Пробный вариант

1. (2 балла) Восемь спортсменов, показавших лучшие результаты на соревнованиях международного уровня в беге на 110 метров с барьерами, приняли участие в забеге на 100 метров без барьеров. В забеге на 100 метров спортсмены заняли дорожки, номера которых совпадали с номерами мест, занятыми спортсменами в беге с барьерами. Результаты забега (по дорожкам) оказались следующими: 11,2; 10,9; 12,1; 11,8; 12,2; 11,6; 11,0; 11,9. Можно ли считать (на уровне значимости 0.05), что имеется зависимость между результатами бега с барьерами и «гладкого» бега? Полностью опишите процедуру проверки соответствующей гипотезы. Какие ещё известные Вам критерии можно применить для решения данной задачи?

X_i	1	2	3	4	5	6	7	8
R_{X_i}	1	2	3	4	5	6	7	8
Y_i	11,2	10,9	12,1	11,8	12,2	11,6	11,0	11,9
R_{Y_i}	3	1	7	5	8	4	2	6
$d = R_{X_i} - R_{Y_i}$	-2	1	-4	-1	-3	2	5	2
$d^2 = (R_{X_i} - R_{Y_i})^2$	4	1	16	1	9	4	25	4

$\sum_{i=1}^{n} d_i = 64$

$\alpha = 0,05$

$n = 8$

$H_0: \rho_S = 0$ (независимы)

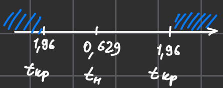
$H_a: \rho_S \neq 0$ (зависимы)

применим ранговый критерий Спирмена

$$\rho_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 64}{8(8^2-1)} = 1 - \frac{384}{504} \approx 0,238$$

$$t_H = \sqrt{n-1}\,\hat{\rho}_S = \sqrt{7} \cdot 0,238 \approx 0,629 \sim N(0;1)$$

$$t_{kp} = z_{1-\frac{\alpha}{2}} = z_{0,975} = 1,96$$



$\Downarrow$

Нет осн. откл. H_0 $\Rightarrow$ между результатами бега зависимости нет

можно также применить критерии Спирмена и Кендалла

2. (3 балла) Случайным образом были выбраны 9 человек одного возраста, проживающих в одном городе, но имеющих различный уровень образования. Доходы (в тысячах рублей) в группе с неполным средним образованием составили – 37, 19, 26, 42, а в группе со средним специальным образованием – 47, 39, 52, 41, 42. Можно ли считать, что работники со средним специальным образованием имеют в среднем более высокие доходы, чем имеющие неполное среднее образование? Уровень значимости принять равным 0.05. Решите задачу, используя следующие предположения: а) все наблюдения имеют нормальное распределение с одинаковыми дисперсиями; б) наблюдения имеют некоторые неизвестные непрерывные распределения, которые могут различаться только математическим ожиданием. Полностью опишите процедуру проверки соответствующих гипотез. Какие ещё известные Вам критерии можно применить для решения данной задачи? Опишите сильные и слабые стороны рассмотренных критериев.

$N = 9$

$X_i = 37, 19, 26, 42$

$n = 4$

$Y_i = 47, 39, 52, 41, 42$

$m = 5$

$\alpha = 0,05$

а) $X \sim N(m_x, \sigma^2)$

$Y \sim N(m_y, \sigma^2)$

$H_0: m_y = m_x$

$H_a: m_y > m_x$

Применим критерий Стьюдента

$$T = \frac{\bar{Y} - \bar{X}}{S_{xy}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2)$$

$$S_{xy}^2 = \frac{(n-1)\cdot S_x^2 + (m-1)\cdot S_y^2}{n+m-2}$$

$$\bar{X} = \frac{37+19+26+42}{4} = 31$$

$$\bar{Y} = \frac{47+39+52+41+42}{5} = 44,2$$

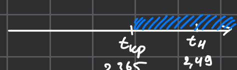
$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(37-31)^2 + (19-31)^2 + (26-31)^2 + (42-31)^2}{3} \approx 108,67$$

$$\hat{S}_y^2 = \frac{\sum (y_i - \bar{y})^2}{m-1} = 27,7$$

$$S_{xy}^2 = \sqrt{\frac{3 \cdot 108,67 + 4 \cdot 27,7}{7}} = 62,4$$

$$t_H = \frac{44,2 - 31}{\sqrt{62,4}\sqrt{\frac{1}{4} + \frac{1}{5}}} = \frac{11\sqrt{78}}{39} \approx 2,49$$

$$t_{kp} = t_{0,95}(n+m-2) = t_{0,95}(4+5-2) = t_{0,95}(7) = 2,365$$



$\Downarrow$

$\cancel{H_0}$ $\Rightarrow$ группа со сред. спец. имеют более высокие доходы б) $H_0: m_y = m_x$

$H_a: m_y > m_x$

x	x	x	y	y	x	y	y	y
19	26	37	39	41	42	42	47	52
1	2	3	4	5	6,5	6,5	8	9

Применим критерий Вилкоксона

$W_{n,m} = \sum_{j=1}^{m} R_j = 4+5+6,5+8+9 = 32,5$

$E(W_{n,m}) = \frac{m}{2}(N+1) = \frac{5}{2}(9+1) = \frac{5}{2}\cdot 10 = 25$

$D(W_{n,m}) = \frac{n\cdot m}{12}(N+1) = \frac{4\cdot 5}{12}\cdot 10 = 16,67$

$D_{св}(W_{n,m}) = D(W_{n,m}) - \frac{m\cdot n\cdot \sum_{k=1}^{l}(t_k(t_k^2-1))}{12N(N-1)} = 16,67 - \frac{4\cdot 5\cdot 2(4-1)}{12\cdot 9\cdot 8} = \frac{7439}{450} \approx 16,531$

$t_н = \frac{W_{m,n} - E(W_{m,n})}{\sqrt{D(W_{m,n})}} = \frac{32,5-25}{\sqrt{16,531}} \approx 1,84464 \sim N(0;1)$

$t_{кр} = Z_{1-\alpha} = Z_{0,95} = Z_{0,45} = 1,64$

$\Downarrow$

$\cancel{H_0} \Rightarrow$ группа со сред. спец. имеют более высокие доходы

3. (3 балла) Имеются данные о себестоимости Y (в у.е.) экземпляра книги в зависимости от тиража X (в тыс. экземпляров). Данные представлены в таблице:

тираж	1	2	3	4	5
себестоимость	6	5	4	4	3

Предполагается, что справедлива модель вида

, где вектор .

1) построить МНК-оценки параметров a и b и линию регрессии;
2) найти оценку дисперсии ;
3) проверить гипотезу ;
4) проверить гипотезу ;
5) построить точечную и интервальную (уровня надёжности 0.95) оценку для себестоимости, если тираж X=6;
6) найти коэффициент детерминации данной модели и скорректированный коэффициент детерминации.

$n = 5$

$x = 1,2,3,4,5$

$y = 6,5,4,4,3$

1) $\hat{y} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$ $\quad \hat{x} = \bar{y} - \hat{y}\bar{x}$

$\bar{x} = \frac{1+...+5}{5} = 3$

$\bar{y} = \frac{6+5+4+4+3}{5} = 4,4$

$\sum(x_i-\bar{x})(y_i-\bar{y}) = ... = -7$

$\sum(x_i-\bar{x})^2 = 10$

$\hat{b} = -0,7 \quad \hat{a} = 6,5$

Линия регресса: $\hat{y} = 6,5 - 0,7x$

2) $\hat{\sigma}_\varepsilon^2 = \frac{RSS}{n-p-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-p-1}$

$\hat{a} = 6,5 \quad \hat{b} = -0,7 \quad \hat{Y}_i = 6,5 - 0,7X_i$

i	y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$
1	6	5,8	0,2	0,04
2	5	5,1	-0,1	0,01
3	4	4,4	-0,4	0,16
4	4	3,7	0,3	0,09
5	3	3	0	0

$RSS = 0,04+0,01+0,16+0,09+0 = 0,3$

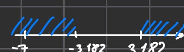
$\hat{\sigma}_\varepsilon^2 = \frac{0,3}{3} = 0,1$

3-4) $H_0: b = 0$

$H_a: b \neq 0$

$T = \frac{\hat{b} - b}{\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\sum(x_i-\bar{x})^2}}} \sim t(n-p-1)$ $\quad t_н = \frac{-0,7-0}{\sqrt{\frac{0,1}{10}}} = -7$

$t_{кр}(n-p-1) = t_{1-\frac{\alpha}{2}}(5-1-1) = t_{0,975}(3) = 3,182$



$\Downarrow$

$\cancel{H_0} \Rightarrow b \neq 0$

5) Точечная оценка:

$x = 6$

$\hat{y} = 6{,}5 - 0{,}7x = 6{,}5 - 0{,}7 \cdot 6 = 2{,}3$

Интервальный оценка:

$\hat{y}(x_0) \pm t_{\frac{\alpha}{2}}(n-2) \cdot SE_{pred}$

$SE_{pred} = \sqrt{\hat{\sigma}_\varepsilon^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)}$

$2{,}3 \pm 3{,}182 \cdot 0{,}458 \approx 2{,}3 \pm 1{,}46 \Rightarrow [0{,}84; 3{,}76]$

6) $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{0{,}3}{5{,}2} \approx 0{,}94$

$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-p-1} = 0{,}92$

4. (2 балла) Руководители телеканала интересуются отношением зрителей к новому телесериалу. Из 100 опрошенных мужчин 30 относятся к сериалу положительно, 40 – отрицательно и 30 – безразлично. Из 150 опрошенных женщин 80 относятся положительно, 50 – отрицательно и 20 –безразлично. Можно ли считать (на уровне значимости 0,05), что отношение к сериалу зависит от пола зрителя? Если зависимость присутствует, то оцените силу этой связи. Полностью опишите процедуру проверки соответствующей гипотезы.

пол (m) \ мнение (k)	+	−	:	
муж	30	40	30	100
жен	80	50	20	150
	110	90	50	250

$\alpha = 0{,}05$

$H_0: P_{ij} = P_{i\cdot} \cdot P_{\cdot j}$ (независимы)

$H_a: \exists_{i,j}: P_{ij} \neq P_{i\cdot} \cdot P_{\cdot j}$ (зависимы)

Применим критерий Пирсона (χ^2)

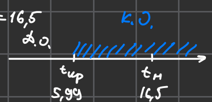
$\chi^2 = n\left(\sum_{i=1}^{m}\sum_{j=1}^{k} \frac{(n_{ij})^2}{n_{i\cdot}\, n_{\cdot j}} - 1\right) \sim \chi^2_{1-\alpha}((k-1)(m-1))$

$\boxed{m = k = 2: \ \chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\cdot}\, n_{2\cdot}\, n_{\cdot 1}\, n_{\cdot 2}} \sim \chi^2(1)}$

$t_{кр} = \chi^2_{1-\alpha}(2) = 5{,}99$

$t_н = n\left(\sum_{i=1}^{m}\sum_{j=1}^{k} \frac{n_{ij}}{n_{i\cdot}\, n_{\cdot j}} - 1\right) = 250\left(\frac{30^2}{100 \cdot 110} + \frac{40^2}{100 \cdot 90} + \frac{30^2}{100 \cdot 50} + \frac{80^2}{150 \cdot 110} + \frac{50^2}{150 \cdot 90} + \frac{20^2}{150 \cdot 50} - 1\right) =$

$= 16{,}5$



$\Downarrow$

$\cancel{H_0} \Rightarrow$ отношение к сериалу зависит от пола зрителя

для оценки зависимости можем использовать коэфф. Пирсона или коэфф. Крамера

↓ для количественных ↓ для номинальных

воспользуемся Крамера

$C = \sqrt{\frac{t_н}{n \cdot \min(m-1; k-1)}} \in [0,1]$

$C = \sqrt{\frac{16{,}5}{250 \cdot 1}} = \sqrt{\frac{33}{500}} = 0{,}257 \in [0; 0{,}3]$ – связь слабая, близка к умеренной