Проверка гипотез: одновыборочные критерии

Ранее была изложена концепция выборочных распределений, которая позднее была использована для построения доверительных интервалов. В настоящей заметке основное внимание уделяется методам проверки гипотез, которые представляют собой часть теории статистического вывода, использующую информацию, содержащуюся в выборке.[1]

> Применение статистики в этой заметке будет показано на сквозном примере **«Процесс расфасовки кукурузных хлопьев».** Будучи управляющим компании Oxford Cereal Company, вы отвечаете за процесс расфасовки кукурузных хлопьев по коробкам. Необходимо убедиться, что конвейер работает нормально, и каждая коробка содержит в среднем 368 г зерна. Для этого вы извлекаете из генеральной совокупности 25 коробок, взвешиваете их и оцениваете отклонение реального веса от номинального. Коробки из этой выборки могут содержать либо слишком мало, либо слишком много хлопьев. В этом случае следует остановить производство и определить причину неполадок. Анализируя разности между реальным весом и номинальным, необходимо решить, равно ли математическое ожидание генеральной совокупности 368 г или нет. Если равно, процесс расфасовки не требует вмешательства, если нет — следует остановить конвейер.

Нулевая и альтернативная гипотеза

Проверка гипотез обычно начинается с некоего утверждения, касающегося конкретного параметра генеральной совокупности. Например, при статистическом анализе процесса расфасовки кукурузных хлопьев естественно предположить, что конвейер работает нормально, и, следовательно, средний вес коробок равен 368 г. Гипотеза о том, что параметр генеральной совокупности равен ожидаемому, называется *нулевой* и обозначается как H_0. В нашем примере нулевая гипотеза заключается в том, что заполнение коробок осуществляется правильно и средний вес коробок равен 368 г. Сформулируем это следующим образом: $H_0: \mu = 368$.

Несмотря на то что нам доступна только информация об отдельной выборке, нулевая гипотеза относится к параметру всей генеральной совокупности, потому что нас интересует процесс расфасовки в целом, а для его оценки используется выборочная статистика. В результате статистического анализа мы можем прийти к выводу, что нулевая гипотеза неверна. Следовательно, необходимо сформулировать ее альтернативу, т.е. гипотезу, которая считается истинной, если нулевая гипотеза оказывается ложной. *Альтернативная* гипотеза H_1 противоположна нулевой гипотезе H_0: $H_1: \mu \neq 368$.

Иными словами, альтернативная гипотеза является отрицанием нулевой. Она оказывается истинной, если существуют статистические данные, свидетельствующие о том, что нулевая гипотеза неверна. Если в нашем примере выяснится, что средний вес коробок в выборке значительно отличается от 368 г, нулевая гипотеза отклоняется, и используется альтернативная гипотеза. В этом случае производство следует остановить и предпринять необходимые действия, направленные на устранение неполадок. Если нулевая гипотеза не отвергается, следует признать, что процесс расфасовки протекает правильно и никакие действия предпринимать не надо. Обратите внимание на то, что во втором сценарии мы не доказываем, что процесс расфасовки выполняется правильно, просто мы не в состоянии доказать обратное и поэтому должны верить (хотя и бездоказательно) в справедливость нулевой гипотезы.

Если выборочная статистика свидетельствует в пользу альтернативной гипотезы, нулевая гипотеза отклоняется. В этом и заключается проверка гипотез. Однако отказ отклонить нулевую гипотезу не означает, что она является истинной. Доказать, что нулевая гипотеза верна, в принципе невозможно, поскольку при ее проверке используется выборка, а не вся генеральная совокупность. Следовательно, отказ отвергнуть нулевую гипотезу означает лишь, что для ее отклонения нет

- Нулевая гипотеза H_0 отражает статус-кво или текущее положение дел.
- Альтернативная гипотеза H_1 является отрицанием нулевой гипотезы и представляет собой исследовательское предположение или особое умозаключение, которое требуется доказать.
- Если нулевая гипотеза отвергается, альтернативная гипотеза считается истинной.

---

[1] Используются материалы книги Левин и др. Статистика для менеджеров. – М.: Вильямс, 2004. – с. 480–578

- Если нулевая гипотеза не отвергается, альтернативная гипотеза считается недоказанной. Однако недоказанность альтернативной гипотезы не означает, что нулевая гипотеза является истинной.
- Нулевая гипотеза H_0 всегда формулируется относительно конкретного значения параметра генеральной совокупности (например, математического ожидания μ), а не выборочной статистики (например, выборочного среднего $\overline{X}$).
- Нулевая гипотеза всегда содержит утверждение о равенстве параметра генеральной совокупности заранее заданному значению (например, $H_0: \mu = 368$).
- Альтернативная гипотеза никогда не содержит утверждения о равенстве параметра генеральной совокупности заранее заданному значению (например, $H_1: \mu \neq 368$).

Критическое значение тестовой статистики

Опишем проверку гипотез на конкретном примере. В сценарии, касающемся компании Oxford Cereal Company, нулевая гипотеза означает, что средний вес коробок равен 368 г (т.е. параметр генеральной совокупности μ равен номинальному). Из генеральной совокупности извлекается выборка, каждая коробка из этой выборки взвешивается, и вычисляется их средний вес. Эта статистика является оценкой соответствующего параметра генеральной совокупности, из которой извлечена выборка. Даже если нулевая гипотеза на самом деле истинна, из-за изменчивости выборочное среднее не может в точности совпадать со средним значением генеральной совокупности. Однако в этом случае можно ожидать, что выборочное среднее будет мало отличаться от математического ожидания генеральной совокупности. Например, если средний выборочный вес коробок равен 367,9, естественно заключить, что математическое ожидание генеральной совокупности очень близко к номинальному (т.е. μ = 368 г). Интуиция подсказывает, что выборка, среднее значение которой равно 367,9, извлечена из генеральной совокупности, математическое ожидание которой равно 368 г.

С другой стороны, если между выборочной статистикой и параметром генеральной совокупности наблюдаются значительные различия, естественно отклонить нулевую гипотезу. Например, если средний выборочный вес коробок равен 320 г, можно прийти к выводу, что математическое ожидание генеральной совокупности не равно номинальному (т.е. $\mu \neq 368$). Следовательно, логично предположить, что математическое ожидание генеральной совокупности не равно 368 г. В любом случае статистический вывод основывается на предположении, что случайные выборки являются репрезентативными и правильно представляют свойства генеральной совокупности, из которых они извлечены.

К сожалению, процесс принятия решения на практике не так прост. Он существенно зависит от субъективного восприятия понятий «большое отклонение» и «небольшое отклонение». Проверка гипотез позволяет формализовать эти понятия и оценить вероятность того, что нулевая гипотеза является истинной. Для этого сначала вычисляется выборочная статистика (например, выборочное среднее), а затем — статистика, положенная в основу критерия, которая, как правило, обладает стандартизованным нормальным или t-распределением.

Области отклонения и принятия гипотез[2]

Распределение статистики, положенной в основу критерия, разделяется на две части — область отклонения гипотезы (иногда называемую критической областью) и область принятия гипотезы (рис. 1). Если тестовая статистика попадает в область принятия гипотезы, нулевую гипотезу отклонить нельзя. В примере, связанном с наполнением коробок, выяснилось, что у менеджера нет основания считать, будто средний вес не равен 368 г. Если тестовая статистика попадает в критическую область, нулевая гипотеза отклоняется. В этом случае менеджер полагает, что средний вес всех коробок не равен 368 г. В критическую область попадают лишь те значения тестовой статистики, при которых нулевая гипотеза неверна. Следовательно, если некое значение тестовой статистики попадает в критическую область, нулевую гипотезу следует отклонить. При проверке гипотез прежде всего следует определить критическое значение тестовой статистики. Это число отделяет область принятия гипотезы от области отклонения гипотезы и зависит от размера критической области. Размер

[2] Напомним, что принятие гипотезы означает лишь, что ее не удалось опровергнуть. По этой причине авторы используют более строгий термин *region of nonrejection*, который для краткости мы перевели как *область принятия гипотезы.*

критической области непосредственно связан с величиной риска, возникающего, когда параметр генеральной совокупности оценивается по выборочным данным.

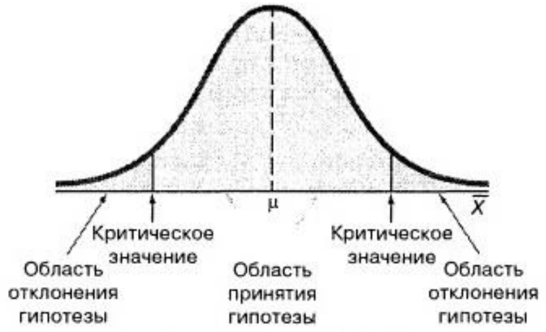

Рис. 1. Области принятия и отклонения гипотез

Риски, возникающие при проверке гипотез

При оценке параметра генеральной совокупности по выборочным значениям существует риск прийти к неверным выводам. При проверке гипотез возможны два типа ошибок: 1- и 2-го рода. В нашем сценарии ошибка 1-го рода возникает, когда менеджер считает, что средний вес всех коробок не равен 368 г, в то время как на самом деле он равен 368 г. С другой стороны, если менеджер придет к выводу, что вес коробок равен 368 г, в то время как на самом деле он не равен 368 г, возникнет ошибка 2-го рода. *Ошибка 1-го рода* возникает, когда отклоняется истинная нулевая гипотеза H_0. Ее вероятность обозначается буквой α. *Ошибка 2-го рода* возникает, когда не отклоняется ложная нулевая гипотеза H_0. Ее вероятность обозначается буквой β.

Уровень значимости (α). Вероятность сделать ошибку 1-го рода обозначается буквой α и называется уровнем значимости статистического критерия. Эта вероятность определяет уровень риска, возникающего при отклонении истинной гипотезы. Поскольку уровень значимости задается заранее, он находится под полным контролем лица, выполняющего проверку. Как правило, уровни значимости равны 0,01, 0,05 и 0,1. Уровень риска зависит от стоимости ошибки 1-го рода. По уровню значимости α можно вычислить размер критической области, а значит, и критическое значение статистики, положенной в основу критерия.

Доверительная вероятность. Вероятность события, противоположного ошибке 1-го рода, называется доверительной вероятностью и равна 1 – α. Умножив ее на 100%, можно вычислить доверительный уровень. Доверительная вероятность 1 – α равна вероятности принять истинную нулевую гипотезу H_0. Доверительный уровень критерия равен (1 – α)x100%. В терминах теории проверки гипотез доверительная вероятность представляет собой вероятность прийти к выводу, что проверяемое значение параметра является достоверным, когда это на самом деле так. В нашем примере доверительная вероятность равна вероятности принять гипотезу, что средний вес коробок равен 368 г, если он действительно равен этому числу.

Уровень риска (β). Вероятность ошибки 2-го рода обозначается буквой β. В отличие от ошибки 1-го рода, которая зависит от уровня значимости α, вероятность ошибки 2-го рода зависит от разности между гипотетическим параметром и фактической выборочной статистикой. Поскольку большую разность легче заметить, чем маленькую, при большой разности вероятность ошибки 2-го рода мала. С другой стороны, если разность между выборочной статистикой и гипотетическим параметром генеральной совокупности мала, вероятность ошибки 2-го рода становится большой.

Мощность критерия. Вероятность противоположного события равна 1 – β и называется мощностью статистического критерия. В нашем сценарии мощность критерия равна вероятности прийти к выводу, что средний вес коробок не равен 368 г, когда он действительно не равен этому числу. Мощность критерия 1 – β равна вероятности отклонить ложную нулевую гипотезу H_0.

Риски, возникающие при принятии решений: точный баланс. В таблице (рис. 2) показаны два возможных решения (принять или отклонить нулевую гипотезу H_0) при проверке гипотез. Как видим, решение может оказаться правильным либо стать причиной ошибки 1-го или 2-го рода.[3]

---

[3] Запомнить соответствие между вероятностями и типами ошибок легко, если учесть, что α — первая буква греческого алфавита и обозначает вероятность ошибки 1-го рода, а β — вторая буква греческого алфавита и обозначает вероятность ошибки 2-го рода.

	Фактическая ситуация	
Статистическое решение	Гипотеза H_0 верна	Гипотеза H_0 неверна
Гипотеза H_0 не отклоняется	Правильное решение	Ошибка 2-го рода
	Доверительная вероятность равна $1-\alpha$	Вероятность ошибки 2-го рода равна β
Гипотеза H_0 отклоняется	Ошибка 1-го рода	Правильное решение
	Вероятность ошибки 1-го рода равна α	Мощность критерия равна $1-\beta$

Рис. 2. Проверка гипотез и принятие решения

Ошибку 1-го рода можно уменьшить, увеличив объем выборки. Более крупные объемы выборки позволяют снизить отклонение выборочных статистик от оцениваемых параметров генеральной совокупности. При заданной ошибке 1-го рода α увеличение объема выборки приводит к уменьшению величины β и, следовательно, к возрастанию мощности критерия. Однако объем выборки нельзя увеличивать бесконечно. Таким образом, необходимо найти компромисс между ошибками двух видов. Поскольку в нашем распоряжении находится лишь вероятность ошибки 1-го рода, следует уменьшить ее величину. Например, если при проверке гипотез ошибка 1-го рода приводит к крайне нежелательным последствиям, необходимо выбрать α = 0,01, а не α = 0,05. Однако при уменьшении величины α увеличивается величина β, следовательно, снижение вероятности ошибки 1-го рода сопровождается увеличением вероятности ошибки 2-го рода. С другой стороны, уменьшая вероятность β, мы увеличиваем вероятность α. Следовательно, если необходимо избежать ошибки 2-го рода, можно выбрать α = 0,05 или α = 0,1, а не α = 0,01.

В нашем сценарии ошибка 1-го рода означает, что менеджер считает процесс заполнения коробок неправильным, в то время как на самом деле он выполняется верно. Ошибка 2-го рода означает, что менеджер считает процесс верным, хотя он выполняется неправильно. Выбор конкретных значений α и β зависит от конкретной стоимости последствий, вызываемых ошибками 1-го и 2-го рода. Например, если процесс заполнения коробок очень трудно перестроить, прежде чем останавливать конвейер, необходимо быть в этом совершенно уверенным. Для этого следует сделать ошибку 1-го рода как можно более маленькой. С другой стороны, если отклонение фактического веса коробок от номинального нельзя допускать ни в коем случае, следует минимизировать ошибку 2-го рода.

Использование Z-критерия для проверки гипотезы о математическом ожидании при известном стандартном отклонении

Напомним, что менеджер компании Oxford Cereal Company должен оценить качество заполнения коробок. Для этого необходимо убедиться, что конвейер работает правильно и средний вес коробок равен 368 г. Для этого из генеральной совокупности извлекается выборка, состоящая из 25 коробок. Затем каждая из них взвешивается, и вычисляется отклонение реального веса от номинального. Нулевая и альтернативная гипотезы формулируются так: H_0: $\mu = 368$, H_1: $\mu \neq 368$. Если стандартное отклонение σ известно, выборочное распределение средних подчиняется нормальному закону. Это позволяет сформулировать Z-критерий:

$$(1)\quad Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Числитель формулы (1) представляет собой отклонение выборочного среднего от математического ожидания генеральной совокупности. Знаменатель равен стандартному отклонению σ, деленному на корень квадратный из объема выборки n. Таким образом, статистика Z выражает разность между $\bar{X}$ и μ, выраженную в единицах стандартного отклонения.

Проверка гипотез с помощью критического значения

Если уровень значимости равен 0,05, то размер критической области также равен 0,05. Следовательно, можно определить критическое значение нормального распределения, выраженное через стандартизованную Z-статистику. Поскольку критическая область разделена на две части (так называемый двусторонний критерий, число 0,05 также следует разделить на два. Таким образом, площадь критической области, ограниченной хвостом гауссовой кривой и нижним критическим значением, равна 0,025. Соответственно, площадь области, ограниченной гауссовой кривой и верхним критическим значением, равна 0,975. А сами критерии легко вычислить с помощью функции Excel =НОРМ.СТ.ОБР(р), где p – вероятность. Z(p=0,025) = –1,96; Z(p=0,975) = 1,96 (рис. 3). Следовательно, если средний вес коробки действительно равен 368 г, как утверждает нулевая гипотеза H_0, Z-статистика имеет стандартизованное нормальное распределение с центром в точке 0 (что соответствует условию $\bar{X}$ =368). Если значение статистики Z меньше –1,96 или больше +1,96, величина $\bar{X}$ настолько далека от μ = 368, что гипотезу H_0 нельзя признать истинной.

Рис. 3. Области принятия и отклонения гипотезы о математическом ожидании при известном стандартном отклонении и уровне значимости, равном 0,05

Следовательно, решающее правило выглядит следующим образом: если Z > +1,96 или Z < –1,96, гипотеза H_0 отклоняется; в противном случае она не отклоняется.

Допустим, что средний вес коробок, содержащихся в выборке из 25 коробок, равен 372,5, а стандартное отклонение равно 15 г. Используя формулу (1), получаем:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{372{,}5 - 368}{\frac{15}{\sqrt{25}}} = +1{,}50$$

Таким образом, гипотезу H_0 отклонять нельзя. Следует признать, что средний вес коробок равен 368 г. Чтобы учесть ошибку 2-го рода, результат необходимо сформулировать так: «Гипотеза о том, что средний вес коробок отличается от 368 г, не имеет достаточных подтверждений».

Проверка гипотезы о математическом ожидании при известном стандартном отклонении выполняется следующим образом:

- Формулируется нулевая гипотеза H_0 о параметрах генеральной совокупности. В задаче о расфасовке кукурузных хлопьев нулевая гипотеза H_0 выглядит так: μ = 368.
- Формулируется альтернативная гипотеза H_1 о параметрах генеральной совокупности. В задаче о расфасовке кукурузных хлопьев альтернативная гипотеза H_1 выглядит так: $\mu \neq 368$.
- Выбирается уровень значимости α. Его конкретная величина определяется относительной важностью риска, вызванного ошибками 1-го и 2-го рода. В нашем примере α = 0,05.
- Определяется объем выборки n, зависящий от величины риска, вызванного ошибками 1-го и 2-го рода (т.е. величин α и β), а также от затрат, необходимых для ее формирования. В нашем примере объем случайной выборки равен 25.
- Выбирается требуемый статистический метод и соответствующая статистика, положенная в основу критерия. Поскольку в нашем примере стандартное отклонение известно заранее, для проверки критерия применяется Z-статистика,
- Устанавливаются критические значения, разделяющие плоскость на области отклонения и принятия гипотезы. В нашем примере критическими значениями являются числа –1,96 и +1,96, поскольку статистика Z имеет стандартизованное нормальное распределение.
- По выборке вычисляется значение статистики, положенной в основу критерия. В нашем примере $\bar{X}$ =372,5, поэтому Z= +1,50.
- Определяется область, в которую попадает вычисленное значение статистики, положенной в основу критерия. Для этого статистика сравнивается с критическими значениями. В нашем

примере Z = +1,50, следовательно, вычисленное значение лежит в области принятия гипотезы, поскольку –1,96 < +1,50 < +1,96.

- Принимается статистическое решение. Если статистика, положенная в основу критерия, попадает в область принятия гипотезы, нулевую гипотезу H_0 отклонять нельзя. В противном случае нулевая гипотеза отклоняется. В нашем примере нулевая гипотеза не была отвергнута.
- Формулируется статистическое решение, учитывающее специфику задачи. В нашем примере гипотеза о том, что средний вес коробок отличается от 368 г, не имеет достаточных подтверждений.

Проверка гипотез по наблюдаемому уровню значимости

В последние годы все большую популярность приобретают критерии проверки гипотез по наблюдаемому уровню значимости, который часто называют *p*-значением. Эта величина соответствует минимальной вероятности того, что нулевая гипотеза H_0 будет отклонена на основе анализа исходного набора данных. Правило отклонения гипотезы H_0 в этом случае выглядит так.

1. Если *p*-значение больше или равно α, нулевая гипотеза не отклоняется.
2. Если p-значение меньше α, нулевая гипотеза отклоняется.

Наблюдаемый уровень значимости, или *p*-значение, представляет собой вероятность того, что тестовая статистика лежит в области отклонения гипотезы при условии, что нулевая гипотеза H_0 верна.

Проиллюстрируем этот подход на нашем примере. Мы по-прежнему хотим знать, равен ли средний вес коробок 368 г. Получив значение Z = +1,50, мы не можем отклонить гипотезу H_0. Применим подход, использующий уровень значимости. Для применения двустороннего критерия необходимо найти вероятность того, что статистика *Z* отстоит от центра стандартизованного нормального распределения не менее, чем на 1,5 стандартных отклонения. Иначе говоря, следует вычислить вероятность того, что статистика Z больше +1,50 или меньше, чем –1,50. Поскольку нормальное распределение симметрично, можно воспользоваться следующей формулой в Excel. Уровень значимости =2*НОРМ.СТ.РАСП(-1,5;ИСТИНА) = 0,1336 (рис. 4). Следовательно, вероятность того, что тестовая статистика лежит в области отклонения гипотезы, равна 0,1336. Поскольку эта величина больше, чем заданная ошибка 1-го рода α, нулевую гипотезу отклонять нельзя.

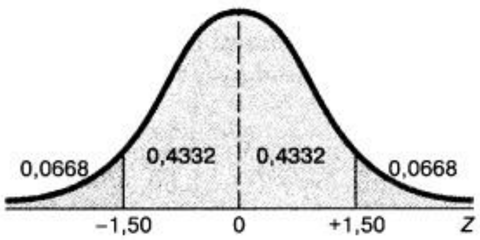

Рис. 4. Вычисление наблюдаемого уровня значимости двустороннего критерия

Связь между построением доверительных интервалов и проверкой гипотез

В этой заметке и [ранее](#) рассмотрены две основные задачи теории статистического вывода: построение доверительного интервала и проверка гипотез. Несмотря на то что они используют одни и те же понятия, эти задачи имеют разный смысл. Доверительные интервалы используются для оценки параметров генеральной совокупности, в то время как проверка гипотез касается конкретных значений этих параметров. В данной главе проверка гипотез применяется для принятия решений, касающихся конкретных значений параметров генеральной совокупности. Однако правильная интерпретация доверительного интервала также может свидетельствовать о том, что исследуемый параметр меньше, больше заданного значения или равен ему. Более того, доверительный интервал является интервальной оценкой исследуемого параметра генеральной совокупности. Наблюдаемый уровень значимости позволяет еще глубже понять преимущества проверки гипотез.

Например, мы проверили гипотезу о том, что средний вес коробок, заполненных на заводе компании Oxford Cereal Company, отличается от 368 г. Для этого мы использовали формулу (1). Вместо проверки этой гипотезы, можно было бы построить доверительный интервал, содержащий величину μ. Если гипотетическая величина μ = 368 попадает в доверительный интервал, нулевая гипотеза не отклоняется, поскольку величина 368 не является чем-то необычным. С другой стороны, если гипотетическая величина не попадает в доверительный интервал, нулевая гипотеза отклоняется, поскольку величина 368 теперь является экстремальной.

Односторонние критерии

До сих пор нулевая гипотеза предполагала равенство исследуемого параметра генеральной совокупности заданному значению. Например, мы проверили, равен ли средний вес коробок 368 г. Альтернативная гипотеза (H_1: $\mu \neq 368$) распадается на две: средний вес коробок может быть меньше или больше 368 г. По этой причине область отклонения гипотезы разделяется на две части, которые соответствуют двум хвостам распределения выборочных средних. Однако в некоторых ситуациях альтернативная гипотеза формулируется конкретнее и предполагает, что параметр генеральной совокупности строго больше заданного значения (или меньше). Рассмотрим один из таких примеров. Допустим, что компания, производящая сыр, желает проверить качество поставляемого молока. В частности, ее интересует, не подмешивает ли поставщик воду в молоко. Как известно, добавление воды снижает температуру замерзания молока, которая равна –0,545°С. Стандартное отклонение температуры замерзания молока равно 0,008°С. Поскольку компанию интересует, не стала ли температура замерзания молока ниже номинальной, область отклонения гипотезы теперь ограничена лишь левым хвостом распределения.[4]

В данном примере нулевая и альтернативная гипотезы формулируются так:

Шаг 1. H_0: $\mu \geq -0{,}545$°С. Шаг 2. H_1: $\mu < -0{,}545$°С.

Обратите внимание на то, что наша цель — доказать альтернативную гипотезу. Если нулевая гипотеза отклоняется, значит, у нас есть веские доказательства того, что температура замерзания молока меньше естественной. Если нулевая гипотеза не отклоняется, значит, у нас нет достаточных доказательств, чтобы утверждать, будто температура замерзания молока меньше естественной.

Шаг 3. $\alpha = 0{,}05$. Шаг 4. $n = 25$.

Шаг 5. Для проверки гипотезы применяется *Z*-критерий, поскольку стандартное отклонение температуры замерзания молока известно.

Шаг 6. Область отклонения гипотезы целиком ограничена левым хвостом распределения выборочных средних, поскольку мы хотим отклонить нулевую гипотезу H_0, только если выборочное среднее значительно меньше –0,545°С. Когда область отклонения гипотезы ограничена только одним хвостом распределения, критерий называется односторонним или направленным. Если альтернативная гипотеза содержит знак «меньше», критическое значение тестовой статистики *Z* должно быть отрицательным. Поскольку вся область отклонения гипотезы ограничена левым хвостом стандартизованного нормального распределения и его площадь превышает 0,05, то площадь, ограниченная кривой и критическим значением тестовой статистики должна быть равной 0,05 (рис. 5). Критическое значение тестовой статистики *Z* = НОРМ.СТ.ОБР(0,05) = –1,645.

Рис. 5. Односторонний критерий для проверки гипотезы о математическом ожидании генеральной совокупности при известном стандартном отклонении и уровне значимости, равном 0,05

Решающее правило выглядит следующим образом: Если $Z < -1{,}645$, гипотеза H_0 отклоняется; в противном случае она не отклоняется.

[4] На мой взгляд, в книге опечатка, так как добавление воды должно повышать температуру замерзания молока. Представьте, если бы молока вообще не было, то вода, как известно, замерзла бы при температуре 0°С.

Шаг 7. Допустим, что из партии молока выбраны 25 бидонов. Выборочная средняя температура замерзания равна –0,550°С. Используя формулу (1), получаем:

$$n = 25,\ \bar{X} = -0{,}550\ ^{\circ}\mathrm{C},\ \sigma = 0{,}008\ ^{\circ}\mathrm{C}$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{-0{,}550 - (-0{,}545)}{\frac{0{,}008}{\sqrt{25}}} = -3{,}125$$

Шаг 8. Поскольку $Z = -3{,}125 < -1{,}645$, гипотеза H_0 отклоняется, так как тестовая статистика лежит в области отклонения гипотезы (рис. 6).

Шаг 9. Таким образом, средняя температура замерзания молока оказалось меньше –0,545°С. Молоко сильно разбавлено, и компания должна провести тщательное расследование этого факта.

Использование t-критерия для проверки гипотезы о математическом ожидании при неизвестном стандартном отклонении

В большинстве ситуаций, касающихся числовых данных, стандартное отклонение σ неизвестно. Однако эту величину можно оценить, вычислив выборочное стандартное отклонение S. Если генеральная совокупность является нормально распределенной, выборочное среднее обладает t-распределением с $n - 1$ степенями свободы. Это дает возможность сформулировать t-критерий для оценки разности между выборочным средним $\bar{X}$ и математическим ожиданием генеральной совокупности μ:

$$(2)\quad t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

где тестовая статистика t имеет t-распределение с $n - 1$ степенями свободы.

Чтобы проиллюстрировать применение t-критерия, вернемся к задаче об аудиторской проверке в компании Saxon Home Improvement Company. Для проверки аудитор извлекает из информационной системы выборку накладных, заполненных в течение последнего месяца. Средняя сумма накладных за последние пять лет равна 120 долл. Аудитор должен оценить, изменилась ли сумма накладных. Иначе говоря, в ходе проверки гипотезы требуется доказать, что средняя сумма накладных увеличивается или уменьшается.

Для проверки гипотезы с помощью двустороннего критерия применяется описанный выше алгоритм.

Шаги 1 и 2. Формулируем нулевую и альтернативную гипотезы: H_0: μ = 120 долл. H_1: μ ≠ 120 долл.

Обратите внимание на то, что предметом доказательства является альтернативная гипотеза. Если нулевая гипотеза отклоняется, значит, у нас есть веские доказательства того, что средняя сумма накладных отличается от 120 долл. Если нулевая гипотеза не отклоняется, значит, у нас нет достаточных доказательств, чтобы утверждать, будто средняя сумма накладных отличается от 120 долл.

Шаг 3. Полагаем α = 0,05.

Шаг 4. Из генеральной совокупности накладных извлекается случайная выборка, состоящая из n = 12 накладных.

Шаг 5. Поскольку предполагается, что генеральная совокупность является нормально распределенной и стандартное отклонение генеральной совокупности не известно, применяется t-критерий.

Шаг 6. При фиксированном объеме выборки n тестовая статистика t имеет t-распределение с $n - 1$ степенями свободы. Поскольку альтернативная гипотеза H_1: μ ≠ 120 долл. является ненаправленной, область отклонения гипотезы в t-критерии разделяется на две части, ограниченные левым и правым хвостами t-распределения. Площадь каждой из областей равна 0,025. Если заданный уровень значимости одностороннего α равен 0,025, а количество степеней свободы = 12 – 1 = 11, критические значения t-распределения составляют: левое = СТЬЮДЕНТ.ОБР(0,025;11) = –2,201, правое = СТЬЮДЕНТ.ОБР(0,975;11) = 2,201 (рис. 6).

Рис. 6. Проверка гипотезы о математическом ожидании при неизвестном стандартном отклонении, уровне значимости, равном 0,05 (двустороннем), и 11 степенях свободы

Решающее правило таково: если $t < -t_{11} = -2{,}201$ или $t > t_{11} = +2{,}201$, нулевая гипотеза отклоняется, в противном случае она не отклоняется.

Шаг 7. Данные о суммах (в долларах) из выборки, состоящей из 12 накладных: 108,98; 152,22; 111,45; 110,59; 127,46; 107,26; 93,32; 91,97; 111,56; 75,71; 128,58; 135,11. Результаты вычислений, связанных с аудиторской проверкой компании Saxon Home Improvement Company:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = 112{,}85 \text{ долл.}, \quad S = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} = 20{,}80 \text{ долл.}$$

В соответствии с формулой (2) *t*-статистика равна:

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{112{,}85 - 120}{\frac{20{,}80}{\sqrt{12}}} = -1{,}19$$

Шаг 8. Поскольку $-2{,}201 < t = -1{,}19 < 2{,}201$, тестовая статистика попадает в область принятия гипотезы (см. рис. 6).

Шаг 9. Гипотеза H_0 не отклоняется.

Шаг 10. Аудитор не имеет оснований утверждать, что средняя сумма накладных за последний месяц значительно отличается от 120 долл.

Одновыборочный *t*-критерий применяется в тех ситуациях, когда стандартное отклонение генеральной совокупности σ неизвестно и оценивается с помощью выборочного стандартного отклонения *S*. Этот критерий является классической параметрической процедурой. Он сопровождается большим количеством строгих ограничений, гарантирующих корректность полученных результатов. Предполагается, что выборка принадлежит нормально распределенной генеральной совокупности. На практике, если выборка достаточно велика, а распределение не слишком асимметрично, *t*-распределение хорошо аппроксимирует выборочное распределение средних.

Условия, налагаемые на применение одновыборочного *t*-критерия, можно проверить с помощью Excel. Как было показано ранее, предположение о нормальности распределения можно проверить несколькими способами. Чтобы убедиться, насколько точно выборочные данные соответствуют теоретическим свойствам нормального распределения, можно воспользоваться методами описательной статистики, а также средствами графического анализа (гистограммой, диаграммой «ствол и листья», блочной диаграммой и графиком нормального распределения). На рис. 7–9 приведены описательные статистики, график нормального распределения и блочная диаграмма, полученные с помощью Excel на основе данных о накладных.

Суммы накладных, долл.		Суммы накладных, долл.	
108,98		Среднее	112,85
152,22		Стандартная ошибка	6,00
111,45		Медиана	111,02
110,59		Мода	#Н/Д
127,46		Стандартное отклонение	20,80
107,26		Дисперсия выборки	432,56
93,32		Эксцесс	0,17
91,97		Асимметричность	0,13
111,56		Интервал	76,51
75,71		Минимум	75,71
128,58		Максимум	152,22
135,11		Сумма	1354,21
		Счет	12,00
		Уровень надежности(95,0%)	13,21

Рис. 7. Описательные статистики на основе данных о накладных, полученные в Excel с использованием Пакета анализа

i	Q_i	Суммы накладных, долл.
1	-1,43	75,71
2	-1,02	91,97
3	-0,74	93,32
4	-0,50	107,26
5	-0,29	108,98
6	-0,10	110,59
7	0,10	111,45
8	0,29	111,56
9	0,50	127,46
10	0,74	128,58
11	1,02	135,11
12	1,43	152,22

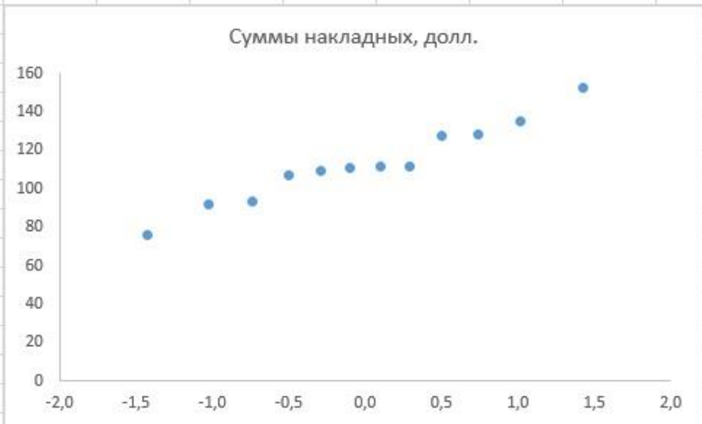

Рис. 8. График нормального распределения на основе данных о накладных

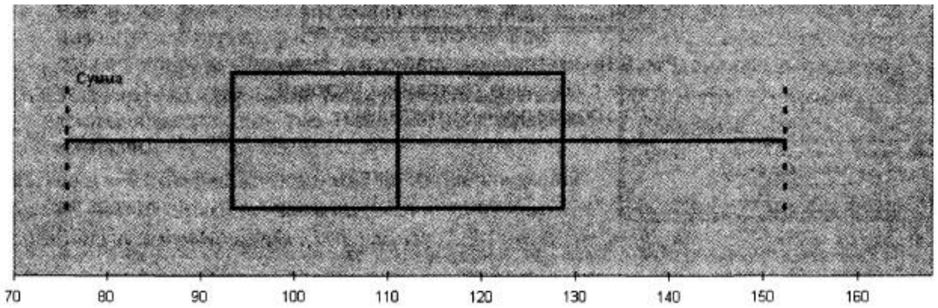

Рис. 9. Блочная диаграмма на основе данных о накладных

Поскольку среднее значение очень близко к медиане, а точки на графике нормального распределения возрастают, колеблясь вокруг прямой линии, нет оснований говорить, что генеральная совокупность сумм накладных не является нормально распределенной. Таким образом, аудиторскую проверку следует признать корректной.

Одновыборочный *t*-критерий является устойчивым. Его мощность не снижается, если кривая распределения, из которой извлечена выборка, отличается от нормальной, особенно если объем выборки достаточно велик (в этом случае для *t*-статистики справедлива центральная предельная теорема). Однако некорректное применение *t*-статистики может привести к ошибочным выводам. Если объем выборки *n* невелик (меньше 30) и распределение генеральной совокупности не является даже приближенно нормальным, следует применять непараметрические процедуры проверки гипотез.

Применение *Z*-критерия для проверки гипотезы о доле признака в генеральной совокупности

В некоторых ситуациях необходимо оценить долю признака p в генеральной совокупности, а не математическое ожидание. Из генеральной совокупности можно извлечь случайную выборку, вычислить *долю признака в выборке* $p_S = X/n$ и сравнить ее с гипотетическим значением параметра p.

Если количество успехов X и количество неудач $n - X$ больше пяти, выборочное распределение доли хорошо аппроксимируется стандартизованным нормальным распределением. Чтобы оценить разность между фактической долей признака p_S и гипотетическим параметром p, применяется критерий для проверки гипотезы о доле признака:

$$(3)\quad Z = \frac{p_S - p}{\sqrt{\frac{p(1-p)}{n}}}$$

где $p_S = X / n$ = количество успехов / размер выборки = наблюдаемая доля успехов , p — гипотетическая доля успехов в генеральной совокупности. *Z*-статистика аппроксимируется стандартизованным нормальным распределением.

Если числитель и знаменатель формулы (3) умножить на n, получим:

$$(4)\quad Z = \frac{X - np}{\sqrt{np(1-p)}}$$

В качестве иллюстрации рассмотрим результаты опроса, опубликованные недавно в журнале Wall Street Journal. В ходе исследования оценивалось количество мужчин и женщин, владеющих надомным бизнесом. В опросе приняли участие 899 респондентов, в том числе 369 женщин. Нулевая и альтернативная гипотезы формулируются следующим образом: H_0: p = 0,50 (доля женщин, владеющих надомным бизнесом, равна 0,5), H_1. p ≠ 0,50 (доля женщин, владеющих надомным бизнесом, не равна 0,5).

Применение критического значения. Поскольку нас интересует, равна ли доля женщин, владеющих надомным бизнесом, числу 0,5, следует применить двусторонний критерий. Если уровень значимости равен 0,05, области отклонения принятия гипотезы определяются так, как показано на рис. 10. Решающее правило формулируется следующим образом: если Z > +1,96 или Z < –1,96, гипотеза H_0 отклоняется; в противном случае она принимается.

Рис. 10. Двусторонний критерий для проверки гипотезы о доле признака с уровнем значимости, равным 0,05

Поскольку среди 899 владельцев надомного бизнеса оказалось 369 женщин, имеем:

$$p_s = \frac{369}{899} = 0,41046$$

Используя формулу (3), получаем:

$$Z = \frac{p_s - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0,41046 - 0,50}{\sqrt{\frac{0,50(1-0,50)}{899}}} = \frac{-0,08954}{0,0167} = -5,37$$

Поскольку –5,37 < –1,96, нулевую гипотезу следует отклонить. Таким образом, есть основания утверждать, что доля женщин среди владельцев надомного бизнеса не равна 0,50.

Потенциальные проблемы и этические вопросы, связанные с проверкой гипотез

Итак, мы рассмотрели основные принципы проверки гипотез. Она используется для анализа разности между выборочными оценками (т.е. статистиками) и характеристиками генеральной совокупности (т.е. параметрами). Кроме того, мы можем оценивать вероятность ошибок 1-го и 2-го рода. На некоторые вопросы необходимо ответить еще на этапе планирования. Для этого следует обратиться к профессиональному статистику. Однако часто это происходит слишком поздно, когда данные уже собраны. Единственное, что может сделать профессионал в этой ситуации, — посоветовать подходящую процедуру обработки данных. Остается лишь надеяться, что плохое планирование опроса не приведет к существенным искажениям результатов, хотя это является большой натяжкой. Корректное исследование невозможно без хорошего планирования. Чтобы избежать искажения результатов опроса, необходимо с самого начала придерживаться правильной стратегии.

Вопросы, на которые необходимо ответить при планировании проверки гипотез:

- Какова цель опроса, исследования или эксперимента? Можно ли сформулировать ее в виде нулевой и альтернативной гипотез?
- Какой вид критерия следует выбрать: двусторонний или односторонний?
- Можно ли извлечь случайную выборку из интересующей нас генеральной совокупности?
- Какие измерения можно выполнить на основе выборки? Какие показатели получаются в результате измерений: числовые или категорийные?
- Какой уровень значимости следует выбрать?
- Достаточно ли велик объем выборки, чтобы достичь желаемой мощности критерия при заданном уровне значимости?
- Какую процедуру выбора следует применить при формировании выборки и почему?
- Какие выводы можно сделать на основе проверки гипотез и как интерпретировать результаты?

Следует различать некорректную методологию и неэтичное поведение, которое, как правило, выражается в нечестном манипулировании процедурой проверки гипотез. Этические проблемы возникают на всех этапах проверки гипотез: при сборе данных, при проведении опроса, при выборе критерия и уровня значимости, при подтасовке, очистке и отбрасывании данных, а также при документировании результатов.

Метод сбора данных — рандомизация. Чтобы избежать возможного искажения результатов, необходимо применять правильный метод сбора данных. Для получения осмысленных результатов выборку следует формировать случайным образом, а в ходе эксперимента использовать процедуру рандомизации. Респондентов нельзя отбирать целенаправленно или по их собственному желанию. Пренебрежение рандомизацией может привести к серьезному искажению результатов и обесценить весь опрос.

Добросовестность респондентов. Этические нормы требуют, чтобы респонденты были проинформированы о цели исследования и потенциальных последствиях опроса. Кроме того, респондент должен быть добросовестным и честно отвечать на вопросы.

Вид критерия — двусторонний или односторонний. Как правило, односторонние критерии обладают более высокой мощностью, чем двусторонние. С другой стороны, если исследователей интересует лишь величина отклонения от нулевой гипотезы, а не ее знак, более приемлемым является двусторонний критерий. Например, если предыдущие исследования или теоретические рассуждения показали, что отклонение от нулевой гипотезы имеет определенный знак, можно применять односторонний критерий.

Выбор уровня значимости. В хорошо продуманном исследовании уровень значимости устанавливают до сбора данных. Его нельзя изменять задним числом, стремясь достичь желаемого результата. Это привело бы к подтасовке данных. Таким образом, формулируя выводы, полученные в результате проверки гипотезы, всегда необходимо указывать *p*-значение.

Подтасовка данных. Никогда не следует подтасовывать данные. Ни в коем случае нельзя сначала проверить гипотезу, проанализировать результат, а затем выбрать вид критерия и/или уровень значимости. Эти этапы выполняют заранее, еще до сбора данных. Иначе выводы исследования потеряют смысл. Нулевая и альтернативная гипотезы, а также уровень значимости должны быть установлены с самого начала.

Очистка и отбрасывание данных. Очистка данных не является подтасовкой. При редактировании, кодировании и переписывании ответов на вопросы исследователи могут обнаружить экстремальные или необычные данные. Если проверка гипотез касается числовых данных, необходимо построить диаграмму «ствол и листья» и блочную диаграмму. Это позволит очистить данные и отбросить экстремальные значения, которые лишь искажают истинную картину. Кроме того, предварительный анализ полученных данных должен сопровождаться проверкой предположения о распределении генеральной совокупности. Процесс очистки данных поднимает важные этические вопросы. Следует ли вообще исключать некие данные из исследования? Разумеется, да, если выяснится, что измерения проводились некорректно. Иногда у исследователей нет выбора — например, респондент может отказаться от дальнейшего участия в опросе, не закончив отвечать на вопросы. В хорошо продуманном исследовании статистик должен заранее сформулировать правила отбрасывания данных.

Документирование результатов. В любом исследовании необходимо честно документировать как положительные, так и отрицательные результаты, чтобы последователи не повторяли ваших ошибок. Нельзя публиковать лишь результаты, обладающие статистической значимостью, игнорируя результаты, не имеющие достаточных подтверждений. Если исследователь не имеет достаточных оснований отклонить нулевую гипотезу, он должен понимать, что это не является доказательством ее истинности. Это лишь означает, что у вас нет достаточной информации, чтобы опровергнуть нулевую гипотезу при данном объеме выборки.

Статистическая значимость и практическая ценность. Исследователь, принимающий решение на основе проверки гипотезы, должен ясно различать статистическую значимость и практическую ценность результата в контексте конкретной прикладной области. Иногда очень большой объем выборки, необходимый для достижения статистически значимого результата, сводит на нет его практическую ценность. Например, предположим, что перед проведением общенациональной рекламной кампании по телевидению считалось, что доля людей, признающих некую торговую марку, равна 0,30. После завершения кампании выяснилось, что рекламируемую торговую марку признают 6168 респондентов из 20 000. Односторонний критерий показывает, что эта доля теперь превышает 0,30 с уровнем значимости p, равным 0,0047, и можно вполне обоснованно утверждать, что доля потребителей, признающих данную торговую марку, увеличилась. Значит ли это, что рекламная кампания была успешной? Результаты проверки гипотез свидетельствуют, что признание торговой марки выросло на статистически значимую величину, но можно ли считать это увеличение значимым с практической точки зрения? Обратите внимание на то, что доля потребителей, признающих указанную торговую марку, равна 6168/20 000 = 0,3084, или 30,84%. Это величина превышает гипотетическое значение, равное 30%, меньше чем на 1%. Привели ли крупные затраты на телевизионную рекламу к существенному росту популярности торговой марки? Учитывая большой объем затраченных средств и незначительный прирост популярности, рекламную кампанию следует признать неудачной. С другой стороны, если доля поклонников рекламируемой торговой марки увеличилась бы на 20%, рекламную кампанию пришлось бы назвать успешной.

Подводя итоги, следует отметить, что главным аспектом в обсуждении проверки гипотез является намерение. Необходимо различать неправильный анализ данных и нечестное поведение, которое проявляется в целенаправленном формировании выборки, манипулировании респондентами, подтасовке данных, выборе подходящего критерия (дву- или одностороннего), подгонке уровня значимости, игнорировании фактов, противоречащих желательной гипотезе, и замалчивании неудобных данных.

Мощность критерия

Обсуждая критерии для проверки гипотез, мы определили два вида риска, возникающего при принятии решений о параметрах генеральной совокупности на основе выборочного анализа: величину α, представляющую собой вероятность того, что будет отклонена истинная нулевая гипотеза, которая не должна быть отвергнута, а величина β, являющаяся вероятностью того, что не будет отвергнута ложная нулевая гипотеза, которую следовало отклонить. Мощность критерия, равная величине $1 - \beta$, характеризует чувствительность статистического критерия — вероятность отклонить ложную нулевую гипотезу, которая должна быть отвергнута. Мощность статистического критерия зависит от того, насколько значительно истинное математическое ожидание генеральной совокупности отличается от гипотетического (принятого в гипотезе H_0), уровня значимости α и объема выборки *n*. Если истинное и гипотетическое математические ожидания существенно отличаются друг от друга, мощность критерия будет выше, а если разность между истинным и гипотетическим математическим ожиданием мала, мощность критерия снижается. Чем выше уровень значимости α, тем легче отвергнуть гипотезу H_0, и, следовательно, выше мощность критерия. Чем больше объем выборки, тем точнее оценки и легче обнаружить разность между истинными и гипотетическими параметрами. Это также увеличивает мощность критерия.

В данном разделе понятие мощности статистического критерия иллюстрируется задачей о расфасовке кукурузных хлопьев. Предположим, что процесс расфасовки подвергается периодическим проверкам, проводимым местным обществом по защите прав потребителей. Цель этих проверок — обнаружить недовес, т.е. выявить коробки, содержащие меньше 368 г кукурузных хлопьев. Таким образом, проверяющие стремятся найти свидетельства, что средний вес коробок меньше 368 г. В этой ситуации нулевая и альтернативная гипотезы формулируются следующим образом: H_0: $\mu \geq 368$ (расфасовка выполняется правильно), H_1: $\mu < 368$ (расфасовка выполняется неправильно).

Проверяющий должен учесть, что стандартное отклонение веса коробки *σ* равно 15 г. Следовательно, можно применить *Z*-критерий. Если уровень значимости α равен 0,05, а случайная выборка состоит из 25 коробок, то значение $\overline{X}$, позволяющее отклонить нулевую гипотезу, можно определить по формуле (1), в которой вместо $\overline{X}$ подставляется величина $\overline{X}_L$:

$$Z = \frac{\overline{X}_L - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \overline{X}_L = \mu + Z\frac{\sigma}{\sqrt{n}}$$

Поскольку этот критерий является односторонним, а его уровень значимости равен 0,05, то Z =НОРМ.СТ.ОБР(0,05) = –1,645 (рис. 11). Следовательно,

$$\overline{X}_L = 368 + (-1{,}645)\frac{15}{\sqrt{25}} = 368 - 4{,}935 = 363{,}065$$

Решающее правило одностороннего критерия таково: гипотеза H_0 отклоняется, если X < 363,065, в противном случае гипотеза H_0 не отклоняется.

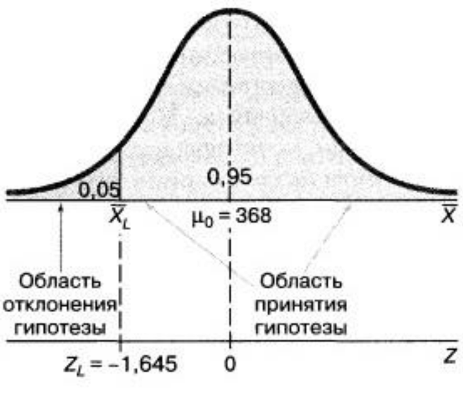

Рис. 11. Определение нижнего критического значения одностороннего *Z*-критерия для проверки гипотезы о математическом ожидании генеральной совокупности при уровне значимости, равном 0,05

Решающее правило устанавливает, что, если выборочное среднее, вычисленное для случайной выборки, состоящей из 25 коробок, меньше 363,065 г, нулевая гипотеза отклоняется, и проверяющий приходит к выводу, что процесс расфасовки осуществляется неправильно. Мощность критерия измеряет вероятность прийти к выводу, что процесс выполняется неверно, на основе анализа величин, отличающихся от истинного математического ожидания генеральной совокупности.

Предположим, требуется определить вероятность отклонить нулевую гипотезу при условии, что истинный средний вес в генеральной совокупности коробок равен 360 г. На основе решающего правила необходимо определить площадь фигуры, лежащей под нормальной кривой слева от точки 363,065. На основе центральной предельной теоремы и предположения о нормальности распределения веса в генеральной совокупности коробок можно допустить, что выборочное распределение является нормальным. Следовательно, площадь фигуры, лежащей под нормальной кривой слева от точки 363,065, можно выразить в единицах стандартного отклонения, так как мы вычисляем вероятности отклонить нулевую гипотезу при условии, что истинный средний вес в генеральной совокупности коробок равен 360,00 г. Используя формулу (1), получаем:

$$Z = \frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}}$$

где μ_1 — истинное математическое ожидание генеральной совокупности. Следовательно,

$$Z = \frac{363,065 - 360}{\frac{15}{\sqrt{25}}} = 1,02$$

$P(Z < +1,02)$ =НОРМ.СТ.РАСП(1,02;ИСТИНА) = 0,8461. Это и есть мощность критерия, равная площади фигуры, лежащей под нормальной кривой слева от точки 363,065 (рис. 12). Вероятность β того, что нулевая гипотеза (μ = 368) будет отклонена, равна 1 – 0,8461 = 0,1539. Следовательно, вероятность ошибки 2-го рода равна 0,1539.

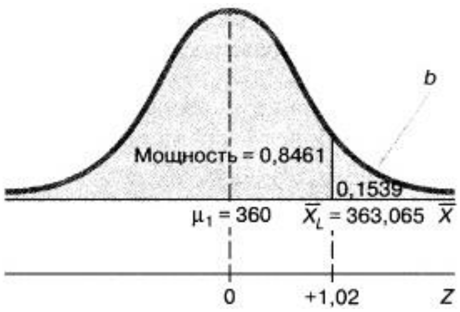

Рис. 12. Определение мощности критерия и вероятности ошибки 2-го рода при условии, что μ_1 = 360 г

Определив мощность критерия при условии, что истинное математическое ожидание генеральной совокупности равно 360 г, мы можем вычислить мощность критерия при любом другом значении μ. Например, какова мощность критерия, если истинное математическое ожидание генеральной совокупности равно 352 г? Предположим, что стандартное отклонение, объем выборки и уровень значимости остаются неизменными. Тогда решающее правило принимает следующий вид: гипотеза H_0 отклоняется, если $\bar{X} < 363,065$, в противном случае гипотеза H_0 не отклоняется.

Поскольку мы проверяем гипотезу о математическом ожидании, снова применим формулу (1):

$$Z = \frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{363,065 - 352}{\frac{15}{\sqrt{25}}} = 3,69$$

(для истинного среднего веса генеральной совокупности коробок μ_1 = 352 г) (рис. 13).

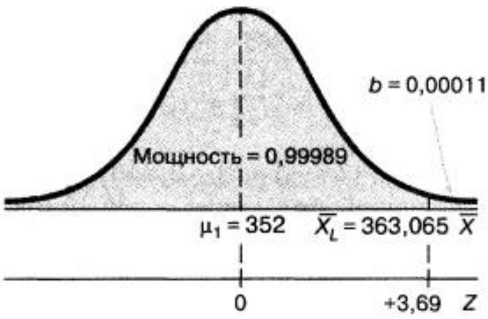

Рис. 13. Определение мощности критерия и вероятности ошибки 2-го рода при условии, что μ_1 = 352 г

В двух рассмотренных выше примерах мощность критерия была довольно высокой, а вероятность ошибки 2-го рода, наоборот, небольшой. В следующем примере мощность критерия вычисляется при условии, что средний вес генеральной совокупности коробок равен 367 г. Это значение очень близко к гипотетическому математическому ожиданию, равному 368 г. Вновь применим формулу (1):

$$Z = \frac{\overline{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{363{,}065 - 367}{\frac{15}{\sqrt{25}}} = -1{,}31$$

(для истинного среднего веса генеральной совокупности коробок μ_1 = 367 г). P(Z < –1,31) = =НОРМ.СТ.РАСП(-1,31;ИСТИНА) = 0,0951. Поскольку в данном примере область отклонения гипотезы лежит в левой части распределения, мощность критерия равна 0,0951, а вероятность совершить ошибку 2-го рода — 0,9049 или 90,5%.

На рис. 14 показана мощность критерия для разных значений μ_1 (включая три рассмотренные выше примера). Этот график называется *кривой мощности*.

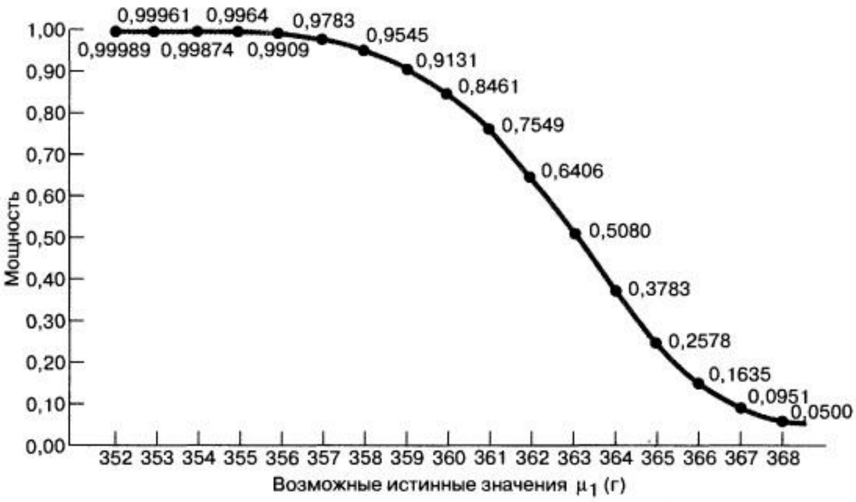

Рис. 14. Кривая мощности критерия для проверки гипотезы о среднем весе коробки кукурузных хлопьев при альтернативной гипотезе H_1: $\mu_1 < 368$ г

Анализ рис. 14 показывает, что мощность рассмотренного одностороннего критерия резко возрастает (и стремится к 100%) по мере того, как истинное математическое ожидание генеральной совокупности удаляется от гипотетического математического ожидания, равного 368 г. Очевидно, что для данного одностороннего критерия, чем меньше истинное значение μ_1 по сравнению с гипотетическим, тем больше вероятность обнаружить это отличие.[5] С другой стороны, если истинное значение μ_1 близко к 368 г, мощность критерия снижается, поскольку он не может эффективно распознавать маленькие отличия между истинным и гипотетическим математическими ожиданиями.

Вычисления, проведенные для всех трех примеров, суммированы на рис. 15.

[5] В ситуациях, относящихся к односторонним критериям, когда истинное среднее μ_1 превышает гипотетическое среднее, наблюдается противоположная зависимость. Чем больше истинное среднее μ_1 по сравнению с гипотетическим средним, тем выше мощность. С другой стороны, для двусторонних критериев, чем больше расстояние между истинным и гипотетическим средними, тем выше мощность критерия.

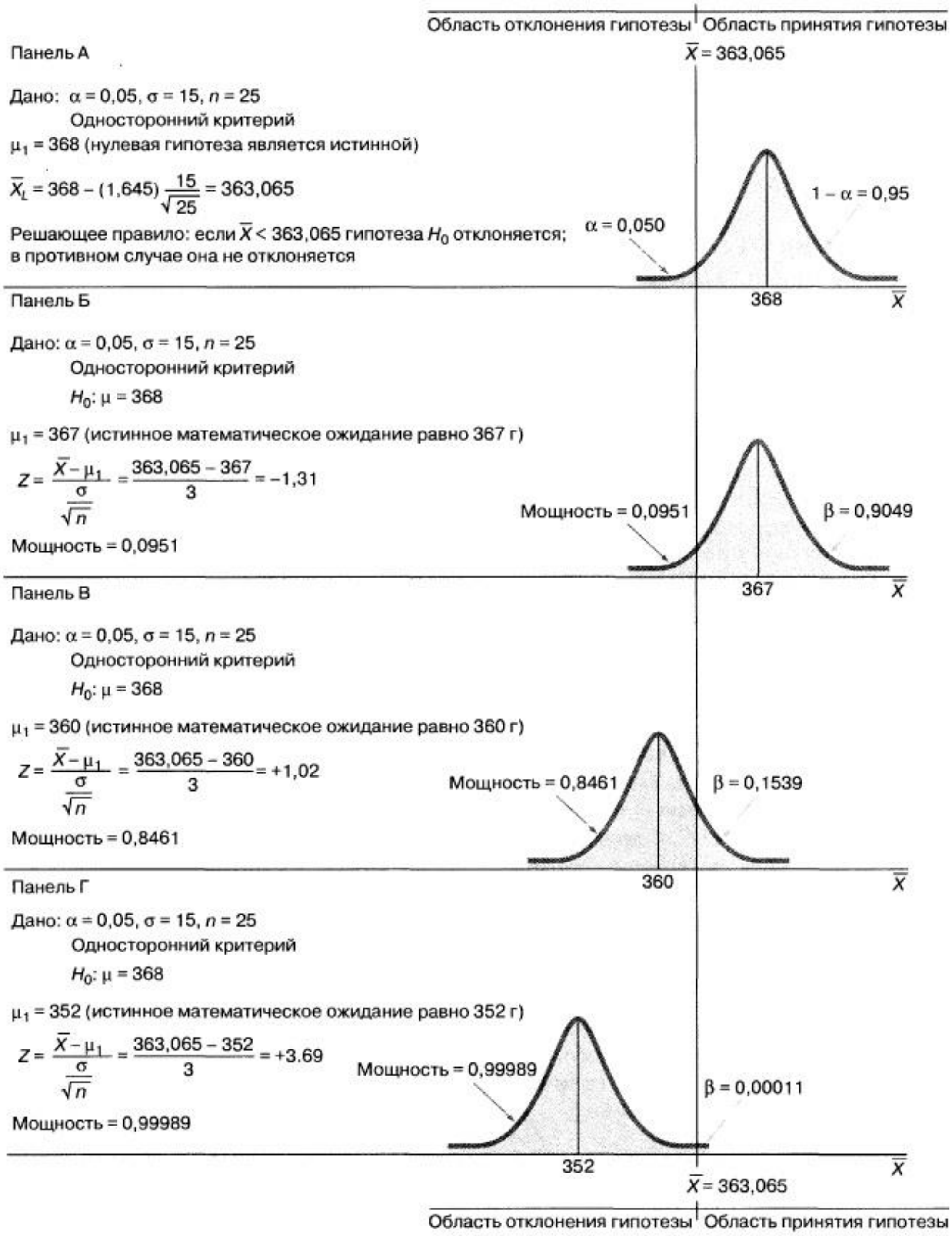

Рис. 15. Вычисление мощности статистических критериев при разных значениях истинного математического ожидания генеральной совокупности

Анализируя рис. 15, можно обнаружить резкие различия между мощностями критерия, соответствующими разным значениям истинного математического ожидания. Как оказано на панелях А и Б, если истинное математическое ожидание генеральной совокупности незначительно отличается от 368 г, вероятность отклонить нулевую гипотезу невелика. Однако, как только истинное математическое ожидание значительно отклоняется от гипотетического, мощность критерия резко возрастает и стремится к максимуму, равному единице (или 100%).

При вычислении мощности одностороннего статистического критерия мы полагали, что уровень значимости равен 0,05, а выборка состоит из 25 коробок. Учитывая это, можно определить эффект, который оказывают на мощность критерия перечисленные ниже параметры, варьируя их по одному.

- Тип статистического критерия — односторонний или двусторонний.
- Уровень значимости α.
- Объем выборки n.

Сформулируем три основных вывода, касающихся мощности критерия.

- Односторонний критерий имеет более высокую мощность, чем двусторонний. Его следует применять, когда требуется определить направление альтернативной гипотезы.
- Поскольку вероятность ошибки 1-го рода (α) и вероятность ошибки 2-го рода (β) противоположны, причем ошибка 2-го рода является параметром, дополнительным к мощности критерия 1 – β, мощность критерия прямо зависит от параметра α. Повышение уровня значимости α увеличивает мощность критерия, а снижение уровня значимости уменьшает ее.
- С увеличением объема выборки *n* мощность критерия повышается, а с уменьшением — снижается.

Итак, как показано на структурной схеме (рис. 16), мы рассмотрели основные принципы проверки гипотез. В частности, изложены *Z*- и *t*-критерии проверки гипотез о математическом ожидании генеральной совокупности, а также *Z*-критерии для проверки гипотез о доле признака в генеральной совокупности. Кроме того, был рассмотрен ряд важных практических примеров. Следующие заметки будут посвящены развитию начатой темы.

Рис. 16. Структурная схема заметки

Предыдущая заметка [Определение объема выборки]

Следующая заметка

К оглавлению [Статистика для менеджеров с использованием Microsoft Excel]