Проверка гипотез однородности

Н.П. Алексеева,

СПбГУ, математико-механический факультет

2020 г.

Выборочные моменты

Выборка случайных наблюдений как дискретная величина с выборочным распределением

$$\begin{array}{ccc} x_1 & \dots & x_n \\ \frac{1}{n} & \dots & \frac{1}{n} \end{array}$$

Выборочное среднее и выборочные смещенная и несмещенная дисперсии имеют соответственно вид

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2; \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2;$$

начальные и центральные выборочные моменты

$$a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k; \quad m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k.$$

Ошибка среднего $m = \frac{S}{\sqrt{n}}$

Выборочные мода и медиана

Выборочная мода унимодального распределения

элемент выборки с наибольшей частотой

Чему равна мода в выборке: 0, 0, 0, 0, 1, 1, 1, 2, 2, 3, 3, 3, 3, 3?

Выборочная медиана

число, которое делит вариационный ряд на две части, содержащие одинаковое число элементов

Если $n = 2k$, то $med = (x^{(k)} + x^{(k+1)})=2:$ Если $n = 2k + 1$, то $med = x^{(k+1)}$.

Пример

Масса до диеты: 68, 80, 92, 81, 70, 79, 78, 66, 57, 76, $\quad n = 10.$

Выборочное среднее

$$x = \sum_{j=1}^{n} x_j = 74{:}9$$

Стандартное отклонение S и ошибка среднего $m = S/\sqrt{n}$

$$\sum_{j=1}^{10} x_j^2 = 56655; \quad \sum_{j=1}^{10} x_j^2 \quad nx^2 = 854{:}1$$

$$S^2 = \frac{1}{n \quad 1} \sum_{j=1}^{10} x_j^2 \quad nx^2 = 94{:}9; \quad S = 9{:}74; \quad m = 3{:}08$$

Чему равна медиана? Упорядочиваем:
57, 66, 68, 70, 76, 78, 79, 80, 81, 92.

В среднем масса до диеты 74:9 $\quad$ 3:08(10),
медиана, квартили, минимум-максимум: 77=68 $\quad$ 80=57 $\quad$ 92.

Интервал, накрывающий истинное значение параметра распределения с заданной вероятностью P называется P 100% - доверительным интервалом.

$$x \quad N \quad ; \frac{}{\sqrt{n}} \qquad (\;) \quad z = \frac{x \qquad}{\sqrt{n}} \quad N(0;1):$$

$$P\{ \; z_{1 \frac{}{2}} < \frac{x \quad}{\sqrt{n}} \quad z_{1 \frac{}{2}} \} = 1$$

$$P \quad x \quad z_{1 \frac{}{2}} \frac{}{\sqrt{n}} < \qquad x + z_{1 \frac{}{2}} \frac{}{\sqrt{n}} \quad = 1 \qquad :$$

При n > 40 параметр   может быть заменен на выборочное стандартное отклонение S.

Доверительный интервал для μ при небольших выборках

$$P\left(\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

$$P\left(\bar{x} - t^{n-1}_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t^{n-1}_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Например, $z_{1-\frac{\alpha}{2}} = 1.96$, $t^{9}_{1-\frac{\alpha}{2}} = 1.83$, $\alpha = 0.05$.

В первом случае 95%-доверительный интервал $(69.63 - 79.77)$, во втором более широкий $(69.05 - 80.34)$.

Для самостоятельной работы

Масса после диеты: 60, 84, 87, 79, 74, 71, 72, 67, 57, 70, $n = 10$.

Среднее

$\bar{x} = ?$

Стандартное отклонение S и ошибка среднего $m = S/\sqrt{n}$

$\sum_{j=1}^{10} x_j^2 = ?$

$\sum_{j=1}^{10} x_j^2 - n\bar{x}^2 = ?$

$S^2 = \frac{1}{n-1}\left(\sum_{j=1}^{10} x_j^2 - n\bar{x}^2\right) = ?$

$S = ?$

$m = ?$

Медиана

57, 60, 67, 70, 71, 72, 74, 79, 84, 87

Для самостоятельной работы

Масса после диеты: 60, 84, 87, 79, 74, 71, 72, 67, 57, 70, $n = 10$.

Среднее

$\bar{x} = 72.1$

Стандартное отклонение S и ошибка среднего $m = S/\sqrt{n}$

$\sum_{j=1}^{10} x_j^2 = 52805$

$\sum_{j=1}^{10} x_j^2 - n\bar{x}^2 = 820.9$

$S^2 = \frac{1}{n-1}\left(\sum_{j=1}^{10} x_j^2 - n\bar{x}^2\right) = 91.21$

$S = 9.55$

$m = 3.02$

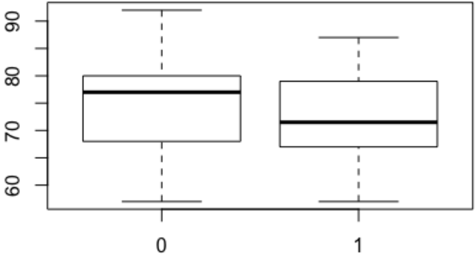
Медиана 71.5

57, 60, 67, 79, 71, 72, 74, 79, 84, 87

Проверка однородности двух зависимых выборок

$(x_i; y_i); i = 1; 2; \dots; n;$ парная выборка одних и тех же наблюдений при разных условиях

	1	2	3	4	5	6	7	8	9	10
Масса до диеты	68	80	92	81	70	79	78	66	57	76
Масса после диеты	60	84	87	79	74	71	72	67	57	70
Потерянный вес	8	4	5	2	4	8	6	1	0	6



Требуется проверить значимость изменений.

до диеты

В среднем масса 74:9   3:08(10),
медиана, квартили, минимум-максимум: 77=68   80=57   92.

после диеты

В среднем масса 72:1   3:02(10),
медиана, квартили, минимум-максимум: 71:5=67   79=57   87.

Вводим $z_i = y_i \quad x_i$ и проверяем гипотезу $H_0 : \quad _z = 0$.

Пусть $z_1, \dots, z_n$ — выборка из нормальной совокупности со средним и дисперсией ^{2}. Тогда отношение

$$= \frac{z}{\sqrt{m_2}}\sqrt{n \quad 1} = \frac{z}{s}\sqrt{n} \qquad (1)$$

имеет распределение Стьюдента с $n \quad 1$ степенью свободы.

Полагая ${}_1 = \frac{z}{=\sqrt{n}} \quad N(0, 1)$ и $\sum_{i=2} {}_i^2 = \frac{nm_2}{^2} \quad {}^2(n \quad 1)$,

получим

$$= s\frac{1}{\sqrt{\frac{1}{n \quad 1}\sum_{k=2} {}_k^2}} = \frac{z}{\frac{}{\sqrt{n}}\sqrt{\frac{1}{n \quad 1}\frac{nm_2}{^2}}} = \frac{z}{\sqrt{m_2}}\sqrt{n \quad 1}.$$

Так как $m_2 = \frac{n \quad 1}{n}s^2$, $\sqrt{m_2} = \frac{\sqrt{n \quad 1}}{\sqrt{n}}s$,

$$= \frac{x}{s}\sqrt{n} \qquad T(n \quad 1). \qquad (2)$$

Потерянный вес: 8; 4; 5; 2; 4; 8; 6; 1; 0; 6

$z = 2.6$

$S = 4.64758$

$n = 10$

$t = \frac{z}{S}\sqrt{n} = \frac{2.6}{4.65}\sqrt{10} = 1.769076$

$P\{ \quad > 1.769076\} = 0.05533384 =$

Гипотеза о том, что средняя разность между выборками равна нулю отвергается с уровнем значимости $\quad > \quad$.