализа можно было бы определить, существуют ли какие-нибудь различил между парными образцами.