**Однофакторный дисперсионный анализ**

В [предыдущей заметке](http://baguzin.ru/wp/?p=5832) были рассмотрены методы проверки гипотез, применяемые для анализа возможных разностей между параметрами двух групп. Однако зачастую необходимо оценить разности между параметрами нескольких групп. Например, может возникнуть необходимость сравнить альтернативные материалы, методы или условия проведения эксперимента на основе заранее установленных критериев. Настоящая заметка посвящена полностью рандомизированному плану эксперимента, в котором рассматривается только один фактор и несколько групп (например, тип шины, рыночная стратегия, марка лекарства или разные поставщики).[[1]](#footnote-1)

Применение статистики в этой заметке будет показано на сквозном примере. Предположим, что вы — руководитель производства в компании Perfect Parachute («Идеальный парашют»). Парашюты изготавливаются из синтетических волокон, поставляемых четырьмя разными поставщиками. Одной из основных характеристик парашюта является его прочность. Вам необходимо убедиться, что все поставляемые волокна обладают одинаковой прочностью. Чтобы ответить на этот вопрос, следует разработать схему эксперимента, в ходе которого измеряется прочность парашютов, сотканных из синтетических волокон разных поставщиков. Информация, полученная в ходе этого эксперимента, позволит определить, какой поставщик обеспечивают наибольшую прочность парашютов.

Многие приложения связаны с экспериментами, в которых рассматривается несколько групп или уровней одного фактора. Некоторые факторы, например, температура обжига керамики, могут иметь несколько числовых уровней (т.е. 300°, 350°, 400° и 450°). Другие факторы, например, местоположение товаров в супермаркете, могут иметь категориальные уровни (например, первый поставщик, второй поставщик, третий поставщик, четвертый поставщик). Однофакторные эксперименты, в ходе которых экспериментальные единицы случайным образом распределяются по группам или уровням фактора, называются полностью рандомизированными.

**Использование F-критерия для оценки разностей между несколькими математическими ожиданиями**

Если числовые измерения фактора в группах являются непрерывными и выполняются некоторые дополнительные условия, для сравнения математических ожиданий нескольких групп применяется дисперсионный анализ (ANOVA — **An**alysis **o**f **Va**riance). Дисперсионный анализ, использующий полностью рандомизированные планы, называется однофакторной процедурой ANOVA. В некотором смысле термин дисперсионный анализ является неточным, поскольку при этом анализе сравниваются разности между математическими ожиданиями групп, а не между дисперсиями. Однако сравнение математических ожиданий осуществляется именно на основе анализа вариации данных. В процедуре ANOVA полная вариация результатов измерений подразделяется на межгрупповую и внутригрупповую (рис. 1). Внутригрупповая вариация объясняется ошибкой эксперимента, а межгрупповая — эффектами условий эксперимента. Символ *с* обозначает количество групп.



Рис. 1. Разделение вариации в полностью рандомизированном эксперименте

Предположим, что *с* групп извлечено из независимых генеральных совокупностей, имеющих нормальное распределение и одинаковую дисперсию. Нулевая гипотеза заключается в том, что математические ожидания генеральных совокупностей одинаковы: *Н0: μ1 = μ2 = ... = μс*. Альтернативная гипотеза гласит, что не все математические ожидания одинаковы: *Н1*: не все μj одинаковы *j* = 1, 2, …, с).

На рис. 2 представлена истинная нулевая гипотеза о математических ожиданиях пяти сравниваемых групп при условии, что генеральные совокупности имеют нормальное распределение и одинаковую дисперсию. Пять генеральных совокупностей, связанных с разными уровнями фактора, идентичны. Следовательно, они накладываются одна на другую, имея одинаковые математическое ожидание, вариацию и форму.



Рис. 2. Пять генеральных совокупностей имеют одинаковое математическое ожидание: *μ1 = μ2 = μ3 = μ4 = μ5*

С другой стороны, предположим, что на самом деле нулевая гипотеза является ложной, причем четвертый уровень имеет наибольшее математическое ожидание, первый уровень — чуть меньшее математическое ожидание, а остальные уровни — одинаковые и еще меньшие математические ожидания (рис. 3). Обратите внимание на то, что за исключением величины математических ожиданий все пять генеральных совокупностей идентичны (т.е. имеют одинаковую изменчивость и форму).



Рис. 3. Наблюдается эффект условий эксперимента: *μ4 > μ1 > μ2 = μ3 = μ5*

При проверке гипотезы о равенстве математических ожиданий нескольких генеральных совокупностей полная вариация разделяется на две части: межгрупповую вариацию, обусловленную разностями между группами, и внутригрупповую, обусловленную разностями между элементами, принадлежащими одной группе. Полная вариация выражается полной суммой квадратов (SST – sum of squares total). Поскольку нулевая гипотеза заключается в том, что математические ожидания всех *с* групп равны между собой, полная вариация равна сумме квадратов разностей между отдельными наблюдениями и общим средним (среднее средних) X̿, вычисленным по всем выборкам. Полная вариация:



где — общее среднее, *Xij* — *i*-e наблюдение в *j*-й группе или уровне, *nj* — количество наблюдений в *j*-й группе, *n* — общее количество наблюдений во всех группах (т.е. *n* = *n1* + *n2* + ... + *nc*), *с* — количество изучаемых групп или уровней.

*Межгрупповая вариация*, называемая обычно межгрупповой суммой квадратов (SSA – sum of squares among groups), равна сумме квадратов разностей между выборочным средним каждой группы *Х̅j* и общим средним *X̿*, умноженных на объем соответствующей группы *nj*:



где *с* — количество изучаемых групп или уровней, *nj* — количество наблюдений в *j*-й группе, *X̅j* — среднее значение *j*-й группы, *X̿* — общее среднее.

*Внутригрупповая вариация*, называемая обычно внутригрупповой суммой квадратов (SSW – sum of squares withing groups), равна сумме квадратов разностей между элементами каждой группы и выборочным средним этой группы *X̅j*:



где *Хij* — *i*-й элемент *j*-й группы, *Х̅j* — среднее значение *j*-й группы.

Поскольку сравнению подвергаются *с* уровней фактора, межгрупповая сумма квадратов имеет *с – 1* степеней свободы. Каждый из *с* уровней обладает *nj – 1* степенями свободы, поэтому Внутригрупповая сумма квадратов имеет *n – с* степеней свободы, и

$$\sum\_{j=1}^{c}(n\_{j} –1)=n –c$$

Кроме того, общая сумма квадратов имеет *n – 1* степеней свободы, поскольку каждое наблюдение *Хij* сравнивается с общим средним *X̿*, вычисленным по всем *n* наблюдениям. Если каждую из этих сумм разделить на соответствующее количество степеней свободы, возникнут три вида дисперсии: *межгрупповая* (mean square among — MSA), *внутригрупповая* (mean square within — MSW) и *полная* (mean square total — MST):

*(4а) MSA =* $\frac{SSA}{c – 1}$ *(4б) MSW =* $\frac{SSW}{n – c}$ *(4в) MST =* $\frac{SST}{n – 1}$

Несмотря на то что основное предназначение дисперсионного анализа — сравнить математические ожидания *с* групп, чтобы выявить эффект условий эксперимента, его название обусловлено тем, что главным инструментом является анализ дисперсий разного типа. Если нулевая гипотеза является истинной, и между математическими ожиданиями *с* групп нет существенных различий, все три дисперсии — MSA, MSW и MST — являются оценками дисперсии *σ2*, присущей анализируемым данным. Таким образом, чтобы проверить нулевую гипотезу *Н0: μ1 = μ2 = ... = μс* и альтернативную гипотезу *Н1*: не все μj одинаковы *j* = 1, 2, …, *с*), необходимо вычислить статистику *F*-критерия, представляющую собой отношение двух дисперсий, MSA и MSW. Тестовая *F*-статистика в однофакторном дисперсионном анализе

*(5) F =* $\frac{MSA}{MSW}$

Статистика *F*-критерия подчиняется *F*-распределению с *с – 1* степенями свободы в числителе *MSA* и *n – с* степенями свободы в знаменателе *MSW*. При заданном уровне значимости α нулевая гипотеза отклоняется, если вычисленная *F*-статистика больше верхнего критического значения *FU*, присущего *F*-распределению с *с – 1* степенями свободы в числителе и *n – с* степенями свободы в знаменателе. Таким образом, как показано на рис. 4, решающее правило формулируется следующим образом: нулевая гипотеза *Н0* отклоняется, если *F > FU*; в противном случае она не отклоняется.



Рис. 4. Критическая область дисперсионного анализа при проверке гипотезы *Н0*

Если нулевая гипотеза *Н0* является истинной, вычисленная *F*-статистика близка к 1, поскольку ее числитель и знаменатель являются оценками одной и той же величины — дисперсии σ2, присущей анализируемым данным. Если нулевая гипотеза *Н0* является ложной (и между математическими ожиданиями разных групп существует значительная разница), вычисленная *F*-статистика будет намного больше единицы, поскольку ее числитель, MSA, помимо естественной изменчивости данных, оценивает эффект условий эксперимента или разности между группами, в то время как знаменатель MSW оценивает лишь естественную изменчивость данных. Таким образом, процедура ANOVA представляет собой *F*-критерий, в котором при заданном уровне значимости α нулевая гипотеза отклоняется, если вычисленная *F*-статистика больше верхнего критического значения *FU*, присущего *F*-распределению с *с – 1* степенями свободы в числителе и *n – с* степенями свободы в знаменателе, как показано на рис. 4.

Для иллюстрации однофакторного дисперсионного анализа вернемся к сценарию, изложенному в начале заметки. Цель эксперимента — определить, имеют ли парашюты, сотканные из синтетического волокна, полученного от разных поставщиков, одинаковую прочность. В каждой из групп соткано по пять парашютов. Группы разделены по поставщикам— Поставщик 1, Поставщик 2, Поставщик 3 и Поставщик 4. Прочность парашютов измеряется с помощью специального устройства, испытывающего ткань на разрыв с двух сторон. Сила, необходимая для разрыва парашюта, измеряется по особой шкале. Чем выше сила разрыва, тем прочнее парашют. [Пакет анализа](http://baguzin.ru/wp/?p=5316) Excel позволяет провести анализ *F*-статистики одним кликом. Пройдите по меню *Данные* → *Анализ данных*, и выберите строку *Однофакторный дисперсионный анализ*, заполните открывшееся окно (рис. 5). Результаты эксперимента (сила разрыва), некоторые описательные статистики и результаты однофакторного дисперсионного анализа представлены на рис. 6.



Рис. 5. Окно *Однофакторный дисперсионный анализ Пакета анализа* Excel



Рис. 6. Показатели прочности парашютов, сотканных из синтетических волокон, полученных от разных поставщиков, описательные статистики и результаты однофакторного дисперсионного анализа

Анализ рисунка 6 показывает, что между выборочными средними наблюдается некоторая разница. Средняя прочность волокон, полученных от первого поставщика, равна 19,52, от второго — 24,26, от третьего — 22,84 и от четвертого — 21,16. Можно ли назвать эту разницу статистически значимой? Распределение силы разрыва продемонстрировано на диаграмме разброса (рис. 7). На ней ясно видны разности как между группами, так и внутри них. Если бы объем каждой группы был больше, для их анализа можно было бы применить диаграмму «ствол и листья», блочную диаграмму или график нормального распределения.



Рис. 7. Диаграмма разброса прочности парашютов, сотканных из синтетических волокон, полученных от четырех поставщиков

Нулевая гипотеза утверждает, что между средними показателями прочности нет существенных различий: *Н0: μ1 = μ2 = μ3 = μ4*. Альтернативная гипотеза заключается в том, что существует по крайней мере один поставщик, у которого средняя прочность волокон отличается от других: *Н1*: не все μj одинаковы *j* = 1, 2, …, *с*).

Общее среднее (см. рис. 6) *X̿* =СРЗНАЧ(D12:D15) = 21,945; для определения *X̿* также можно усреднить все 20 исходных чисел: *X̿* =СРЗНАЧ(A3:D7). Значения дисперсий рассчитываются *Пакетом анализа* и отражаются в табличке *Дисперсионный анализ* (см. рис. 6): SSA = 63,286, SSW = 97,504, SST = 160,790 (см. колонку *SS* таблицы *Дисперсионный анализ* рисунка 6). Средние значения вычисляются путем деления этих сумм квадратов на соответствующее количество степеней свободы. Поскольку *с* = 4, а *n* = 20, получаем следующие значения степеней свободы; для SSA: *с – 1* = 3; для SSW: *n – c* = 16; для SST: *n – 1* = 19 (см. колонку *df*). Таким образом: MSA = SSA / (*с – 1)* = 21,095; MSW = SSW / (*n – c)* = 6,094; MST = SST / (*n – 1*) = 8,463 (см. колонку *MS*). *F*-статистика = MSA / MSW = 3,462 (см. колонку *F*).

Верхнее критическое значение *FU*, характерное для *F*-распределения, определяется по формуле =F.ОБР(0,95;3;16) = 3,239. Параметры функции =F.ОБР(): α = 0,05, числитель имеет три степени свободы, а знаменатель — 16. Таким образом, вычисленная *F*-статистика, равная 3,462, превышает верхнее критическое значение *FU* = 3,239, нулевая гипотеза отклоняется (рис. 8).



Рис. 8. Критическая область дисперсионного анализа при уровне значимости, равном 0,05, если числитель имеет три степени свободы, а знаменатель —16

*р*-значение, т.е. вероятность того, что при истинной нулевой гипотезе *F*-статистика не меньше 3,46, равно 0,041 или 4,1% (см. колонку *р-Значение* таблицы *Дисперсионный анализ* рисунка 6). Поскольку эта величина не превышает уровень значимости α = 5%, нулевая гипотеза отклоняется. Более того, *р*-значение свидетельствует о том, что вероятность обнаружить такую или большую разность между математическими ожиданиями генеральных совокупностей при условии, что на самом деле они одинаковы, равна 4,1%.

Итак. Между четырьмя выборочными средними существует разница. Нулевая гипотеза заключалась в том, что все математические ожидания четырех генеральных совокупностей равны между собой. В этих условиях мера полной изменчивости (т.е. полная вариация SST) прочности всех парашютов вычисляется путем суммирования квадратов разностей между каждым наблюдением *Xij* и общим средним *X̿*. Затем полная вариация разделялась на два компонента (см. рис. 1). Первый компонент представлял собой межгрупповую вариацию SSA, а второй — внутригрупповую SSW.

Чем объясняется изменчивость данных? Иначе говоря, почему все наблюдения не одинаковы? Одна из причин заключается в том, что разные фирмы поставляют волокна разной прочности. Это частично объясняет, почему группы имеют разные математические ожидания: чем сильнее эффект условий эксперимента, тем больше разность между математическими ожиданиями групп. Другой причиной изменчивости данных является естественная изменчивость любого процесса, в данном случае — производства парашютов. Даже если бы все волокна приобретались у одного и того же поставщика, их прочность была бы неодинаковой при прочих равных условиях. Поскольку этот эффект проявляется в каждой из групп, он называется внутригрупповой вариацией.

Разности между выборочными средними называются межгрупповой вариацией SSA. Часть внутригрупповой вариации, как уже указывалось, объясняется принадлежностью данных разным группам. Однако даже если бы группы были совершенно одинаковыми (т.е. нулевая гипотеза была бы истинной), межгрупповая вариация все равно существовала. Причина этого заключается в естественной изменчивости процесса производства парашютов. Поскольку выборки разные, их выборочные средние отличаются друг от друга. Следовательно, если нулевая гипотеза является истинной, как межгрупповая, так и внутригрупповая изменчивость представляют собой оценку изменчивости генеральной совокупности. Если нулевая гипотеза является ложной, межгрупповая гипотеза будет больше. Именно этот факт лежит в основе *F*-критерия для сравнения разностей между математическими ожиданиями нескольких групп.

После выполнения однофакторного дисперсионного анализа и обнаружения значительной разницы между фирмами остается неизвестным, какой же из поставщиков существенно отличается от остальных. Нам известно лишь, что математические ожидания генеральных совокупностей не равны. Иначе говоря, по крайней мере одно из математических ожиданий существенно отличается от других. Чтобы определить, какой из поставщиков отличается от других, можно воспользоваться *процедурой Тьюки*, использующей попарное сравнение между поставщиками. Эта процедура была разработана Джоном Тьюки. Впоследствии он и К. Крамер независимо друг от друга модифицировали эту процедуру для ситуаций, в которых объемы выборок отличаются друг от друга.

**Множественное сравнение: процедура Тьюки-Крамера**

В нашем сценарии для сравнения прочности парашютов использовался однофакторный дисперсионный анализ. Обнаружив значительные различия между математическими ожиданиями четырех групп, необходимо определить, какие именно группы отличаются друг от друга. Хотя существует несколько способов решить эту задачу, мы опишем лишь процедуру множественного сравнения Тьюки-Крамера. Этот метод является примером процедур апостериорного сравнения (post hoc comparison), поскольку проверяемая гипотеза формулируется после анализа данных. Процедура Тьюки-Крамера позволяет одновременно сравнить все пары групп. На первом этапе вычисляются разности *Xj – Xj’*, где *j ≠ j’*, между математическими ожиданиями *с(с – 1)/2* групп. *Критический размах* процедуры Тьюки-Крамера вычисляется по формуле:

*(6) Критический размах = QU*$\sqrt{\frac{MSW}{2}(\frac{1}{n\_{j}}+\frac{1}{n\_{j'}})}$

где *QU* — верхнее критическое значение распределения стьюдентизированного размаха, имеющего *с* степеней свободы в числителе и *n – с* степеней свободы в знаменателе.

Если объемы выборок не одинаковы, критический размах вычисляется для каждой пары математических ожиданий отдельно. На последнем этапе каждая из *с(с – 1)/2* пар математических ожиданий сравнивается с соответствующим критическим размахом. Элементы пары считаются значимо различными, если модуль разности |*Xj – Xj’*| между ними превышает критический размах.

Применим процедуру Тьюки-Крамера к задаче о прочности парашютов. Поскольку компания, производящая парашюты, имеет четыре поставщика, следует проверить 4(4 – 1)/2 = 6 пар поставщиков (рис. 9).



Рис. 9. Попарные сравнения выборочных средних

Поскольку все группы имеют одинаковый объем (т.е. все *nj = nj’*), достаточно вычислить только один критический размах. Для этого по таблице *Дисперсионного анализа* (рис. 6) определим величину MSW = 6,094. Затем найдем величину *QU* при α = 0,05, *с* = 4 (число степеней свободы в числителе) и *n – с* = 20 – 4 = 16 (число степеней свободы в знаменателе). К сожалению, я не нашел соответствующей функции в Excel, так что воспользовался таблицей (рис. 10).



Рис. 10. Критическое значение стьюдентизированного размаха *QU*

Получаем:

*Критический размах = 4,05*$\sqrt{\frac{6,094}{2}(\frac{1}{5}+\frac{1}{5})}$ *= 4,47*

Поскольку лишь 4,74 > 4,47 (см. нижнюю таблицу рис. 9), статистически значимая разница существует между первым и вторым поставщиком. Все остальные пары имеют выборочные средние, которые не позволяют говорить о их различии. Следовательно, средняя прочность парашютов, сотканных из волокон, приобретенных у первого поставщика, значимо меньше, чем у второго.

**Необходимые условия однофакторного дисперсионного анализа**

При решении задачи о прочности парашютов мы не проверяли, выполняются ли условия, при которых можно использовать однофакторный *F*-критерий. Как же узнать, можно ли применять однофакторный *F*-критерий при анализе конкретных экспериментальных данных? Однофакторный *F*-критерий можно применять, только если выполняются три основных предположения: экспериментальные данные должны быть случайными и независимыми, иметь нормальное распределение, а их дисперсии должны быть одинаковыми.

Первое предположение — *случайность и независимость данных* — должно выполняться всегда, поскольку корректность любого эксперимента зависит от случайности выбора и/или процесса рандомизации. Чтобы избежать искажения результатов, необходимо, чтобы данные извлекались из *с* генеральных совокупностей случайно и независимо друг от друга. Аналогично данные должны быть случайным образом распределенными по *с* уровням интересующего нас фактора (экспериментальным группам). Нарушение этих условий может серьезно исказить результаты дисперсионного анализа.

Второе предположение — *нормальность* — означает, что данные извлечены из нормально распределенных генеральных совокупностей. Как и для *t*-критерия, однофакторный дисперсионный анализ на основе *F*-критерия относительно мало чувствителен к нарушению этого условия. Если распределение не слишком значительно отличается от нормального, уровень значимости *F*-критерия изменяется мало, особенно если объем выборок достаточно велик. Если же условие о нормальности распределения нарушается серьезно, следует применять непараметрические процедуры дисперсионного анализа (будут рассмотрены позже).

Третье предположение — *однородность дисперсии* — означает, что дисперсии каждой генеральной совокупности равны между собой (т.е. σ12 = σ22 = … = σj2). Это предположение позволяет решить, разделять или объединять внутригрупповые дисперсии. Если объемы групп совпадают, условие однородности дисперсии слабо влияет на выводы, полученные с помощью *F*-критерия. Однако, если объемы выборок неодинаковы, нарушение условия о равенстве дисперсий может серьезно исказить результаты дисперсионного анализа. Таким образом, следует стремиться к тому, чтобы объемы выборок были одинаковыми. Одним из методов проверки предположения об однородности дисперсии является критерий *Левенэ*, описанный ниже.

Если из всех трех условий нарушается лишь условие об однородности дисперсии, можно применять процедуру, аналогичную *t*-критерию, использующему раздельную дисперсию (подробнее см. [Проверка гипотез: двухвыборочные критерии](http://baguzin.ru/wp/?p=5832)). Однако, если предположения о нормальном распределении и однородности дисперсии нарушаются одновременно, необходимо выполнить нормализацию данных и уменьшить разности между дисперсиями или применить непараметрическую процедуру.

**Критерий Левенэ для проверки однородности дисперсии**

Несмотря на то что *F*-критерий относительно устойчив к нарушениям условия о равенстве дисперсий в группах, грубое нарушение этого предположения существенно влияет на уровень значимости и мощность критерия. Возможно, одним из наиболее мощных является критерий *Левенэ*. Для проверки равенства дисперсий *с* генеральных совокупностей проверим следующие гипотезы:

*Н0: σ12 = σ22 = … = σj2*

*Н1*: не все *σj2* одинаковы (*j* = 1, 2, …, *с*)

Модифицированный критерий Левенэ основан на утверждении, что если изменчивость в группах одинакова, для проверки нулевой гипотезы о равенстве дисперсий можно применить анализ дисперсии абсолютных величин разностей между наблюдениями и медианами групп. Итак, сначала следует вычислить абсолютные величины разностей между наблюдениями и медианами в каждой группе, а затем выполнить однофакторный дисперсионный анализ полученных абсолютных величин разностей. Для иллюстрации критерия Левенэ вернемся к сценарию, изложенному в начале заметки. Используя данные, представленные на рис. 6, проведем аналогичный анализ, но в отношении модулей разниц исходных данных и медиан по каждой выборке отдельно (рис. 11).



Таблица 11. Абсолютные величины разностей между медианами прочности волокон и наблюдениями в каждой группе поставщиков

Предыдущая заметка [Проверка гипотез: двухвыборочные критерии](http://baguzin.ru/wp/?p=5832)

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. 641–664 [↑](#footnote-ref-1)