2. Проверка гипотез о параметрах распределения

Покажем, как задачу A_1 решить с помощью доверительного оценивания. Построим доверительный интервал для параметра θ:

$$(\bar{x} - x_{\alpha/2}\sigma/\sqrt{n};\ \bar{x} + x_{\alpha/2}\sigma/\sqrt{n}) =$$
$$= (31.8 - 1.64 \cdot 4/5;\ 31.8 + 1.64 \cdot 4/5) = (30.488;\ 33.112).$$

Величина $\theta_0 = 30$, соответствующая гипотезе H_0, не принадлежит данному интервалу.

Является ли непринадлежность параметра θ_0 гипотезы доверительному интервалу необходимым и достаточным условием отклонения гипотезы? Ответ утвердительный. Действительно, гипотеза при данном уровне значимости α принимается тогда и только тогда, когда значение параметра θ_0 принадлежит доверительному интервалу с уровнем доверия $1 - \alpha$.

Это доказывается довольно просто. При задании значения $x_{\alpha/2}$ для доверительного интервала с уровнем доверия $1 - \alpha$ мы исходили из уравнения

$$\mathbf{P}\left(\frac{\sqrt{n}|\bar{X} - \theta_0|}{\sigma} > x_{\alpha/2}\right) = 1 - \alpha$$

и показывали, что соотношение

$$\hat{T}_n = \frac{\sqrt{n}|\bar{X} - \theta_0|}{\sigma} > x_{\alpha/2}$$

эквивалентно

$$\theta_0 \in (\bar{X} - x_{\alpha/2}\sigma/\sqrt{n},\ \bar{X} + x_{\alpha/2}\sigma/\sqrt{n}).$$

Но выражение $\sqrt{n}|\bar{X} - \theta_0|/\sigma$ задает тестовую статистику критерия проверки гипотезы H_0: $\theta = \theta_0$ против альтернативы

H_1: $\theta \neq \theta_0$, причем $x_{\alpha/2}$ задается тем же самым уравнением (3.2).

Таким образом, мы можем сформулировать общий принцип. *Если доверительный интервал накрывает значение* $\theta = \theta_0$ *гипотезы, то гипотеза принимается. В противном случае гипотеза отклоняется.*

Как этот принцип действует в случае односторонних альтернатив? Оказывается, что бывают односторонние доверительные интервалы, а именно левый $(-\infty,\ \bar{X} + x_\alpha \sigma/\sqrt{n})$ он соответствует постановке задачи проверки гипотезы $H_0 : \theta < \theta_0$ против альтернативы H_1: $\theta > \theta_0$ и правый $(\bar{X} - x_\alpha \sigma/\sqrt{n},\ \infty)$ соответствует проверке гипотезы $H_0 : \theta > \theta_0$ против альтернативы H_1: $\theta < \theta_0$.

Как следствие, рассуждая аналогично, мы получаем тестовые статистики и для других задач проверки гипотез.

а) Проверка гипотезы о среднем нормального распределения с неизвестной дисперсией.

При проверке гипотезы H_0: $\theta = \theta_0$ против альтернативы H_1: $\theta \neq \theta_0$ критерий значимости имеет вид

$$\hat{T}_n = \frac{\sqrt{n}|\bar{X} - \theta_0|}{\bar{s}} > t_{\alpha/2,n-1},$$

где $t_{\alpha/2,n-1}$ находится из таблиц распределения Стьюдента с $n-1$ степенями свободы.

б) Проверка гипотезы о вероятности p биномиального распределения

При проверке гипотезы H_0: $p = p_0$ против альтернативы H_1: $p \neq p_0$ критерий значимости имеет вид

$$\hat{T}_n = \frac{|\hat{p}_n - p_0|\sqrt{n}}{\sqrt{\hat{p}_n(1-\hat{p}_n)}} > x_{\alpha/2},$$

где $x_{\alpha/2}$ удовлетворяет соотношению (3.2) и находится из таблиц стандартного нормального распределения.

Так же как и в случае построения доверительного интервала, если $n \leq 15$ при проверке гипотезы надо пользоваться таблицами биномиального распределения и если $n\hat{p}$ мало то пуассоновской аппроксимацией биномиального распределения.

Для односторонних гипотез и альтернатив вида $p > p_0$ критерии отличаются только тем, что в тестовой статистике $\hat{T}_n$ убирается модуль, критическое значение $x_{\alpha/2}$ заменяется на x_α и дальше критическая область находится также как в случае задач проверки гипотез о среднем нормального распределения с известной дисперсией.

в) Проверка гипотезы о дисперсии $\sigma = \sigma_0$ против альтернативы
$\sigma \neq \sigma_0$ в случае нормального распределения.

Используя аргументы из вывода формулы для границ доверительного интервала для дисперсии σ^2 получаем, что гипотеза принимается, если

$$\chi^2_{\nu,1-\alpha/2} < \frac{(n-1)\hat{s}^2}{\sigma_0^2} < \chi^2_{\nu,\alpha/2}.$$

Величины $\chi^2_{\nu,\alpha/2}$ и $\chi^2_{\nu,1-\alpha/2}$, удовлетворяющие соотношению (2.11), находим с помощью таблиц χ^2-распределения.

При использовании пакетов статистических программ обычно используется понятие *p-значения (probability value)* вместо понятия уровня значимости α. Именно оно обычно появляется на экране монитора.

p-значение α_p является наибольшей вероятностью ошибки первого рода, при которой принимается гипотеза.

Таким образом, правило принятия гипотезы таково.

в1) Если α_p больше уровня значимости α ($\alpha_p > \alpha$), то гипотеза H_0 принимается.

в2) Если α_p меньше уровня значимости α ($\alpha_p < \alpha$), то гипотеза H_0 отвергается.

Так если $\alpha_p = 0.4$, то для $\alpha < 0.4$, т.е. практически всегда, гипотеза принимается. Если $\alpha_p = 0.005$, то гипотеза принимается для $\alpha < 0.005$, т.е. она не принимается практически никогда.

По величине p-значения мы можем сразу сказать для всех возможных значений вероятностей ошибок первого рода принимаем мы гипотезу или нет. В этом и заключается удобство p-значения.

Поясним как находится p-значение. По исходным данным вычисляется значение тестовой статистики. Допустим оно равно $t = \hat{T}_n$. Для этого t находим вероятность α_p такую, что

$$\mathbf{P}(\hat{T}_n > t) = \alpha_p. \tag{3.3}$$

Эта вероятность α_p и является p-значением.

В случае задач проверки гипотезы о среднем нормального распределения с известной дисперсией, если гипотеза $H_0 : \theta = \theta_0$ двусторонняя, то $\alpha_p = 2\Phi(-|T_n|)$, если гипотеза одностороння $H_0 : \theta < \theta_0$, то $\alpha_p = 1-\Phi(T_n)$ и если $H_0 : \theta > \theta_0$, то $\alpha_p = \Phi(T_n)$. Если дисперсия в данной постановке неизвестна, то во всех выражениях для α_p нормальное распределение заменяется на соответствующее распределение Стьюдента с $n-1$ степенью свободы. Для задач проверки гипотез для биномиального распределения формулы вычисления p-значения те же, что и для нормального распределения.

Так, в примере A_1 мы получили $t = 2.25$, отсюда

$$\alpha_p = \mathbf{P}\left(\sqrt{n}\frac{|\bar{X} - \theta_0|}{\sigma} > 2.25\right) = \frac{2}{\sqrt{2\pi}}\int_{2.25}^{\infty} \exp\{-s^2/2\}ds = 0.025.$$

Так как в условии задачи уровень значимости $\alpha = 0.1$ и, следовательно, $\alpha > \alpha_p$, то гипотеза H_0 отвергается.

Как обосновать уравнение (3.3), задающее p-значение α_p? Для этого достаточно обосновать утверждения *в1)* и *в2)*.

Из уравнения

$$\mathbf{P}(\hat{T}_n > x_\alpha) = \alpha$$

мы видим, что чем больше x_α, тем меньше α, поэтому если $t = \hat{T}_n > x_\alpha$, то $\alpha_p < \alpha$. С другой стороны, неравенство $t = \hat{T}_n > x_\alpha$ означает, что значение t тестовой статистики $\hat{T}_n$, найденное по выборке, попало в критическую зону, поэтому мы отвергаем гипотезу и тем самым доказываем пункт *в2)*. Пункт *в1)* доказывается аналогично.

г) Проверка гипотезы о равенстве математических ожиданий двух нормально распределенных совокупностей

Довольно часто менеджеру приходится сравнивать значения средних двух выборок, чтобы прийти к определенному решению. Так, например, в задаче A_2, в принципе, мы должны были бы сравнивать числовые данные о спросе до и после рекламной кампании. Другими примерами являются перечисленные ниже задачи.

Действительно ли в компании мужчины и женщины за одинаковую работу и получают одну и ту же заработную плату?

Является ли продолжительность жизни продукции (телевизоров, компьютеров и т. п.) в двух компаниях одинаковой?

Одинакова ли в среднем стоимость продовольственных товаров в двух магазинах и т. п.?

Пусть две независимые случайные величины X и Y распределены нормально и $X \in N(m_x, \sigma_x^2),\ Y \in N(m_y, \sigma_y^2)$. Пусть имеются две независимые выборки

$$X_1, \ldots, X_{n_1}, \quad Y_1, \ldots, Y_{n_2}$$

объемов n_1 и n_2 соответственно для X и Y и

$$\bar{X} = \frac{1}{n_1}\sum_{j=1}^{n_1} X_j, \quad \bar{Y} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_j,$$

$$s_x^2 = \frac{1}{n_1}\sum_{j=1}^{n_1}(X_j - \bar{X})^2, \quad s_y^2 = \frac{1}{n_2}\sum_{j=1}^{n_2}(Y_j - \bar{Y})^2.$$

Нужно проверить гипотезу H_0, состоящую в том, что $EX = EY$, т. е.

$$H_0\colon \quad m_x = m_y.$$

Для проверки гипотезы H_0 следует воспользоваться различными критериями значимости в зависимости от того, известны или не известны дисперсии случайных величин X и Y. Рассмотрим оба случая.

г1) Пусть σ_x и σ_y известны. Так как

$$\bar{X} \in N\left(m_x, \frac{\sigma_x^2}{n_1}\right), \quad \bar{Y} \in N\left(m_y, \frac{\sigma_y^2}{n_2}\right),$$

а случайные величины $\bar{X}$ и $\bar{Y}$ независимы, то

$$\sigma^2_{(\bar{X}-\bar{Y})} = \mathbf{D}(\bar{X}-\bar{Y}) = \mathbf{D}\bar{X} + \mathbf{D}\bar{Y} = \frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}$$

(дисперсия разности равна сумме дисперсий, так как по свойству дисперсий $\mathbf{D}(X) = \mathbf{D}(-X)$). Отсюда следует, что

$$\bar{X} - \bar{Y} \in N(m_x - m_y, \sigma^2(\bar{X} - \bar{Y})).$$

Для проверки гипотезы H_0 в качестве критерия рассмотрим величину

$$z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}}. \tag{3.4}$$

Если гипотеза H_0 верна , то $z \in N(0,1)$ и в качестве критической области для двусторонней альтернативы следует взять область больших по модулю отклонений z, т. е. $|z| > x_{\alpha/2}$, где $x_{\alpha/2}$ определяется из (3.2), а α — уровень значимости.

Пример 3.1. Пусть $\bar{X} = 18.4,\ \bar{Y} = 19\ \sigma_X^2 = 1.2,\ \sigma_Y^2 = 3,\ n_1 = 20,\ n_2 = 30$. Найдем значение критерия z:

$$z = \frac{18.4 - 19}{\sqrt{1.2/20 + 3/30}} = -1.5.$$

Пусть $\alpha = 0.05$, тогда $z_\alpha = 1.96$, и так как $|z| < z_\alpha(|-1.5| < 1.96)$, то принимаем гипотезу о равенстве математических ожиданий.

г2) Рассмотрим гипотезу H_0: $m_x = m_y$, когда дисперсии σ_x и σ_y не известны, но выполнено условие

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2. \tag{3.5}$$

При выполнении (3.5) величина z в (3.4) принимает вид

$$\xi = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{3.6}$$

причем при выполнении гипотезы H_0 случайная величина $\xi \in N(0,1)$. Поэтому для построения критерия достаточно заменить дисперсию σ^2 ее оценкой. Построим такую оценку σ^2. Согласно п. 5г) гл. 2 случайные величины $\eta_1 = n_1 s_x^2/\sigma^2$ и $\eta_2 = n_2 s_y^2/\sigma^2$ имеют χ^2-распределение соответственно с $n_1 - 1$ и $n_2 - 1$ степенями свободы, а так как X и Y независимы, то $\eta_1 + \eta_2$ имеет χ^2 – распределение с $n_1 + n_2 - 2$ степенями свободы. Разделив ξ на $\frac{\sqrt{\eta_1+\eta_2}}{\sqrt{n_1+n_2-2}}$ получаем тестовую статистику

$$\frac{\xi\sqrt{n_1 + n_2 - 2}}{\sqrt{\eta_1 + \eta_1}}, \tag{3.7}$$

которая уже не зависит от σ и имеет распределение Стьюдента с n_1+n_2-2 степенями свободы Ее можно записать в более явном виде

$$t = \frac{\bar{X} - \bar{Y}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{3.8}$$

где

$$s_p = \frac{n_1 s_1^2 + n_2 s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Далее поступаем следующим образом. Задаем уровень значимости $\alpha = 0.05; 0.01$ и т.д. и затем по таблице распределения Стьюдента находим $t_{\alpha/2,k}$, где $k = n_1+n_2-2$. Если значения $|t| > t_{\alpha/2,k}$, то отвергаем H_0, если $|t| < t_{\alpha/2,k}$, то говорим, что гипотеза H_0 не противоречит данным.

г3). Рассмотрим гипотезу $H_0 : m_x = m_y$, когда дисперсии σ_x^2, σ_y^2 не известны и $\sigma_x \neq \sigma_y$. Тогда в тестовой статистике (3.4) дисперсии σ_x^2 и σ_y^2 заменяем их оценками s_x^2 и s_y^2 соответственно. В результате получаем тестовую статистику

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}}.$$

Распределение подкоренного выражения имеет довольно сложный вид. Поэтому довольно сложный вид имеет и распределение тестовой статистики t. Рекомендуется находить критические значения для t из таблиц распределения Стьюдента с $k = \min n_1 - 1, n_2 - 1$ степенями свободы, поскольку как можно показать

$$P(t > t_{\alpha,k}) \leq \alpha.$$

д).Проверка гипотезы о равенстве двух пропорций. Пусть $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$ две выборки независимых случайных величин имеющих биномиальное распределение. Пусть $P(X_i = 1) = p_1, P(X_i = 0) = 1 - p_1, 1 \leq i \leq n$ и $P(Y_j = 1) = p_2, P(Y_j = 0) = 1-p_2, 1 \leq j \leq m$. Необходимо проверить гипотезу

$$H_0 : p_1 = p_2$$

против альтернатив

$$H_1 : p_1 \neq p_2.$$

Тестовая статистика строится на основе относительных частот

$$\hat{p}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{p}_2 = \frac{1}{m}\sum_{j=1}^{m} Y_j.$$

При этом используется тот факт, что при справедливости гипотезы, для больших значений $n_1, n_2 (n_1, n_2 > 20)$ распределение статистики

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1-\bar{p})}\sqrt{\frac{1}{n}+\frac{1}{m}}}, \quad \bar{p} = p_1 = p_2$$

хорошо аппроксимируется стандартным нормальным распределением. Данное утверждение следует из центральной предельной теоремы.

Взяв в качестве оценки $\bar{p}$ статистику

$$\tilde{p} = \frac{\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j}{n+m} = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m},$$

получаем вид тестовой статистики для проверки гимпотезы

$$T = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\tilde{p}(1-\tilde{p})}\sqrt{\frac{1}{n}+\frac{1}{m}}}.$$

Критические значения x_α находятся из таблиц стандартного нормального распределения.

Пример. Корпорация "Сигма"заинтересовалась изменился ли спрос на ее продукцию. В прошлом году было опрошено 50 человек и оказалось, что 9 человек предпочитают продукцию данной фирмы. Опрос 60 человек показал, что 10 человек предпочитают продукцию данной фирмы. Было решено проверить гипотезу с уровнем значимости 5 процентов. Найти p-значение.

Таким образом надо проверить гипотезу, что доля рынка продукции не изменилась, предполагая что в предыдущем году она также была оценена с некоторой погрешностью. Имеем

$$\hat{p}_X = \frac{9}{50} = 0.18, \quad \hat{p}_Y = \frac{10}{60} = 0,1666,$$

$$\tilde{p} = \frac{9+10}{50+60} = \frac{19}{11}.$$

Подставляя значения $\hat{p}_X, \hat{p}_Y$ и $\tilde{p}$, получаем значение тестовой статистики

$$\hat{T} = \frac{|0.18 - 0.1666|}{\sqrt{\frac{19}{110}}\sqrt{\frac{1}{50}+\frac{1}{60}}} = \frac{0.01333}{0.378 \cdot 0.0366} = \frac{0.0133}{0.0138} = 0.96.$$

Соответствующее p-значение равно $2\Phi(-0.96) = 2{\cdot}0.168 = 0.336$ и гипотеза принимается.

e) Проверка гипотезы о равенстве дисперсий двух нормально распределенных совокупностей

Пусть у нас есть та же самая постановка задачи, что и в предыдущем пункте. Даны две выборки независимых нормально распределенных случайных величин с неизвестными математическими ожиданиями и дисперсиями. Стоит задача проверки гипотезы о равенстве дисперсий

$$H_0\colon \quad \sigma_x = \sigma_y.$$

В данной задаче в качестве тестовой статистики возьмем

$$\hat{T}_n = s_x^2/s_y^2.$$

Статистика $\hat{T}_n$ имеет распределение Фишера с $n_1 - 1$ и $n_2 - 1$ степенями свободы.

Пример 3.3. Компания, производящая электрические батарейки, сравнивает две технологии их производства по продолжительности жизни батареек. Результаты испытаний дали следующие данные

$$n_1 = 25, \quad \bar{X}_1 = 3.6, \quad s_1 = 0.36,$$

$$n_2 = 27, \quad \bar{X}_2 = 3.8, \quad s_2 = 0.32.$$

Проверить гипотезу вторая технология лучше первой с уровнем значимости 5 процентов. Найти p-значение.

Из условий задачи следует, что нужно проверить гипотезу $H_0 : m_{X_2} > m_{X_1}$ против альтернативы $H_1 : m_{X_1} \geq m_{X_2}$.

Значения дисперсий неизвестны. Поэтому в первую очередь надо решить какой тестовой статистикой воспользоваться – с равными или неравными дисперсиями. Для этого надо проверить гипотезу о равенстве дисперсий.

Проверка гипотезы о равенстве дисперсий $\sigma_1 = \sigma_2$. Вычислим статистику Фишера

$$F = \frac{s_1^2}{s_2^2} = \frac{0.36^2}{0.32^2} = \frac{0.1286}{0.1024} = 1.265625.$$

Она имеет распределение Фишера с $n_1 - 1 = 24$ и $n_2 - 1 = 26$ степенями свободы. Возьмем уровень значимости $\alpha = 0.05$. Тогда $x_{1-\alpha/2} = 2.27$ и $x_{\alpha/2} = 0.44$. Так как

$$0.44 < 1.2656 < 2.57,$$

гипотеза о равенстве дисперсий принимается.

Проверка гипотезы о равенстве средних. Дисперсии равны. Вычислим тестовую статистику. Имеем

$$s_p^2 = \frac{25 \cdot 0.36^2 + 27 \cdot 0.32^2}{(25-1) + (27-1)} = \frac{3.24 + 2.76}{50} = 0.12,$$

$$s_p = 0.346,$$

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{1}{27} + \frac{1}{25}} = \sqrt{0.037 + 0.04} = \sqrt{0.077} = 0.277,$$

$$t = \frac{3.6 - 3.8}{0.346 \cdot 0.277} = -\frac{0.2}{0.096} = -2.0834.$$

Тестовая статистика имеет распределение Стьюдента с $n_1 - 1 + n_2 - 1 = 50$ степенями свободы. Поскольку число степеней свободы больше 30 для вычисления p-значения используем таблицы нормального распределения. Оно равно $\Phi(-2.08) = 0.0188 < 0.05 = \alpha$. Таким образом вторая технология лучше первой.

К решению этой задачи мы можем также применить критерий проверки гипотезы для разных дисперсий. Естественно качество критерия будет хуже, поскольку мы не знаем точного распределения тестовой статистики.

в). Проверка гипотезы о равенстве средних. Различные дисперсии. Тестовая статистика равна.

$$t = \frac{3.6 - 0.3.8}{\sqrt{\frac{0.36^2}{25} + \frac{0.32^2}{27}}} = -\frac{0.2}{\sqrt{0.00518 + 0.0038}} = -\frac{-0.2}{\sqrt{0.009}} = -\frac{0.2}{0.094} = -2.12.$$

Используя таблицы распределения Стьюдента с $\min\{n_1 - 1, n_2 - 1\} = 24$ степенями свободы, получаем, что p-значение приближенно равно 0.025. Следовательно гипотеза отклоняется.

Как мы видим, поскольку дисперсии мало отличаются друг от друга, мало отличаются друг от друга и значения тестовых статистик. Однако неточное знание распределения тестовой статистики в случае в) приводит к существенно завышенной величине p-значения.

Английские аналоги русских терминов в оценивании и проверке гипотез

оценка — estimate, estimator,
смещение — bias,
доверительный интервал — confidence interval,
уровень доверия — confidence level,
уровень значимости — significance level,
квадратичный риск — mean — squared error,
контрольная карта — control chart,
гипотеза — hypothesis,
проверка гипотез — hypothesis testing,
нулевая гипотеза — null hypothesis,
альтернатива — alternative hypothesis,
ошибка первого рода — type I error,
ошибка второго рода — type II error,
односторонний критерий — one-tailed test,
двусторонний критерий — two-tailed test,
критическое значение — critical value,
p-значение — probability value (p-value).