# Расчетное задание по математической статистике

Редько Анна, **выборка 18:** $\alpha = 2, \sigma^2 = 0.7, \varepsilon = 0.18$

24 мая 2020 г.

Задание 1. По числовой выборке объема 50 из нормальной совокупности с параметрами $\alpha$ и $\sigma^2$ (первая выборка) построить доверительные интервалы уровня доверия $1-\varepsilon$ для параметра:

**$\alpha$, если $\sigma^2$ известно:**

Дана выборка $X = (X_1, \ldots, X_n) \sim \Phi_{\alpha,\sigma^2}$. Тогда, из **теоремы о свойствах выборок из нормального распределения**,

$$\frac{\bar{X}-\alpha}{\sigma}\sqrt{n} \sim \Phi_{0,1}$$

Из таблиц стандартного нормального распределения находим число $q > 0$ такое, что $\Phi_{0,1}(-q) = \frac{\varepsilon}{2}$. Это значит, что

$$P\left(-q < \frac{\overline{X}-\alpha}{\sigma}\sqrt{n} < q\right) = \Phi_{0,1}(q) - \Phi_{0,1}(-q) = 1-\varepsilon$$

Это соотношение эквивалентно тому, что

$$P\left(\bar{X} - \frac{q\sigma}{\sqrt{n}} < \alpha < \bar{X} + \frac{q\sigma}{\sqrt{n}}\right) = 1-\varepsilon$$

После подсчетов(см. t1a.R) получили доверительный интервал: $(1.78278, 2.10006)$.

**$\alpha$, если $\sigma^2$ неизвестно:**

По **следствию из теоремы о свойствах выборок из нормального распределения** случайная величина

$$\frac{\bar{X}-\alpha}{S}\sqrt{n-1} \sim T_{n-1}$$

распределена по закону Стьюдента с $n-1$ степенью свободы. Подставим $S^2 = \frac{n-1}{n}S_0^2$, тогда

$$\frac{\bar{X}-\alpha}{S_0}\sqrt{n} \sim T_{n-1}$$

Из таблиц распределения $T_{n-1}$ находим q такое, что $T_{n-1}(-q) = \frac{\varepsilon}{2}$. Тогда

$$P\left(-q < \frac{\bar{X} - \alpha}{S_0}\sqrt{n} < q\right) = T_{n-1}(q) - T_{n-1}(-q) = 1 - \varepsilon,$$

Это соотношение эквивалентно тому, что

$$P\left(\bar{X} - \frac{qS_0}{\sqrt{n}} < \alpha < \bar{X} + \frac{qS_0}{\sqrt{n}}\right) = 1 - \varepsilon$$

Здесь несмещенную дисперсию вычисляем как

$$S_0 = \frac{1}{\sqrt{n-1}} \sum_{i=1}^{n} \left(X_i - \bar{X}\right).$$

После подсчетов(см. t1b.R) получили доверительный интервал: $(1.794088, 2.088752)$.

**$\sigma^2$, если $\alpha$ известно:**

Случайные величины $\frac{X_i - \alpha}{\sigma}$, $i = 1, \ldots, n$, независимы, и имеют стандартное нормальное распределение, поэтому

$$\sum_{i=1}^{n} \left(\frac{X_i - \alpha}{\sigma}\right)^2 \sim \chi^2_n.$$

Из таблиц распределения $\chi^2_n$ находим $q_1$ и $q_2$ такие, что $\chi^2_n(q_1) = \frac{\varepsilon}{2}$, $\chi^2_n(q_2) = 1 - \frac{\varepsilon}{2}$. Тогда

$$P\left(q_1 < \sum_{i=1}^{n} \frac{(X_i - \alpha)^2}{\sigma^2} < q_2\right) = \chi^2_n(q_2) - \chi^2_n(q_1) = 1 - \varepsilon$$

Это соотношение эквивалентно тому, что

$$P\left(\frac{\sum_{i=1}^{n}(X_i - \alpha)^2}{q_2} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i - \alpha)^2}{q_1}\right) = 1 - \varepsilon$$

После подсчетов(см. t1c.R), получили, что доверительный интервал: $(0.7773202, 0.4528119)$.

**$\sigma^2$, если $\alpha$ неизвестно:**

Из **теоремы о свойствах выборок из нормального распределения**, $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$. Из таблиц распределения $\chi^2_{n-1}$ находим $q_1$ и $q_2$ такие, что $\chi^2_{n-1}(q_1) = \frac{\varepsilon}{2}$ и $\chi^2_{n-1}(q_2) = 1 - \frac{\varepsilon}{2}$. Тогда

$$P\left(q_1 < \frac{nS^2}{\sigma^2} < q_2\right) = \chi^2_{n-1}(q_2) - \chi^2_{n-1}(q_1) = 1 - \varepsilon.$$

Это соотношение эквивалентно тому, что

$$P\left(\frac{nS^2}{q_2} < \sigma^2 < \frac{nS^2}{q_1}\right) = 1 - \varepsilon.$$

После подсчетов(см. t1d.R), получили, что доверительный интервал: $(0.458274, 0.7910913)$.

Задание 2. По данным числовым наблюдениям (вторая выборка объема 30) проверить основную гипотезу о равномерности распределения с помощью а) критерия Колмогорова, б) критерия $\chi^2$ (асимптотического размера $\varepsilon$). Построить график эмпирической функции распределения. Найти реально достигнутый уровень значимости.

**используя критерий Колмогорова:**

Функция распределения K(y) называется функцией Колмогорова, она абсолютно непрерывна; для нахождения ее значений имеются таблицы.

Перейдем к построению критерия.

Пусть $X \sim F$ и проверяются гипотезы $H_1 : F = U_{0,1}$ против $H_2 : F \neq U_{0,1}$, где $U_{0,1}$ непрерывна.

Наша задача: построить асимптотический критерий уровня $1 - \varepsilon$.

Для начала вычислим величину $D_n$ в предположении, что верна гипотеза $H_1$, т. е. $F = U_{0,1}$:

$$D_n = \sup_t |F_n^*(t) - U_{0,1}(t)|$$

$$D_n = 0.887$$

$$\sqrt(n) D_n = 4.858299$$

Тут $F_n^*$ - эмперическая функция распределения, $U_{0,1}$ - теоретическая функция распределения. В силу теоремы Колмогорова, при больших $n$ функция распределения случайной величины $\sqrt{n}D_n$ мало отличается от $K(y)$, поэтому заранее по таблицам функции Колмогорова мы можем найти такое число q > 0, что $K(q) = 1 - \varepsilon$. Нашли, что $q = 1.1$. Следовательно, если верна $H_1$, то $P_1\left(\sqrt{n}D_n < q\right) \simeq K(q) = 1 - \varepsilon$. Поэтому мы будем отвергать гипотезу $H_1$, если окажется, что $\sqrt{n}D_n \geqslant q$ , т. е. если расхождение между эмпирической и гипотетической функциями распределения достаточно велико. В нашем случае это именно так, ведь $4.858299 > 1.1$, гипотезу $H_1$ мы отвергаем. Ясно, что при этом

$$\beta_1 = P_1\left(\sqrt{n}D_n \geqslant q)\right) = 1 - P_1\left(\sqrt{n}D_n < q\right) \simeq 1 - K(q) = \varepsilon = 0.18$$

Критическое множество для построенного критерия:

$$K = (X_1, \ldots, X_n) \in \mathbb{R}^n : \sqrt{n}D_n \geqslant q.$$

Достигаемый уровень значимости критерия Колмогорова равен:

$$\alpha^* = 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2nD_n^2} < 0.1$$

.

**используя критерий $\chi^2$:**

Пусть $X \sim F$ и проверяются гипотезы $H_1 : F = U_{0,1}$ против $H_2 : F \neq U_{0,1}$. По-прежнему наша задача состоит в построении асимптотического критерия уровня $1-\varepsilon$. В предположении, что $X \sim U_{0,1}$, разобьем область возможных значений $X_1 = [0, 1)$ на k непересекающихся промежутков(здесь k ищем по формуле Стеджеса $k = \lfloor log_2 30 \rfloor + 1 = 4 + 1 = 5$):

$$P_1(X_1 \in \Delta_1 \cup \cdots \cup \Delta_k) = 1,$$

где $\Delta_i$ имеет вид $\Delta_i = [a_i; b_i)\,, i = 1, \ldots, k$.

Пусть $\nu_i$ - число наблюдений, попавших в $\Delta_i, i = 1, \ldots, k, \nu_1 + \cdots + \nu_k = n$.

Обозначим также

$$p_i = P_1\left(X_1 \in \Delta_i\right) = U_{0,1}(b_i) - U_{0,1}(a_i), i = 1, \ldots, k.$$

Из закона больших чисел следует, что

$$\frac{\nu_i}{n} \xrightarrow{\mathbb{P}} p_i, n \to \infty,$$

при каждом $i$, если верна $H_1$. В качестве меры близости совокупностей $\nu_1/n, \ldots, \nu_k/n$ и $p_1, \ldots, p_k$ предлагается использовать величину

$$\Psi_n = n \sum_{i=1}^{k} \frac{1}{p_i} \left(\frac{\nu_i}{n} - p_i\right)^2 = \sum_{i=1}^{k} \frac{\left(\nu_i - np_i\right)^2}{np_i}.$$

Т.к. $p_i = 0.2\ \forall i$, можем представить $\Psi_n$ как

$$\Psi_n = \frac{1}{np_i} \sum_{i=1}^{k} \left(\nu_i - np_i\right)^2.$$

$$\Psi_n = 4.666667$$

**Теорема Пирсона.** Если $0 < p_i < 1$ при всех $i = 1, \ldots, k$, то для любого $y > 0$

$$P_1(\Psi_i < y) \to \chi^2_{k-1}(y), n \to \infty.$$

Займемся построением критерия. Найдем число $q$ такое, что $\chi^2_{k-1}(q) = 1 - \varepsilon$. Получим $q = 6.268116$.

Если верна гипотеза $H_1$, то с вероятностью, близкой к $1 - \varepsilon$, значение случайной величины $\Psi_n$ должно быть меньше $q$. Поэтому мы отвергаем гипотезу, если $\Psi_n \geqslant q$, и принимаем ее в противном случае. В нашем случае $\Psi_n < q$, а значит гипотезу $H_1$ мы принимаем.

Это значит, что мы принимаем $H_1$, если нет явного противоречия этой гипотезы с наблюденными значениями. Критическое множество в данном случае:

$$K = (X_1, \ldots, X_n) : \Psi_n \geqslant q.$$

Для вероятности ошибки первого рода имеем

$$\beta_1 = P_1\left(\Psi_n \geqslant q\right) = 1 - P_1\left(\Psi_n < q\right) \simeq 1 - \chi^2_{k-1}(q) = \varepsilon = 0.18$$

Достигаемый уровень значимости критерия $\chi^2$ равен:

$$\varepsilon^* = 0.323239.$$

Задание 3. По данным двум выборкам из нормальных совокупностей (первые 20 и следующие 30 элементов первой выборки) проверить, с помощью критериев размера $\varepsilon$, гипотезу:

**о совпадении дисперсий при неизвестных средних:**

Имеем две независимые выборки

$$X = (X_1, \ldots, X_n) \sim \Phi_{\alpha_1, \sigma_1^2}$$

$$Y = (Y_1, \ldots, Y_m) \sim \Phi_{\alpha_2, \sigma_2^2}$$

В этом пункте проверяем гипотезу $H_1 : \sigma_1^2 = \sigma_2^2$ против $H_2 : \sigma_1^2 \neq \sigma_2^2$. По условию $\varepsilon = 0.18$ и пусть

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2,$$

$$\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i,$$

$$S_Y^2 = \frac{1}{m}\sum_{i=1}^{m}\left(Y_i - \bar{Y}\right)^2,$$

По теореме о свойствах выборок из нормального распределения

$$\frac{nS_X^2}{\sigma_1^2} \sim \chi^2_{n-1},$$

$$\frac{mS_Y^2}{\sigma_2^2} \sim \chi^2_{m-1},$$

причем эти случайные величины независимы, поскольку построены по независимым выборкам. Из них можно построить случайную величину, имеющую распределение Фишера:

$$\frac{1}{n-1}\frac{nS_X^2}{\sigma_1^2} : \frac{1}{m-1}\frac{mS_{Y^2}}{\sigma_2^2} = \frac{n(m-1)\sigma_2^2 S_x^2}{m(n-1)\sigma_1^2 S_y^2} \sim F_{n-1,m-1}.$$

Если верна гипотеза $H_1$, т.е. $\sigma_1^2 = \sigma_2^2$, то

$$\eta = \frac{n(m-1)S_X^2}{m(n-1)S_Y^2} \sim F_{n-1,m-1}$$

$$\eta = 1$$

С помощью таблиц распределения $F_{n-1,m-1}$ можно найти числа $q_1$ и $q_2$ такие, что $F_{n-1,m-1}(q_1) = \varepsilon/2, F_{n-1,m-1}(q_2) = 1 - \varepsilon/2$. Получили, что $q_1 = 0.5523387$, $q_2 = 1.726676$. Тогда

$$P_1 \left(q_1 < \eta < q_2\right) = F_{n-1,m-1}(q_2) - F_{n-1,m-1}(q_1) = 1 - \varepsilon = 0.82.$$

Поэтому логично отвергать $H_1$, если $\eta \notin (q_1, q_2)$; вероятность такого события равна в точности $\varepsilon$, если верна $H_1$.

В нашем случае $1 \in (0.5523387, 1.726676)$, так что гипотезу $H_1$ мы принимаем.

Здесь

$$K = (X_1, \dots, X_n, Y_1, \dots, Y_m) : \eta \notin (q_1, q_2).$$

**о совпадении средних, если известно, что неизвестные дисперсии совпадают:**

Имеем две независимые выборки

$$X = (X_1, \dots, X_n)$$

$$Y = (Y_1, \dots, Y_m)$$

По условию, дисперсии совпадают: $\sigma_1^2 = \sigma_2^2 = \sigma^2$, при этом $\sigma^2$ неизвестна.

Необходимо проверить гипотезу $H_1 : \alpha_1 = \alpha_2$ против $H_2 : \alpha_1 \neq \alpha_2$.

Воспользуемся распределением Стьюдента. В силу того, что $\bar{X}$ и $\bar{Y}$ независимы, и

$$\bar{X} \sim \Phi_{\alpha_1, \sigma^2/n}$$

$$\bar{Y} \sim \Phi_{\alpha_2, \sigma^2/m},$$

имеем

$$\bar{X} - \bar{Y} \sim \Phi_{\alpha_1 - \alpha_2, \sigma^2(1/n + 1/m)}$$

После стандартизации:

$$\frac{\bar{X} - \bar{Y} - (\alpha_1 - \alpha_2)}{\sqrt{\sigma^2 \left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} \sim \Phi_{0,1}$$

Далее, по свойству распределения хи-квадрат,

$$\frac{nS_X^2}{\sigma^2} + \frac{mS_Y^2}{\sigma^2} \sim \chi^2_{n+m-2},$$

эта случайная величина не зависит от $\bar{X} - \bar{Y}$. Таким образом:

$$\frac{\bar{X} - \bar{Y} - (\alpha_1 - \alpha_2)}{\sigma \sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} : \sqrt{\frac{1}{n+m-2} \frac{nS_X^2 + mS_Y^2}{\sigma^2}} \sim T_{n+m-2}.$$

Если верна гипотеза $H_1$, то $\alpha_1 - \alpha_2 = 0$

$$\psi = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}\sqrt{\dfrac{nS_X^2 + mS_Y^2}{n+m-2}}} \sim T_{n+m-2}.$$

$$\psi = 1.754003$$

Из таблиц распределения $T_{n+m-2}$ находим $q$ такое, что $T_{n+m-2}(-q) = \dfrac{\varepsilon}{2}$.
Нашли, что $-q = -1.360585$.

Тогда

$$P_1(-q < \psi < q) = T_{n+m-2}(q) - T_{n+m-2}(-q) = 1 - \varepsilon.$$

В нашем случае, получаем, что $1.754003 \notin (-1.360585, 1.360585)$. Гипотеза $H_1$ не принимается.

Следовательно, выбрав

$$K = (X_1, \ldots, X_n, Y_1, \ldots, Y_m) : |\psi| \geqslant q,$$

мы будем иметь

$$\beta_1 = P_1((X_1, \ldots, X_n, Y_1, \ldots, Y_m) \in K) = \varepsilon = 0.18.$$