**Проверка гипотез: двухвыборочные критерии**

Проверка гипотез основана на подтверждающем подходе к анализу данных. В [предыдущей заметке](http://baguzin.ru/wp/?p=5790) рассмотрены широко распространенные процедуры проверки гипотез на основе одной выборки, извлеченной из одной генеральной совокупности. В этой заметке описываются процедуры проверки гипотез на основе двух числовых выборок, извлеченных из двух генеральных совокупностей. Например, равны ли средние недельные объемы продаж BLK-колы, размещенной на специализированных стеллажах и на обычных полках?[[1]](#footnote-1)

Применение статистики в этой заметке будет показано на сквозном примере **«Зависит ли объем продаж от вида полок в магазине?»** Представьте себе, что вы — региональный менеджер по продажам компании BLK Foods и хотите сравнить объемы продаж BLK-колы, выставленной на обычных полках и на специализированных стеллажах. Для этого вы создаете выборку, состоящую из 20 магазинов компании BLK Foods, в которых объявлена полная распродажа товаров. Затем вы случайным образом делите эту выборку пополам: 10 магазинов относите к первой группе, а остальные 10 — ко второй. Менеджеры магазинов из первой группы размещают бутылки с BLK-колой на обычных полках среди других прохладительных напитков. В то же время менеджеры магазинов из второй группы должны расположить бутылки с BLK-колой на специализированных стеллажах и разместить на них рекламу. Как определить, одинаковы ли объемы продаж BLK-колы в магазинах из этих двух групп? Совпадает ли изменчивость объемов продаж в этих магазинах? Как использовать ответы на эти вопросы, чтобы повысить объемы продаж BLK-колы?

**Использование Z-критерия для оценки разности между двумя математическими ожиданиями**

Предположим, что из первой генеральной совокупности извлекается случайная выборка, имеющая объем *n1* а из второй — случайная выборка, объем которой равен *n2*. Требуется проанализировать данные, принадлежащие каждой выборке. Обозначим математическое ожидание первой генеральной совокупности через *μ1*, а стандартное отклонение — через *σ1*. Аналогично математическое ожидание второй генеральной совокупности обозначим символом *μ1*, а стандартное отклонение — σ*2*. Статистика, положенная в основу критерия для проверки равенства математических ожиданий двух генеральных совокупностей, основана на разности между выборочными средними *Х̅1* – *Х̅2*. По центральной предельной теореме, сформулированной [ранее](http://baguzin.ru/wp/?p=5663), при достаточно больших объемах выборок эта статистика имеет стандартизованное нормальное распределение. Следовательно, для оценки разности между двумя математическими ожиданиями можно сформулировать следующий *Z*-критерий:

*(1) Z =*

где *Х̅1* — среднее значение выборки из первой генеральной совокупности, *μ1* — математическое ожидание первой генеральной совокупности, — дисперсия первой генеральной совокупности, *n1* — объем выборки, извлеченной из первой генеральной совокупности, *Х̅2* — среднее значение выборки из второй генеральной совокупности, *μ2* — математическое ожидание второй генеральной совокупности, — дисперсия второй генеральной совокупности, *n2* — объем выборки, извлеченной из второй генеральной совокупности. Статистика Z имеет стандартизованное нормальное распределение.

**Применение *t*-критерия для оценки разности между математическими ожиданиями с помощью суммарной дисперсии**

В большинстве ситуаций дисперсии и стандартные отклонения двух генеральных совокупностей неизвестны. Единственная информация, доступная исследователю, — выборочные средние, выборочные дисперсии и выборочные стандартные отклонения. Если выборки являются случайными, независимыми и извлечены из нормально распределенных генеральных совокупностей, имеющих одинаковую дисперсию (т.е. = ), для проверки гипотезы о значимом различии между математическими ожиданиями двух генеральных совокупностей можно применять *t-критерий, использующий суммарную дисперсию*. Нулевая гипотеза состоит в том, что математические ожидания двух независимых генеральных совокупностей не отличаются друг от друга:

H0: μ1 = μ2 или μ1 – μ2 = 0

Альтернативная гипотеза заключается в том, что математические ожидания не совпадают:

H1: μ1 ≠ μ2 или μ1 – μ2 ≠ 0

t-критерий для оценки разности между двумя математическими ожиданиями с помощью суммарной дисперсии

*(2) t =*

где = суммарная дисперсия, *Х̅1* — среднее значение выборки из первой генеральной совокупности, *μ1* — математическое ожидание первой генеральной совокупности, — дисперсия выборки из первой генеральной совокупности, *n1* — объем выборки, извлеченной из первой генеральной совокупности, *Х̅2* — среднее значение выборки из второй генеральной совокупности, *μ2* — математическое ожидание второй генеральной совокупности, — дисперсия выборки из второй генеральной совокупности, *n2* — объем выборки, извлеченной из второй генеральной совокупности.

Статистика t имеет *t*-распределение Стьюдента с *n1 + n2 – 2* степенями свободы.

При заданном уровне значимости α двусторонний критерий отклоняет нулевую гипотезу, если *t*-статистика больше верхнего критического значения или меньше нижнего критического значения (рис. 1). Ограниченный сверху критерий отклоняет нулевую гипотезу, если *t*-статистика больше верхнего критического значения, а ограниченный снизу критерий — если она меньше нижнего критического значения.

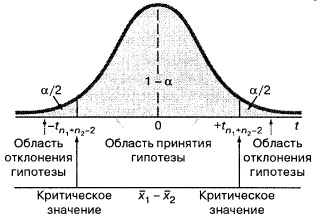


Рис. 1. Области принятия и отклонения гипотез при использовании двустороннего *t*-критерия для оценки разности между двумя математическими ожиданиями

Продемонстрируем применение *t*-критерия, использующего суммарную дисперсию, на примере сценария, описанного в начале заметки. Совпадают ли средние объемы продаж BLK-колы, размещенной на обычных полках и специализированных стеллажах. В этой задаче рассматриваются две генеральные совокупности. Первая генеральная совокупность состоит из всевозможных еженедельных объемов продаж BLK-колы, если все супермаркеты компании BLK используют обычные стеллажи. Во вторую генеральную совокупность входят всевозможные еженедельные объемы продаж BLK-колы, если все супермаркеты компании BLK используют специализированные стеллажи (рис. 2).



Рис. 2. Сравнение еженедельных продаж BLK-колы, размещенной на разных стеллажах (количество покупок)

Нулевая и альтернативная гипотезы формулируются следующим образом: H0: μ1 = μ2 или μ1 – μ2 = 0, H1: μ1 ≠ μ2 или μ1 – μ2 ≠ 0. Предполагая, что выборки извлечены из нормально распределенных генеральных совокупностей, имеющих одинаковую дисперсию (т.е. = ), применим *t*-критерий, использующий суммарную дисперсию. Эта статистика имеет *t*-распределение, имеющее 10 + 10 – 2 = 18 степеней свободы. Если уровень значимости двустороннего критерия α равен 0,05, критическая область разбивается на две части, каждая из которых соответствует вероятности, равной 0,025. Критические значения *t*-статистики: нижняя =СТЬЮДЕНТ.ОБР(0,025;18) = –2,1009, верхняя =СТЬЮДЕНТ.ОБР(0,975;18) = +2,1009 (рис. 3). Решающее правило имеет следующий вид: если t > +2,1009 или t < –2,1009, нулевая гипотеза *H0*отклоняется, в противном случае она не отклоняется

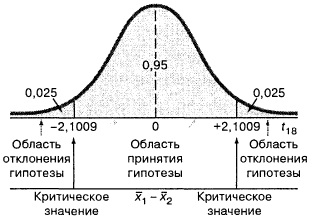


Рис. 3. Области принятия и отклонения гипотез при использовании *двустороннего* *t*-критерия для оценки разности между двумя математическими ожиданиями с уровнем значимости, равным 0,05, при 18 степенях свободы

Используя данные, содержащиеся на рис. 1 и [Пакет анализа](http://baguzin.ru/wp/?p=5316) Excel рассчитаем описательные статистики двух выборок и двухвыборочный t-тест для случая с одинаковыми дисперсиями (рис. 4).

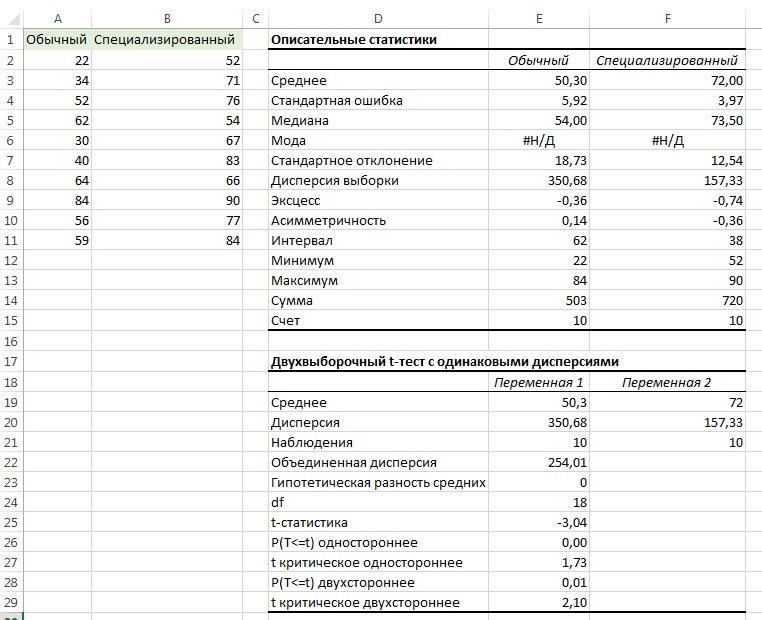


Рис. 4. Описательные статистики (панель А) и результаты применения *t*-критерия (панель Б) для двух разновидностей стеллажей

Поскольку уровень значимости равен 0,05, нулевая гипотеза отклоняется, так как *t* = ­–3,04 < *t18* < –2,10 (см. строку *t-статистика* таблицы «Двухвыборочный t-тест» нижней части рис. 4). Наблюдаемый уровень значимости (*р*-значение), вычисленный с помощью Excel, равен 0,01 (см. строку *P(T<=t) двухстороннее* таблицы «Двухвыборочный t-тест» нижней части рис. 4). Иначе говоря, вероятность того, что t > 3,04 или t < –3,04, равна 0,01. Значит, если математические ожидания обеих генеральных совокупностей на самом деле равны, вероятность обнаружить статистически значимую разность между ними равна 0,01. Поскольку *р*-значение меньше 0,05, у нас есть основания отклонить нулевую гипотезу. Таким образом, можно утверждать, что объем продаж BLK-колы, размещенной на обычных полках, значительно меньше объема продаж BLK-колы, расположенной на специализированных стеллажах.

Чтобы провести в Excel двухвыборочный t-тест пройдите по меню *Данные* → *Анализ данных*; в открывшемся окне *Анализ данных* выберите строку *Двухвыборочный t-тест с одинаковыми дисперсиями*; откроется окно *Двухвыборочный t-тест с одинаковыми дисперсиями* (рис. 5). Заполните его, как указано на рисунке.

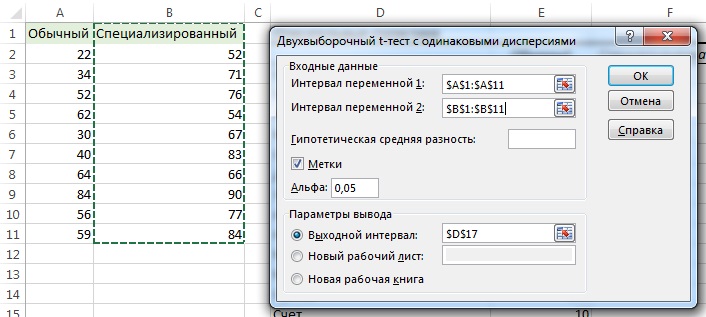


Рис. 5. Процедура Excel: проверка гипотезы о разности математических ожиданий двух генеральных совокупностей на основе выборок с помощью t-критерия, использующего суммарную дисперсию

При проверке гипотезы о разности математических ожиданий двух генеральных совокупностей с помощью *t*-критерия предполагается, что обе генеральные совокупности распределены нормально и имеют одинаковую дисперсию. Если объемы выборок достаточно велики, *t*-критерий, использующий суммарную дисперсию, является устойчивым и мало чувствительным к отклонению от предположения о нормальности генеральных совокупностей. В этих ситуациях *t*-критерий можно использовать без существенной потери мощности. С другой стороны, если предположение о нормальном распределении генеральных совокупностей не выполняется, существуют две возможности:

* можно использовать непараметрическую процедуру, например, ранговый критерий Уилкоксона (будет описан позднее), который не зависит от предположения о нормальности распределения генеральной совокупности.
* к каждой выборке можно применить нормирующее, а затем — *t*-критерий, использующий суммарную дисперсию.

Для проверки предположения о нормальном распределении каждой генеральной совокупности можно применить блочную диаграмму (рис. 6). Видно, что предположение о нормальном распределении генеральных совокупностей нарушается незначительно, следовательно, применение *t*-критерия не приведет к серьезным ошибкам.

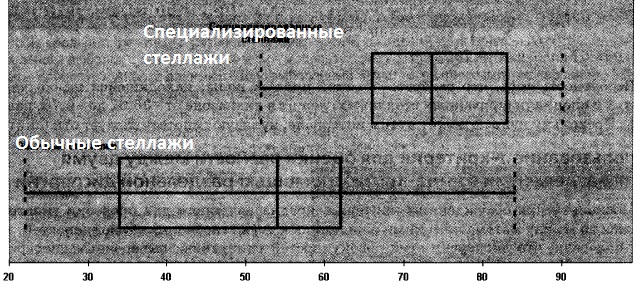


Рис. 6. Блочные диаграммы для двух разновидностей стеллажей

**Доверительный интервал для разности между математическими ожиданиями**

Вместо проверки гипотезы о разности между математическими ожиданиями двух генеральных совокупностей (или в дополнение к ней) можно построить доверительный интервал, содержащий среднюю разность:

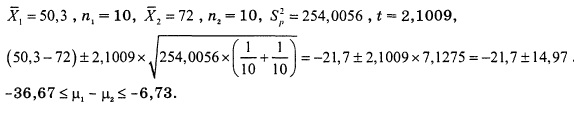
*(3) (Х̅1 – Х̅2) ±*

или



где — критическое значение *t*-распределения с n1 + n2 – 2 степенями свободы для области, ограниченной верхним хвостом распределения, площадь которой равна α/2.

Используя формулу (3) и данные, показанные на рис. 4, получаем следующий 95%-ный доверительный интервал:



Вероятность того, что разность объемов продаж колы, находящейся на обычных полках и специализированных стеллажах, лежит в диапазоне от -36,67 до -6,73, равна 95%. Поскольку ноль не лежит в этом интервале, нулевую гипотезу следует отклонить.

**Использование *t*-критерия для оценки разности между двумя математическими ожиданиями с помощью раздельной дисперсии**

Поскольку при обсуждении *t*-критерия, предназначенного для проверки гипотезы о разности между математическими ожиданиями двух генеральных совокупностей, мы предполагали, что дисперсии этих совокупностей одинаковы, выборочные дисперсии были объединены в одну величину — суммарную дисперсию . Однако, если это предположение ошибочно, суммарная дисперсия становится неприемлемой. Для решения этой проблемы Саттерсвейт предложил *t*-критерий, использующий раздельную дисперсию. В процедуре Саттерсвейта для аппроксимации *t*-статистики используются две выборочные дисперсии. Эта процедура довольна сложна, но в Пакете анализа для ее выполнения предусмотрена отдельная строка (рис. 7). Видно, что дисперсия продаж колы, расположенной на специализированных стеллажах, вдвое превосходит дисперсию продаж колы, выставленной на обычных полках. Таким образом, для проверки гипотезы о равенстве средних объемов продаж следует применить *t*-критерий, использующий раздельную дисперсию.



Рис. 7. Результат применения *t*-критерия, использующего раздельную дисперсию, полученный в Excel с помощью Пакета анализа

В частности, из рис. 7 следует, что *t*-статистика равна *t* = -3,04, причем двустороннее *р*-значение равно 0,0077 < 0,05. Следовательно, результаты применения *t*-критерия, использующего раздельную дисперсию, практически не отличаются от результатов, полученных с помощью *t*-критерия, использующего суммарную дисперсию.

Обратите внимание на то, что два разных *t*-критерия привели к одинаковым результатам. Предположение о равенстве дисперсий в этой задаче практически не влияет на результат. Однако в других ситуациях эти критерии могут привести к противоположным выводам. Именно поэтому следует уделять много внимания проверке предположения о равенстве дисперсий и лишь затем выбирать критерий. Эта проблема является весьма важной частью анализа данных. Для ее решения можно применять *F*-критерий, описанный ниже. Это позволит правильно выбрать *t*-критерий (использующий либо суммарную, либо раздельную дисперсию).

**Сравнение двух зависимых выборок: критерии для оценки разности между двумя математическими ожиданиями**

До сих пор мы рассматривали процедуры проверки гипотез о двух независимых генеральных совокупностях на основе извлеченных из них выборок. Далее описывается критерий, позволяющий оценить разность между математическими ожиданиями двух генеральных совокупностей, связанных между собой. Иначе говоря, показатели первой группы зависят от показателей второй. Эта зависимость возникает, поскольку элементы выборок являются парными результатами повторных измерений, выполненных в одном и том же множестве элементов. В этой ситуации интерес представляет разность между величинами, а не сами величины как таковые.

Первый подход к решению задачи о зависимых выборках основывается на попарном сравнении элементов, имеющих определенные свойства. Например, при сравнении результатов двух рекламных кампаний используется объем генеральной совокупности и/или другие экономические и демографические переменные. Исследуя эти переменные, можно измерить эффект двух разных рекламных стратегий.

Второй подход к анализу зависимых выборок использует повторные измерения одних и тех же элементов. Если предположить, что одни и те же элементы при разных воздействиях ведут себя по-разному, следует выявить любые отличия между двумя измерениями одних и тех же элементов. Например, при оценке вкуса некоего продукта каждый элемент выборки подвергается повторным испытаниям одним и тем же дегустатором.

Независимо от подхода к решению задачи, цель исследования двух зависимых выборок — выявить различия между результатами двух измерений, уменьшив влияние изменчивости, присущей элементам выборки.

Для того чтобы определить, существует ли разница между двумя группами, сначала вычисляют разности между отдельными элементами каждой группы (рис. 8).



Рис. 8. Вычисление разностей между элементами двух зависимых групп

Для оценки средней разности между средними значениями двух зависимых выборок величины *Di* рассматриваются как наблюдения, принадлежащие одной и той же выборке. Если стандартное отклонение разностей известно, применяется *Z*-статистика, вычисляемая по формуле:[[2]](#footnote-2)

*(4) Z =*

где *D̅ =* , *μD* – гипотетическое математическое ожидание, *σD* — стандартное отклонение генеральной совокупности разностей, *n* — объем выборки.

Z-статистика имеет стандартизованное нормальное распределение.

В большинстве ситуаций стандартное отклонение генеральной совокупности неизвестно. Единственным параметром, доступным исследователю, являются выборочные статистики, например, выборочное среднее, выборочная дисперсия и выборочное стандартное отклонение. Если разности предполагаются случайными и независимыми величинами, имеющими нормальное распределение, для оценки разности между математическими ожиданиями зависимых генеральных совокупностей можно применить *t*-критерий. Для этого следует вычислить *t*-статистику, имеющую *t*-распределение с *n – 1* степенями свободы. Несмотря на то что генеральная совокупность предполагается нормально распределенной, на практике при достаточно больших объемах выборки и умеренной асимметрии выборочное распределение средней разности можно аппроксимировать *t*-распределением.

Чтобы проверить нулевую и альтернативную гипотезы: H0: μ1 = μ2 или μ1 – μ2 = 0, H1: μ1 ≠ μ2 или μ1 – μ2 ≠ 0, необходимо вычислить *t*-статистику:

*(5) t =*

где SD =

По определению *t*-статистика имеет *t*-распределение с *n – 1* степенями свободы.

При заданном уровне значимости α нулевая гипотеза отклоняется, если *t*-статистика больше верхнего критического значения *tn – 1* или меньше нижнего критического значения *tn – 1* из *t*-распределения с *n* степенями свободы. Иначе говоря, решающее правило выглядит следующим образом: нулевая гипотеза *Н0* отклоняется, если t > *tn – 1* или t < *tn – 1*; в противном случае нулевая гипотеза не отклоняется.

Чтобы продемонстрировать применение *t*-критерия для оценки разности между двумя математическими ожиданиями, предположим, что некая компания разрабатывает новое программное обеспечение для финансовых расчетов. Поскольку одним из основных критериев качества программного обеспечения является скорость вычислений, разработчики стремятся к тому, чтобы их пакет не уступал по своим возможностям лидерам рынка программ, но превосходил их по скорости расчетов. Если новый пакет окажется эффективным, он будет приводить к тем же результатам, что и другие программы, но за более короткое время.

Для оценки программного обеспечения разработчики провели эксперимент, в ходе которого один и тот же набор задач решали, как с помощью стандартных программ, так и с помощью нового пакета. Поскольку измерения для каждой конкретной задачи проводились согласованно, для оценки эффективности пакета необходимо сравнить не средние значения двух независимых выборок, а среднюю разность между соответствующими элементами (рис. 9).

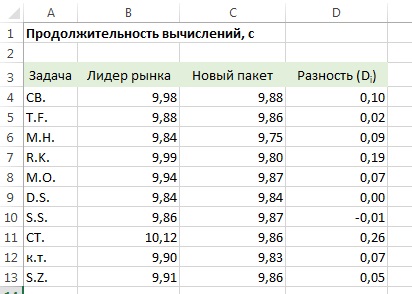


Рис. 9. Попарные измерения продолжительности работы двух конкурирующих пакетов при решении финансовых задач

Можно ли утверждать, что новое программное обеспечение работает быстрее? Иначе говоря, существуют свидетельства того, что на решение финансовых задач стандартный пакет затрачивает больше времени, чем новый? Нулевая и альтернативная гипотеза формулируются следующим образом: H0: μD ≤ 0 (в среднем стандартный пакет работает быстрее, чем новый), H1: μD > 0 (в среднем стандартный пакет работает медленнее, чем новый). Установим уровень значимости α равным 0,05 и предположим, что разности распределены нормально. Это позволяет применить *t*-критерий для парных выборок – формулу (5). Для выборки, состоящей из 10 задач, решающее правило имеет следующий вид: нулевая гипотеза *Н0* отклоняется, если t > t9 =СТЬЮДЕНТ.ОБР(0,05;9) = 1,8331, в противном случае она не отклоняется.

Средняя разность между результатами, полученными в ходе попарных сравнений (рис. 10) *D̅* = 0,084, а стандартное отклонение *SD* = 0,0844, *t* = +3,149. Поскольку значение t = +3,15 лежит в критической области (рис. 11), нулевая гипотеза *Н0* отклоняется. Таким образом, в среднем новый пакет работает быстрее стандартного.

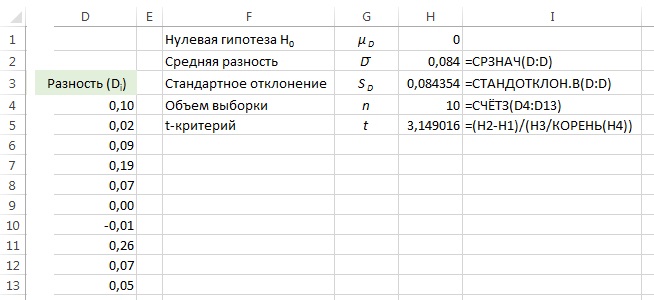


Рис. 10. Расчет *t*-критерия



Рис. 11. Критическая область одностороннего *t*-критерия с 5%-ным уровнем значимости и 9 степенями свободы

Для вычисления *t*-статистики (и *р*-значения) можно воспользоваться Пакетом анализа (рис. 12). Обратите внимание, что в этом случае можно не находить разности (столбец *Разности* (*Di)* не требуется). Пройдите по меню *Данные* → *Анализ данных* и выберите строку *Парный двухвыборочный t-тест для средних*. Поскольку *р*-значение равно 0,006 и меньше α < 0,05, нулевую гипотезу *Н0* следует отклонить. Вычисленное *р*-значение означает следующее: если на самом деле оба пакета имеют одинаковую среднюю продолжительность работы при решении финансовых задач, то вероятность обнаружить превосходство нового пакета более чем на 0,084 с не превышает 0,6%. Поскольку эта величина крайне мала, степень уверенности в нулевой гипотезе весьма невысока, и следует принять альтернативную гипотезу (т.е. стандартный пакет работает медленнее).

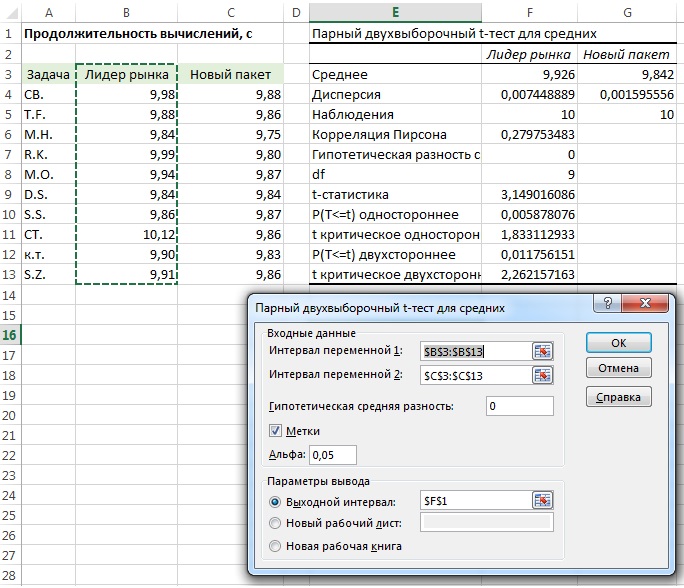


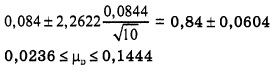
Рис. 12. Расчет *t*-статистики и *р*-значения с помощью опции *Парный двухвыборочный t-тест для средних* Пакета анализа

**Доверительный интервал, содержащий разность между двумя математическими ожиданиями**

Вместо применения парного двухвыборочного *t*-критерия можно построить доверительный интервал, содержащий разность между математическими ожиданиями двух генеральных совокупностей:



Используя формулу (6), в нашем примере получаем: *D̅* =0,084, *SD* = 0,0844, *n* = 10, *t* = 2,2622. В этом случае 95%-ный доверительный интервал имеет следующие границы:

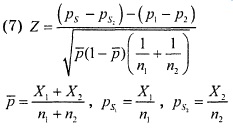


Таким образом, при доверительном уровне 95% средняя разность между результатами измерения эффективности двух пакетов колеблется в интервале от 0,0236 до 0,1444 с. Поскольку ноль не принадлежит этому интервалу, следует сделать вывод, что эффективность нового пакета выше.

**Использование *Z*-критерия для оценки разности между двумя долями признака**

Иногда необходимо выполнить анализ различий между двумя генеральными совокупностями, используя категорийные данные. Оценку разности между двумя долями признака в независимых выборках можно осуществить двумя способами. В данной заметке мы рассмотрим процедуру, в которой тестовая *Z*-статистика аппроксимируется стандартизованным нормальным распределением. Позже описывается процедура, в которой используется тестовая χ2-статистика, аппроксимированная χ2-распределением с одной степенью свободы. Как мы убедимся, эти два критерия эквивалентны.

Для оценки различий между двумя генеральными совокупностями на основе независимых выборок можно применять *Z*-критерий. На основе разности между двумя выборочными долями признака *Ps1* – *Ps2* вычисляется *Z*-статистика, используемая для оценки разности между двумя долями признака в генеральных совокупностях. Если объем выборок достаточно велик, эта тестовая статистика имеет стандартизованное нормальное распределение. Z-критерий для оценки разности между двумя долями:



где *ps1* — доля успехов в первой выборке, *Х1* — количество успехов в первой выборке, *n1* — объем выборки из первой генеральной совокупности, *p1* — доля успехов в первой генеральной совокупности, *ps2* — доля успехов во второй выборке, *Х2* — количество успехов во второй выборке, *n2* — объем выборки из второй генеральной совокупности, *р2* — доля успехов во второй генеральной совокупности, *р̅* – оценка доли успехов в объединенной генеральной совокупности.

При достаточно большом объеме выборок тестовая *Z*-статистика подчиняется стандартизованному нормальному распределению.

Нулевая гипотеза заключается в том, что доли признака в двух генеральных совокупностях одинаковы. Следовательно, проверку равенства долей признака в двух генеральных совокупностях можно свести к оценке доли признака в объединенной генеральной совокупности. Оценка объединенной доли равна результату деления количества успехов в обеих выборках *Х1+Х2* на сумму объемов выборок *n1+n2*.

С помощью *Z*-критерия можно определить, существуют ли различия между долями успеха в двух группах (двусторонний тест), а также установить, превышает ли доля успехов в одной группе долю успехов в другой (односторонний критерий) (рис. 13).

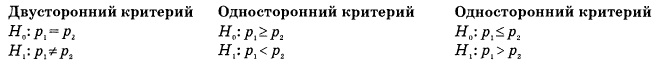


Рис. 13. Три варианта *Z*-критерия

Чтобы проверить нулевую и альтернативные гипотезы H0: р1 = р2, H1: р1 ≠ р2, следует использовать тестовую *Z*-статистику – формулы (7). При заданном уровне значимости α нулевая гипотеза отклоняется, если вычисленная *Z*-статистика больше верхнего или меньше нижнего критического значения стандартизованного нормального распределения.

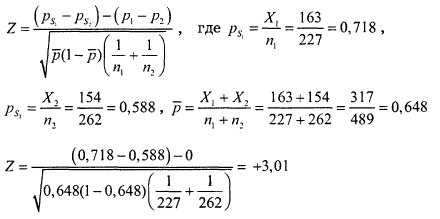
Для того чтобы проиллюстрировать *Z*-критерий для проверки гипотезы о равенстве двух долей, предположим, вы — менеджер компании Т. С. Resort Properties. На одном из островов компании Т. С. Resort Properties принадлежат два отеля: Beachcomer и Windsurfer. На вопрос «Планируете ли вы вернуться в наш отель снова?» 163 из 227 постояльцев отеля Beachcomer ответили: «Да», в то же время 154 из 262 постояльцев отеля Windsurfer на этот вопрос ответили: «Нет». Можно ли утверждать, что при уровне значимости, равном 0,05, между степенью удовлетворенности постояльцев обоих отелей (вероятностью, что в следующем сезоне они вернутся в отель) значимой разницы нет? Нулевая и альтернативная гипотезы формулируются следующим образом: H0: р1 = р2, H1: р1 ≠ р2.

Поскольку уровень значимости равен 0,05, критические значения *ZL* =НОРМ.СТ.ОБР(0,025) = –1,96 и *ZU* =НОРМ.СТ.ОБР(0,975) = +1,96 (рис. 14), а решающее правило имеет следующий вид: нулевая гипотеза *Н0* отклоняется, если Z < –1,96 или Z > +1,96, в противном случае нулевая гипотеза *Н0* не отклоняется.



Рис. 14. Проверка гипотезы о разности между двумя долями при уровне значимости α = 0,05

Вычислим *Z*-статистику:



При уровне значимости, равном 0,05, нулевая гипотеза *Н0* отклоняется, поскольку Z = +3,01 > +1,96. Если нулевая гипотеза является истинной, вероятность того, что *Z*-статистика будет больше +1,96 и меньше –1,96 стандартного отклонения от центра *Z*-распределения, равна 0,05. Наблюдаемый уровень значимости представляет собой вероятность того, что разность между двумя выборочными долями р(Z = 3,01) =(1-НОРМ.СТ.РАСП(3,01;ИСТИНА))\*2 = 0,00262. Таким образом, можно утверждать, что два отеля значительно различаются по качеству обслуживания. Иначе говоря, доля гостей, планирующих вернуться, в отеле Beachcomer больше, чем в гостинице Windsurfer.

**Использование *F*-критерия для оценки разности между двумя дисперсиями**

Довольно часто возникает необходимость проверить, имеют ли две независимые генеральные совокупности одинаковую дисперсию. Например, это требуется для того, чтобы выбрать правильный *t*-критерий — использующий суммарную или раздельную дисперсию. Проверка разности между дисперсиями двух генеральных совокупностей основана на исследовании их отношения. Если каждая генеральная совокупность является нормально распределенной, отношение S12/S22 подчиняется *F*-распределению, получившему свое название в честь знаменитого статистика Р. Фишера. Критическое значение *F*-распределения зависит от двух множеств степеней свободы. Степени свободы числителя относятся к первой выборке, а степени свободы знаменателя — ко второй. Для проверки равенства двух дисперсий в критерии используется *F*-статистика, вычисляемая по формуле:

*(8) F =*

где *S12* — дисперсия выборки из первой генеральной совокупности, *n1* — объем выборки, извлеченной из первой генеральной совокупности, *S22* — дисперсия выборки из второй генеральной совокупности, *n2* — объем выборки, извлеченной из второй генеральной совокупности, *n1 – 1* — количество степеней свободы числителя, *n2 – 1* — количество степеней свободы знаменателя.

*F*-статистика имеет *F*-распределение с *n1 – 1* и *n2 – 1* степенями свободы. При заданном уровне значимости α нулевая и альтернативная гипотеза: H0: σ12 = σ22, H1: σ12 ≠ σ22. Если *F*-статистика больше верхнего критического значения *FU* или меньше нижнего критического значения *FL* из *F*-распределения с *n1 – 1* степенями свободы в числителе и *n2 – 1* степенями свободы в знаменателе, нулевая гипотеза отклоняется. Таким образом, решающее правило выглядит следующим образом: нулевая гипотеза *Н0* отклоняется, если F > *FU* или F < *FL*; в противном случае нулевая гипотеза не отклоняется (рис. 15).



Рис. 15. Критическая область двустороннего F-критерия

Продемонстрируем применение *F*-критерия на примере сценария, описанного выше. Напомним, что в нем требовалось определить, совпадают ли средние объемы продаж BLK-колы, выставленной на обычных полках и специализированных стеллажах. Чтобы выбрать правильный *t*-критерий (с суммарной или раздельной дисперсией), необходимо сначала проверить гипотезу о равенстве дисперсий двух генеральных совокупностей. Следовательно, нулевая и альтернативная гипотеза формулируются так: H0: σ12 = σ22, H1: σ12 ≠ σ22.

Поскольку критерий является двусторонним, критическая область разбивается на две части, ограниченные левым и правым хвостом *F*-распределения. Если уровень значимости α = 0,05, каждая из этих областей соответствует вероятности, равной 0,025. Поскольку выборки содержат по 10 магазинов с разными видами полок, в первой и второй группах существуют 10 – 1 = 9 степеней свободы. Верхнее критическое значения *F*-распределения =F.ОБР(0,975;9;9) = 4,026; нижнее критическое значения *F*-распределения =F.ОБР(0,025;9;9) = 0,248 (рис. 16).



Рис. 16. Критическая область двустороннего *F*-критерия с уровнем значимости, равным 0,05, и 9 степенями свободы в числителе и знаменателе

Таким образом, решающее правило: нулевая гипотеза *Н0* отклоняется, если F > *FU* =4,026 или F < *FL* = 0,248, в противном случае нулевая гипотеза не отклоняется.

Для расчета статистик в нашем примере снова удобно воспользоваться Пакетом анализа, выбрав строку *Двухвыборочный F-тест для дисперсии* (рис. 17). Поскольку *FL* = 0,248 < F = 2,229 < *FU* =4,026 у нас нет оснований отклонять нулевую гипотезу. Если необходимо применить подход, основанный на определении *р*-значения, то параметр *р(F<=f)* также выводится при расчете (см. соответствующую строку на рис. 17). Поскольку *р*-значение для двустороннего критерия равно 0,248 (удвоенное *р*-значение для одностороннего критерия), приходим к выводу, что продажи колы с разных стеллажей обладают практически одинаковой изменчивостью. Итак, *t*-критерий для сравнения математических ожиданий двух групп на основе суммарной дисперсии является вполне корректным.

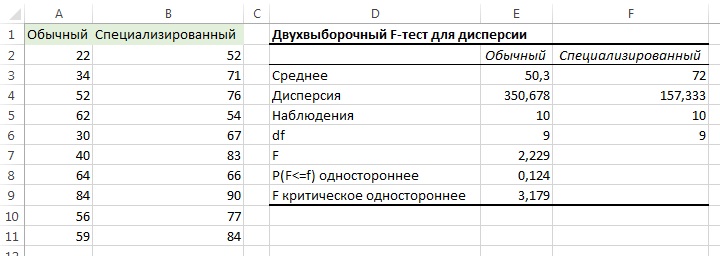


Рис. 17. Расчет F-статистики с помощью Пакета анализа

При оценке разности между двумя дисперсиями с помощью *F*-критерия предполагается, что обе генеральные совокупности имеют нормальное распределение. *F*-критерий очень чувствителен к нарушению этого условия. Если блочная диаграмма или график нормального распределения демонстрируют значительное отклонение от указанного требования, *F*-критерий применять нельзя. В таких ситуациях следует применять непараметрические процедуры.

При выборе разновидности *F*-критерия, как правило, применяется двусторонний *F*-критерий. Однако, если исследователя интересует собственно изменчивость данных, можно применять односторонний *F*-критерий. Таким образом, для сравнения дисперсии двух генеральных совокупностей можно применять как двусторонний, так и односторонний *F*-критерии. Эти ситуации изображены на рис. 18.

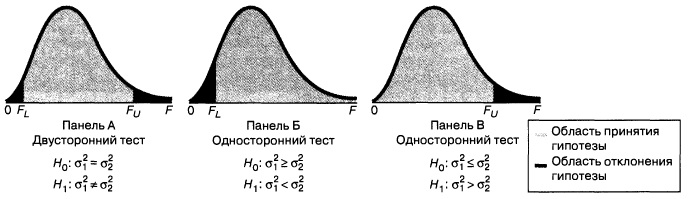


Рис. 18. Критические области при проверке гипотез о равенстве дисперсий двух генеральных совокупностей

Довольно часто объемы выборок не равны. Например, из нормально распределенной генеральной совокупности извлечена выборка, имеющая объем *n1* = 8. Дисперсия *S12* этой выборки равна 56,0. Из второй распределенной генеральной совокупности, независимой от первой, извлечена выборка, имеющая объем *n2* = 10. Дисперсия *S22* этой выборки равна 24,0. Проверьте нулевую гипотезу, заключающуюся в том, что между дисперсиями этих генеральных совокупностей нет существенной разницы.

Верхнее критическое значения *F*-распределения FU =F.ОБР(0,975;7;9) = 4,197; нижнее критическое значения *F*-распределения FL =F.ОБР(0,025;7;9) = 0,207. *F*-статистика = *S12/ S22* = 56 / 24 = 2,333. Решающее правило выглядит так: нулевая гипотеза *Н0* отклоняется, если F > FU = 4,197 или F < FL = 0,207; в противном случае нулевая гипотеза не отклоняется. Поскольку *F*-статистика, равная 2,333, лежит в интервале между *FL* = 0,207 и *FU* = 4,197, нулевую гипотезу *Н0* отклонять нельзя. Итак, при заданном уровне значимости α = 0,05 между дисперсиями двух независимых генеральных совокупностей нет статистически значимой разницы. Хотя дисперсия одной из генеральных совокупностей в 2,33 раза превышает другую, этот факт может оказаться случайным.

**Резюме**

Основное различие между критериями для сравнения двух групп заключается в свойствах генеральных совокупностей: независимы они или взаимозависимы, а также в особенностях исследуемой переменной, числовой или категорийной. Классифицировав критерии по группам, необходимо обратить особое внимание на условия, которые должны выполняться при их применении (рис. 19).

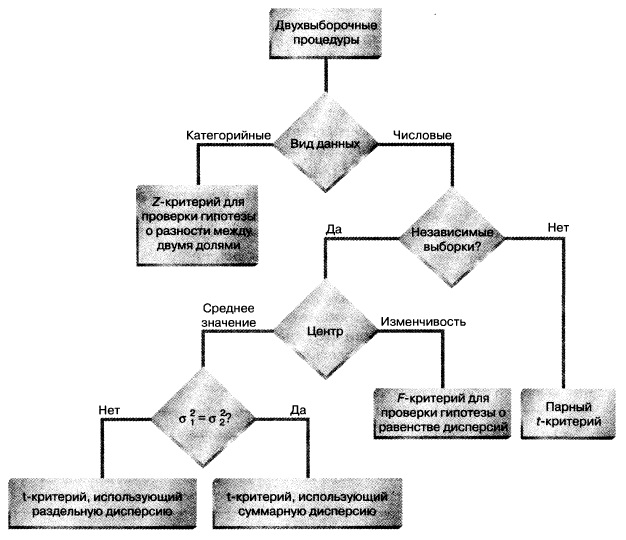


Рис. 19. Структурная схема заметки

Предыдущая заметка [Проверка гипотез: одновыборочные критерии](http://baguzin.ru/wp/?p=5790)

Следующая заметка

К оглавлению [Статистика для менеджеров с использованием Microsoft Excel](http://baguzin.ru/wp/?p=5285)

1. Используются материалы книги Левин и др. Статистика для менеджеров. – М.: Вильямс, 2004. – с. 579–640 [↑](#footnote-ref-1)
2. Если объем выборки достаточно велик, центральная предельная теорема утверждает, что средняя разность *D̅* имеет нормальное распределение. [↑](#footnote-ref-2)