1 Конспект по проверке гипотез

1.1 Задача проверки гипотез

Рассмотрим, такую задачу: провели слепую дегустацию двух сортов чая, каждый респондент выбрал из двух неподписанных чашек чая более вкусный. Необходимо сказать, какой из двух чаев лучше. Понятно, что просто сказать, что лучше тот чай, который выбрало больше народу, нельзя — потому что наблюдаемый эффект может оказаться чистой случайностью (при этом наиболее предпочитаемым случайно может оказаться более плохой чай).

Пусть всего было N респондентов и результаты представлены в виде выборки $X_1, \ldots, X_N$, где каждый элемент имеет значение 0 или 1 в зависимости от того, какой сорт чая выбрал респондент.

На вероятностном языке, задача состоит в том, чтобы по конечной выборке из неизвестного (или, скажем так, не полностью известного) распределения проверить некоторое утверждение касательно этого распределения. Задача является статистической в том смысле, что случайным наблюдениям необходимо сделать вывод о детерминированных свойствах. Аналогичная ситуация была при оценивании параметров, но там постановка задачи несколько иная, хотя между этими задачами есть тесная связь.

Понятно, что (за исключением вырожденных случаев) мы можем дать только вероятностный ответ на поставленный вопрос, так же как при оценивании параметров мы могли только указать доверительный интервал для неизвестного параметра, но не могли ничего сказать точно.

Нам нужно получить вычислительную схему, которая по переданным данным может дать в каком-то смысле ответ на поставленный вопрос.

1.2 Статистические критерии

Рассмотрим общую схему построения одновыборочного критерия[1]. Пусть имеется выборка из какого-то распределения (предполагается, что она получена в результате серии экспериментов). Предполагается, что наблюдения независимы как случайные величины и одинаково распределены.

Формулируется основная гипотеза (еще говорят нулевая гипотеза, H_0), а также альтернативная гипотеза H_1. В примере выше в качестве нулевой гипотезы можно рассмотреть "чаи одинаково привлекательны", а в качестве альтернативы — "чаи отличаются по привлекательности" или "чай «Twinings» более привлекателен, чем «Чай со слоном»". Гипотезу называют *простой*, если она соответствует только одному распределению в рамках рассматриваемой модели, в противном случае гипотезу называют *сложной*. Как правило, нулевая гипотеза является простой, а альтернатива — сложной, объединением целого класса простых гипотез. В рассматриваемом примере нулевой гипотезе соответствует только одно распределение — симметричное распределение Бернулли, в то время как альтернативным гипотезам соответствуют целые классы распределений (в первом случае — любое несимметричное распределение Бернулли, во втором — распределение Бернулли с $p > 0.5$).

Будем строить некоторую схему ("критерий"), которая по данной выборке будет отвечать на вопрос, отвергается ли нулевая гипотеза в пользу альтернативы или нет. Критерий может ошибаться. Более того, так как сами данные случайны, то (кроме

[1] По-английски критерии называют tests, в русском так тоже говорят

некоторых вырожденных случаев) даже хороший критерий будет ошибаться с ненулевой вероятностью. Будем называть *ошибкой первого рода* ошибочное отклонение верной нулевой гипотезы. В нашем случае, эта ошибка имеет следующий смысл: мы приняли случайное различие за систематическое. Ошибкой *второго рода* будем называть не отклонение неверной нулевой гипотезы при верной альтернативе. В примере с чаями ошибкой второго рода будет не увидеть имеющего места различия из-за чисто случайных эффектов (или принять наблюдаемое различие за случайный эффект). Качество критерия определяется вероятностями двух этих ошибок. Эти вероятности обозначают α_I и α_{II} соответственно. Величину $1-\alpha_{II}$ называют *мощностью*.

При построении критерия мы фиксируем величину α, называемую *уровнем значимости*, и будем строить критерий так, чтобы $\alpha_I = \alpha$ (тогда критерий будет называться точным) или хотя бы $\alpha_I \to \alpha$ с ростом объема выборки (тогда критерий будет называться асимптотическим).

Мы можем рассмотреть пространство всех возможных выборок $\Omega \sim \mathsf{R}^N$. Далее на этом пространстве мы выделим некоторое подмножество S (называемое *критическим множеством* или *критической областью*), такое, что вероятность случайной выборки попасть в него, при условии, что верна нулевая гипотеза, равна (или асимптотически равна) α. Дополнение $\Omega \backslash S$ будем называть *доверительным* множеством.

Теперь определим критерий следующим образом: проверка гипотезы H_0 сводится к проверке $X \in S$. Если выборка попала в критическую область, то отвергаем гипотезу. Очевидно, что вероятность ошибки первого рода, т.е. вероятность ошибочно отвергнуть верную гипотезу равна в точности α. Т.е. построенный нами критерий является корректным, в том смысле, что его уровень значимости соответствует желаемому (или асимптотически по N соответствует).

Однако, корректность критерия не является единственным требованием. Очевидно, что мы можем построить критическое множество S “корректно” очень многими способами. Какой же из них лучше? Лучшим будет тот способ, который гарантирует нам наибольшую мощность, т.е. наименьшую вероятность ошибки второго рода (наименьшую вероятность принять неверную гипотезу при верной альтернативе).

Очевидно, что значение мощности не определяется нами, а зависит от альтернативы. Критерий, мощный против одной альтернативы, может быть немощен против какой-то другой; а критерий, мощный против объединения нескольких альтернатив сразу, скорее всего, против каждой конкретной из них будет менее мощным, чем критерий, построенный только против нее (проблема разводного ключа).

Более того, мощность зависит от **конкретной** простой альтернативы, в то время как альтернатива может (как в нашем примере) быть сформулирована достаточно размыто, в виде целого класса простых альтернатив. В самом деле, привлекательность чаев может отличаться, но насколько сильно? Чем больше отличие, тем легче его обнаружить (т.е. тем больше мощность).

Несмотря на то, что мощность зависит от конкретной альтернативы, часто мы можем построить критерий, мощность которого будет стремиться к 1 при росте объема выборки против любой простой альтернативы из выбранного класса альтернатив H_1. Такие критерии называют *критерии асимптотической мощности* 1.

Замечание 1. Часто можно слышать фразу “гипотезу нельзя принять, можно только отвергнуть”. Это утверждение следует понимать так. Если мы проводим тест и попали в критическую область, то мы уверены, что с вероятностью α нулевая гипотеза неверна (следовательно, верна альтернатива). И мы можем сказать “гипотеза отвергнута в пользу такой-то альтернативы”, и мы будем знать вероятность ошибиться ($\alpha_I = \alpha$) и, более того, мы можем сами ее задавать.

Если же мы попали в доверительную область, то это может означать как то, что нулевая гипотеза верна, так и то, что верна альтернатива, но нам не хватило мощности, чтобы это распознать.

Принять гипотезу, если она не отвергалась, мы можем, если знаем мощность критерия против всех возможных альтернатив и она нас устраивает. Однако в общем случае мощность нам неизвестна и повлиять мы на нее не можем, поэтому грамотно в этом случае говорить не "принимается нулевая гипотеза", а "нет оснований отвергнуть нулевую гипотезу".

1.3 Язык статистик критерия

Понятно, что строить критерий по вышеописанной схеме на практике неудобно по многим причинам (выборка, как правило, состоит более, чем из одного элемента, следовательно, пространство Ω будет многомерным). Хотелось бы иметь возможность строить одномерные доверительные области для произвольного значения N и и свести проверку гипотезы к проверке неравенства.

Предлагается следующая идея. Рассмотрим некоторую статистику (т.е. функцию от выборки) $t = t(X)$, распределение которой при условии, что верна нулевая гипотеза H_0, известно (или известно асимптотически). Это известное распределение иногда называют *идеальное распределение статистики*, т.е. распределение, когда все хорошо и исходное предположение верно. Кроме этого распределения также известно должно быть распределение или **поведение** статистики критерия в том случае, когда верна альтернативная гипотеза H_1 (напомню, что часто альтернативной гипотезе соответствует целый класс распределений, поэтому указать распределение статистики для альтернативной гипотезы не всегда возможно).

Теперь все множество возможных значений статистики (уже одномерное) разбивается на доверительную и критическую области. Разбиение осуществляется с помощью квантилей известного идеального распределения, потому как нужно получить заявленный уровень значимости (т.е. ошибку первого рода), а положение квантилей выбирается исходя из известного поведения статистики при верной альтернативе.

Конкретно, критическая область должна находиться там, где появление статистики критерия при верной альтернативе наиболее вероятно. При таком выборе мощность против данной альтернативы будет максимальной.

1.4 Классические одновыборочные критерии

Мы начнем наше знакомство с конкретными критериями с наиболее простого случая, который хотя и редко встречается на практике, позволяет проиллюстрировать основные принципы и понятия, используемые при практической проверке гипотез.

1.4.1 Тест для среднего при известной дисперсии

Пусть дана выборка $X_1, \ldots, X_N$ из нормального распределения с известной дисперсией σ^2 и неизвестным средним. Необходимо проверить нулевую гипотезу $H_0\colon \mathsf{E}X = a$.

Известно, что в случае, когда нулевая гипотеза верна, статистика[2]

$$Z = \sqrt{N} \cdot \frac{\overline{X} - a}{\sigma}$$

имеет стандартное нормальное распределение. Даже если исходное предположение о нормальности данных не выполнено, то распределение статистики будет асимптотически нормальным, в силу ЦПТ; т.е. точный тест станет асимптотическим.

Теперь надо разметить критические области. Выбор критических областей зависит от альтернативы, против которой мы хотим, чтобы наш критерий был мощен. Рассмотрим возможные варианты.

Двусторонний критерий Рассмотрим альтернативную гипотезу $H_1\colon \mathsf{E}X = a' \neq a$. Пусть верна альтернатива H_1. Тогда в силу ЗБЧ:

$$\overline{X} - a \to a' - a \neq 0 \quad \Rightarrow \quad |Z| \to \infty.$$

Следовательно, наибольшей мощности мы достигнем, разметив критические области на левом и правом хвосте распределения статистики, именно такие значения характерны для этой альтернативы.

Разумно взять симметричные хвосты — в самом деле, мы же не знаем, больше реальное значение среднего или меньше, чем предполагаемое. Для того, чтобы критерий имел нужный уровень значимости, нам нужно отложить слева и справа квантили уровня $\alpha/2$.

Итого, имеем: если $Z \in S = (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$, то гипотезу H_0 следует отвергнуть в пользу альтернативы H_1[3].

Правосторонний критерий Рассмотрим альтернативную гипотезу $H_1\colon \mathsf{E}X = a' > a$. Пусть верна альтернатива H_1, тогда в силу ЗБЧ:

$$\overline{X} - a \to a' - a > 0 \Rightarrow Z \to +\infty.$$

Следовательно, наибольшей мощности мы достигнем, разметив критическую область на правом хвосте идеального распределения статистики.

Итого, если $Z \in S = (z_{1-\alpha}, \infty)$, то гипотезу H_0 следует отвергнуть в пользу альтернативы H_1.

Левосторонний критерий Рассмотрим альтернативную гипотезу $H_1\colon \mathsf{E}X = a' < a$. Рассуждая аналогично предыдущему случаю, получаем, что если $Z \in S = (-\infty, z_\alpha)$, то гипотезу H_0 необходимо отвергнуть в пользу альтернативы H_1.

Простая альтернатива Если в качестве альтернативы дано конкретное значение (т.е. $H_1\colon \mathsf{E}X = b \neq a$), то следует применить односторонний критерий, в зависимости от того, больше это значение или меньше, чем в нулевой гипотезе.

[2]Здесь и ниже: $\overline{X}$ означает выборочное среднее, $s(X)$ — выборочное **подправленное** стандартное отклонение

[3]Здесь и везде ниже: $z_\varkappa$ обозначает квантиль уровня $\varkappa$ стандартного нормального распределения

Замечание 2. Касательно выбора того, какую именно альтернативу использовать в конкретном случае. Основная мысль — нужно формулировать нулевую и альтернативную гипотезы до того, как мы "посмотрели на данные"[4]. В противном случае скорее всего будет подгон и результат получится неверным.

Например, если среднее уже посчитано, то очень хочется провести односторонний тест, потому что "если мы видим, что среднее меньше, то зачем рассматривать случай, когда оно может быть больше". **Так рассуждать нельзя!** В этом случае реальный уровень значимости окажется вдвое больше заданного, а тест фактически окажется двусторонним, мы просто отсекли одну половину двусторонней критической области сразу, посмотрев на среднее.

Замечание 3. Понятие нулевой и альтернативной гипотез довольно условно. Выбор обычно обусловлен следующим: нулевая гипотеза должна быть простой, т.к. мы должны знать идеальное распределение статистики критерия, чтобы разметить критические области. А для альтернативной гипотезы нам достаточно только знать поведение статистики.

Если же задача допускает две постановки, то часто рассуждают следующим образом. Вероятность ошибки первого рода (вероятность ошибочно отвергнуть верную гипотезу) для правильно построенного критерия равна α, мы можем сами ее задавать. Вероятность же ошибки второго рода и зависит от конкретной альтернативы и в общем случае не вычисляется. Поэтому за нулевую гипотезу принимают обычно что-то "плохое" (например, что новое лекарство ничуть не лучше старого). Тогда в случае, если мы отвергли гипотезу, мы можем сказать, что "все хорошо" с вероятностью, не ниже, чем $1 - \alpha$, а α мы можем задать из соображений стандартов индустрии (традиционно считают $\alpha = 0.05$, это считается приемлемой вероятностью принять неверное решение).

1.5 Тест для среднего с неизвестной дисперсией

Теперь пришло время рассмотреть один из самый известных и используемых тестов в статистике.

Пусть дана выборка $X_1, \ldots, X_N$ из нормального распределения с неизвестной дисперсией. Необходимо проверить нулевую гипотезу $\mathsf{E}X = a$.

Известно, что в случае, когда нулевая гипотеза верна, статистика

$$t = \sqrt{N} \cdot \frac{\overline{X} - a}{s(X)},$$

где $s(X)$, здесь и далее, **подправленное** выборочное стандартное отклонение, имеет распределение t-Стюдента с $(N-1)$-й степенью свободы. Для больших N, а также для случая, когда не выполнено предположение о нормальности, но дисперсия распределения значений X конечна, в силу ЦПТ вместо распределения Стюдента можно использовать стандартное нормальное распределение.

Для различных альтернатив критерий принимает двусторонний или односторонний вид, аналогично случаю с известной дисперсией:

[4] В идеале — до проведения любых экспериментов. Но к учебным задачам это не относится; да и в жизни — далеко не всегда

Двусторонний критерий $H_1\colon \mathsf{E}X \neq a$. Критическая область: $S = (-\infty, t_{\alpha/2,N-1}) \cup (t_{1-\alpha/2,N-1}, \infty)$ для точного критерия[5] или $S = (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2})$ для нормальной аппроксимации.

Правосторонний критерий $H_1\colon \mathsf{E}X > a$. Критическая область: $S = (t_{1-\alpha,N-1}, \infty)$ для точного критерия или $S = (z_{1-\alpha})$ для нормальной аппроксимации.

Левосторонний критерий $H_1\colon \mathsf{E}X < a$. Критическая область: $S = (-\infty, t_{\alpha,N-1})$ для точного критерия или $S = (-\infty, z_\alpha)$ для нормальной аппроксимации.

1.6 Z-Тест для пропорции (доли) в модели Бернулли

Этот тест подходит для примера с дегустацией чая, с которого мы начали рассказ о проверке гипотез.

Пусть дана выборка $X_1, \ldots, X_N$ из распределения Бернулли с неизвестным параметром p. Необходимо проверить нулевую гипотезу $p = a$.

В случае, когда нулевая гипотеза верна, статистика

$$Z = \sqrt{N} \cdot \frac{\overline{X} - a}{\sqrt{a(1-a)}}$$

асимптотически имеет стандартное нормальное распределение.

Как и в случае проверки гипотез о среднем, тест критерий принимает односторонний или двусторонний вид в зависимости от альтернативы.

Двусторонний критерий $H_1\colon p \neq a$. Критическая область: $S = (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$.

Правосторонний критерий $H_1\colon p > a$. Критическая область: $S = (z_{1-\alpha}, \infty)$.

Левосторонний критерий $H_1\colon p < a$. Критическая область: $S = (-\infty, z_\alpha)$.

1.7 Тест для дисперсии при неизвестном среднем в нормальной модели

Замечание 4. Важно! Тесты для дисперсии в нормальной модели можно применять только для нормально распределенных данных, в отличие от тестов для среднего, они неверны при отсутствии нормальности, даже асимптотически

Пусть дана выборка $X_1, \ldots, X_N$ из нормального распределения с неизвестным средним. Необходимо проверить нулевую гипотезу $\mathsf{D}X = a$.

Известно, что в случае, когда нулевая гипотеза верна, статистика

$$t = (N-1) \cdot \frac{s^2(X)}{a}$$

имеет распределение χ^2 с $N-1$-й степенью свободы.

В зависимости от альтернативы H_1, как и ранее, получаются двусторонние и односторонние критерии:

[5]Здесь и везде: $t_{\varkappa,d}$ обозначает $\varkappa$-квантиль распределения t-Стьюдента с d степенями свободы

Двусторонний критерий $H_1\colon \mathsf{D}X \neq a$. Критическая область:[6] $S = (0, \chi^2_{\alpha/2,N-1}) \cup (\chi^2_{1-\alpha/2,N-1}, \infty)$.

Правосторонний критерий $H_1\colon \mathsf{D}X > a$. Критическая область: $S = (\chi^2_{1-\alpha,N-1}, \infty)$.

Левосторонний критерий $H_1\colon \mathsf{D}X < a$. Критическая область: $S = (0, \chi^2_{\alpha,N-1})$.

1.8 Тест для дисперсии при известном среднем в нормальной модели

Пусть дана выборка $X_1, \ldots, X_N$ из нормального распределения с известным средним μ. Необходимо проверить нулевую гипотезу $\mathsf{D}X = a$.

Известно, что в случае, когда нулевая гипотеза верна, статистика

$$t = N \cdot \frac{\overline{(X-\mu)^2}}{a}$$

имеет распределение χ^2 с N степенями свободы.

В зависимости от альтернативы H_1, как и ранее, получаются двусторонние и односторонние критерии:

Двусторонний критерий $H_1\colon \mathsf{D}X \neq a$. Критическая область: $S = (0, \chi^2_{\alpha/2,N}) \cup (\chi^2_{1-\alpha/2,N}, \infty)$.

Правосторонний критерий $H_1\colon \mathsf{D}X > a$. Критическая область: $S = (\chi^2_{1-\alpha,N}, \infty)$.

Левосторонний критерий $H_1\colon \mathsf{D}X < a$. Критическая область: $S = (0, \chi^2_{\alpha,N})$.

2 Двухвыборочные тесты

Довольно часто данные эксперимента представляют собой не одну выборку, а несколько независимых, которые необходимо сравнить.

Рассмотрим, например, такую задачу: провели испытание двух лекарств на двух разных тестовых группах. Необходимо определить, является ли одно лекарство более эффективным, чем другое. Понятно, что если просто посчитать процент выздоровевших в каждой группе и сравнить, мы не сможем ответить на вопрос, потому что различие в группах может быть не систематическим, а вызвано чистой случайностью (при этом вполне возможно, что различие будет не в пользу реально более эффективного лекарства).

На этот случай теория и практика доверительных-критических областей естественно обобщается. Нужно просто рассмотреть вероятностное пространство всех выборок, далее также разбить все пространство возможных исходов на области, в зависимости от гипотез и данного уровня значимости.

Идея использовать статистики критерия здесь тоже применима, только статистика уже будет функцией от нескольких выборок, т.е. $t = t(X^{(1)}, X^{(2)})$.

[6] Здесь и ниже: $\chi^2_{\varkappa,d}$ обозначает квантиль уровня $\varkappa$ распределения χ^2 с d степенями свободы

2.1 Двухвыборочный тест для независимых выборок для проверки равенства средних в нормальной модели с различными дисперсиями (Welch t-test)

Вообще, для проверки равенства средних в двух выборках есть много тестов для разных случаев (когда совпадают/не совпадают размеры выборок, совпадают/не совпадают дисперсии, etc), но мы рассмотрим наиболее часто полезный на практике самый общий вариант (когда неизвестно ничего).

Пусть даны выборки $X_1^{(1)}, \ldots, X_{N_1}^{(1)}$ и $X_1^{(2)}, \ldots, X_{N_2}^{(2)}$ из нормальных распределений с неизвестными (и необязательно совпадающими) дисперсиями. Необходимо проверить нулевую гипотезу $\mathsf{E}X^{(1)} = \mathsf{E}X^{(2)}$.

Известно, что в случае, когда нулевая гипотеза верна, статистика

$$Z = \frac{\overline{X^{(1)}} - \overline{X^{(2)}}}{\sqrt{\frac{s^2(X^{(1)})}{N_1} + \frac{s^2(X^{(2)})}{N_2}}}$$

имеет распределение t-Стюдента с ν степенями свободы[7], но для больших N_1, N_2 можно воспользоваться нормальной аппроксимацией. В случае, когда предположение о нормальности наблюдений не выполняется, результат становится асимптотическим и также можно применить нормальную аппроксимацию.

И снова, в зависимости от альтернативы H_1, получаются двусторонние и односторонние критерии (используем нормальную аппроксимацию):

Двусторонний критерий $H_1\colon \mathsf{E}X^{(1)} \neq \mathsf{E}X^{(2)}$. Критическая область: $S = (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$.

Правосторонний критерий $H_1\colon \mathsf{E}X^{(1)} > \mathsf{E}X^{(2)}$. Критическая область: $S = (z_{1-\alpha}, \infty)$.

Левосторонний критерий $H_1\colon \mathsf{E}X^{(1)} < \mathsf{E}X^{(2)}$. Критическая область: $S = (-\infty, z_\alpha)$.

2.2 Двухвыборочный тест для независимых выборок для проверки равенства средних в нормальной модели с равными дисперсиями (pooled t-test)

Наряду с тестом Велча рассмотрим также менее общий, но, тем не менее, полезный случай, когда дисперсии неизвестны, но известно, что они совпадают.

Пусть даны выборки $X_1^{(1)}, \ldots, X_{N_1}^{(1)}$ и $X_1^{(2)}, \ldots, X_{N_2}^{(2)}$ из нормальных распределений с неизвестными, но совпадающими дисперсиями. Необходимо проверить нулевую гипотезу $\mathsf{E}X^{(1)} = \mathsf{E}X^{(2)}$.

Известно, что в случае, когда нулевая гипотеза верна, статистика

$$t = \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_1}} \cdot \frac{\overline{X^{(1)}} - \overline{X^{(2)}}}{\sqrt{s_{\mathrm{p}}^2(X^{(1)}, X^{(2)})}}$$

[7]Значение ν имеет вид:

$$\nu \approx \frac{\left(\frac{s^2(X^{(1)})}{N_1} + \frac{s^2(X^{(2)})}{N_2}\right)^2}{\frac{(s^2(X^{(1)})/N_1)^2}{N_1-1} + \frac{(s^2(X^{(2)})/N_2)^2}{N_2-1}}.$$

имеет распределение t-Стюдента с $(N_1 + N_2 - 2)$-мя степенями свободы. Для больших N_1, N_2, а также для случая, когда предположение о нормальности не выполняется, допустимо применять нормальную аппроксимацию. В данном случае "-2" возникает из-за того, что мы оцениваем *два* средних (и из-за этого теряем две степени свободы).

Здесь

$$s_{\mathrm{p}}^2 = \frac{s^2(X^{(1)}) \cdot (N_1 - 1) + s^2(X^{(2)}) \cdot (N_2 - 1)}{(N_1 - 1) + (N_2 - 1)}$$

так называемая *pooled variance.* Вообще, слово *pooled* означает, что мы рассматриваем статистику, полученную по нескольким выборкам, *слитым (pool)* в одну.

И снова, в зависимости от альтернативы H_1, получаются двусторонние и односторонние критерии:

Двусторонний критерий $H_1\colon \mathsf{E}X^{(1)} \neq \mathsf{E}X^{(2)}$. Критическая область: $S = (-\infty, z_{\alpha/2}) \cup (t_{1-\alpha/2, N_1+N_2-2}, \infty)$.

Правосторонний критерий $H_1\colon \mathsf{E}X^{(1)} > \mathsf{E}X^{(2)}$. Критическая область: $S = (t_{1-\alpha, N_1+N_2-2}, \infty)$.

Левосторонний критерий $H_1\colon \mathsf{E}X^{(1)} < \mathsf{E}X^{(2)}$. Критическая область: $S = (-\infty, t_{\alpha, N_1+N_2-2})$.

2.3 Проверка равенства пропорций (долей) успехов в двух выборках

Этот тест поможет нам решить задачу с двумя группами пациентов, которая была сформулирована в начале раздела.

Для этой задачи существует несколько похожих тестов, мы рассмотрим один из них — так называемый pooled Z-test Пирсона.

Пусть даны выборки $X_1^{(1)}, \ldots, X_{N_1}^{(1)}$ и $X_1^{(2)}, \ldots, X_{N_2}^{(2)}$ из распределений Бернулли с параметрами p_1 и p_2. Необходимо проверить нулевую гипотезу о равенстве пропорций $H_0\colon p_1 = p_2$.

Известно, что в случае, когда нулевая гипотеза верна, статистика

$$Z = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \cdot \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1 - \widehat{p})}} \sim \mathcal{N}(0, 1),$$

имеет асимптотически стандартное нормальное распределение. Здесь

$$\widehat{p}_1 = \overline{X^{(1)}}; \quad \widehat{p}_2 = \overline{X^{(2)}};$$
$$\widehat{p} = \frac{N_1 \widehat{p}_1 + N_2 \widehat{p}_2}{N_1 + N_2}.$$

$\widehat{p}$ называется *pooled proportion.* Вообще, слово *pooled* означает, что мы рассматриваем статистику, полученную по нескольким выборкам, *слитым (pool)* в одну.

И снова, в зависимости от альтернативы H_1, получаются двусторонние и односторонние критерии:

Двусторонний критерий $H_1\colon p_1 \neq p_2$. Критическая область: $S = (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$.

Правосторонний критерий $H_1\colon p_1 > p_2$. Критическая область: $S = (z_{1-\alpha}, \infty)$.

Левосторонний критерий $H_1\colon p_1 < p_2$. Критическая область: $S = (-\infty, z_\alpha)$.

3 Критерии согласия

На практике часто возникает задача проверки соответствия выборки какому-то распределению. Критерии выше позволяли проверить или сравнить распределения имеющихся выборок, но в смысле конкретных характеристик. Тест для среднего, конечно, можно использовать для проверки согласия с каким-то распределением с известным средним, критерий даже будет корректен, но он будет мощен только против альтернатив-распределений с другим средним и совершенно не будет обнаруживать различие в формах распределений.

Для проверки соответствия распределения выборки какому-то другому известному распределению существуют специальные критерии, так называемые *критерии согласия (goodness-of-fit)*. Обычно статистика критерия согласия имеет смысл расстояния между предполагаемым распределением и имеющимся. Под расстоянием имеется ввиду некоторая метрика на множестве распределений. Существует несколько употребимых метрик, каждой из которых соответствует критерий.

3.1 Критерий Колмогорова-Смирнова

Одним из наиболее известных критериев согласия является критерий Колмогорова-Смирнова. В качестве статистики используется нормированное уклонение Чебышева между эмпирической (выборочной) функцией распределения и теоретической ("идеальной").

Формально: пусть дана выборка $X_1, \ldots, X_N$ и нам нужно проверить эту выборку на соответствие непрерывному распределению с функцией распределения F. Если нулевая гипотеза, $H_0\colon X \sim F$, верна, то статистика

$$t_{\mathrm{K-S}} = \sqrt{N} \sup_x |F(x) - F_N(x, X)|,$$

где $F_N(\cdot, X)$ — эмпирическая функция распределения выборки X, асимптотически имеет распределение Колмогорова-Смирнова[8].

Если же нулевая гипотеза неверна, то расстояние между функциями распределения не стремится к нулю, следовательно, статистика $t_{\mathrm{K-S}}$ стремится к бесконечности.

Альтернативная гипотеза в данном случае возможна только одна: $H_1\colon X \not\sim F$. Критическая область: $S = (q_{1-\alpha}, \infty)$, где $q_\varkappa$ — $\varkappa$-квантиль распределения Колмогорова-Смирнова [9]. Для уровня $\alpha = 0.05$: $q_{0.95} \approx 1.36$, для уровня $\alpha = 0.01$: $q_{0.99} \approx 1.63$.

Замечание 5. Важно! Критерии типа расстояния: Колмогорова-Смирнова, Смирнова-Крамера-фон Мизеса (aka омега-квадрат, ω^2), Пирсона (aka хи-квадрат, χ^2) и другие всегда имеют критическую область справа, т.е. мы отвергаем гипотезу, если расстояние между предполагаемым и наблюдаемым распределениями большое.

[8]Это распределение неслучайно так названо, разумеется. Его плотность известна, есть возможность вычислить любые его квантили

[9]Конечно, лучше использовать функцию из статистического пакета, тем более, что в хороших пакетах для маленьких N используются квантили не асимптотического распределения, а точного

Критерий Колмогорова детектирует любое различие в распределениях и альтернативная гипотеза для него: "Распределение выборки какое-то другое". Односторонний в смысле альтернативы (обнаруживающий только смещение распределения в определенную сторону, а не произвольное отклонение) критерий Колмогорова-Смирнова существует, но он отличается от двустороннего не положением критической области, а видом статистики критерия.