2025 Пробный вариант экзаменационного билета

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?
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.
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.

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.

КОММЕНТАРИИ

В данном варианте отсутствуют (но в других вариантах могут быть) задачи на следующие темы:

1) Проверка гипотезы об однородности двух выборок против альтернативы масштаба;
2) Задача однофакторного дисперсионного анализа;
3) Задача на меры прогноза в номинальной шкале
4) Построение асимптотического доверительного интервала для коэффициента корреляции показателей, имеющих гауссовское распределение;
5) Задача о согласованности экспертной группы;
6) Построение оценок и проверка гипотез в моделях множественной линейной регрессии (в качестве дополнительной информации будут даны некоторые промежуточные вычисления);
7) Частные коэффициенты корреляции;
8) Критерий Чоу;
9) Фиктивные переменные.

Несмотря на то, что в экзаменационных задачах будут заданы выборки малых объёмов, разрешается (из-за отсутствия нужного набора таблиц) использовать асимптотический подход.

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?

стр 142 моего файлика (похожая задача)

1) Ставим гипотезы $\alpha = 0{,}05$ использовали ранговый критерий Спирмена

$H_0: \rho_s = 0$ (независимы) $n = 8$

$H_a: \rho_s \neq 0$ (зависимы)

номер участника (i)	1	2	3	4	5	6	7	8
место начальное (X_i)	1	2	3	4	5	6	7	8
место конечное (Y_i)	11,2	10,9	12,1	11,8	12,2	11,6	11	11,9
ранг Y_i (R_i)	3	1	7	5	8	4	2	6
$X_i - R_i$	−2	1	−4	−1	−3	2	5	2
$(X_i - R_i)^2$	4	1	16	1	9	4	25	4

← Тут у нас ранг X_i совпадает с X_i

$$\sum \underbrace{(X_i - R_i)^2}_{d_i^2} = 4+1+16+1+9+4+25+4 = 64$$

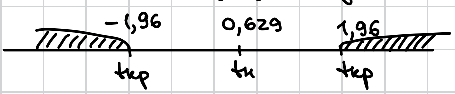
5 21 22 31 35 60

$$\hat{\rho}_s = 1 - \frac{6\sum d_i^2}{n^3 - n} = 1 - \frac{6\cdot \cancel{64}\,8}{\cancel{8}(8^2-1)} = 1 - \frac{48}{63} = 1 - \frac{16}{21} = \frac{5}{21} \approx 0{,}238$$

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

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

Н0Но ⇒ независима



Для решения этой задачи мы можем также использовать Кендалла или Пирсона

Ответ: нет, нельзя утверждать, что имеется зависимость. В этой задаче можно также применить критерии Пирсона и Кендалла.

Решение 2 (через Кендалла)

$X = 1, 2, 3, 4, 5, 6, 7, 8$

$Y = 11{,}2\ \ 10{,}9\ \ 12{,}1\ \ 11{,}8\ \ 12{,}2\ \ 11{,}6\ \ 11\ \ 11{,}9$

для каждой $i < j$

для $i=1$: $C=5$ $D=2$ C – конкордантных

для $i=2$: $C=6$ $D=0$ D – дискордантных

для $i=3$: $C=1$ $D=4$

для $i=4$: $C=2$ $D=2$

для $i=5$: $C=0$ $D=3$

для $i=6$: $C=1$ $D=1$

для $i=7$: $C=1$ $D=0$

$\sum C = 5+6+1+2+1+1 = 16$ $\sum D = 2+4+2+3+1 = 12$

В сумме $\sum C + \sum D$ должно быть как всего пар

Для 8 всего пар $\frac{8\cdot 7}{2} = 28$. $\sum C + \sum D = 16 + 12 = 28$

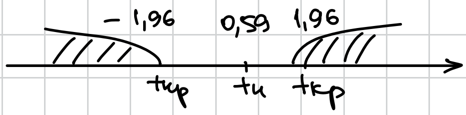
У нас нет связок

$\hat{\tau} = 1 - \frac{4Q}{n(n-1)} = 1 - \frac{4\cdot 12}{8\cdot 7} = \frac{1}{7} \approx 0,14$

H_0 ок $\Rightarrow$ независим

$t_{н} = \frac{3}{2}\sqrt{n}\,\hat{\tau} = \frac{3}{2}\sqrt{8}\cdot 0,14 \approx 0,59$

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



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.

стр 143 моего большого файла (похожие)

а) Т.к. выборка имеет нормальное распределение можем применить кр. Стьюдента (группе 237 не нужно проверять равенство дисперсий)

Группа	n	доход в тыс. руб
A (неполное среднее)	$n_1=4$	37, 19, 26, 42
B (среднее специальное)	$n_2=5$	47, 39, 52, 41, 42

$H_0: \mu_A = \mu_B$

$H_a: \mu_B > \mu_A$

$$\bar{x}_A = \frac{37+19+26+42}{4} = \frac{124}{4} = 31$$

$$\bar{x}_B = \frac{47+39+52+41+42}{5} = \frac{221}{5} = 44{,}2$$

$$\sigma_A^2 = \frac{\sum (x_{Ai}-31)^2}{n_1-1} = \frac{(37-31)^2+(19-31)^2+(26-31)^2+(42-31)^2}{4-1} = \frac{6^2+(-12)^2+(-5)^2+11^2}{3} = \frac{36+144+25+121}{3} =$$

$$= \frac{326}{3} \approx 108{,}67$$

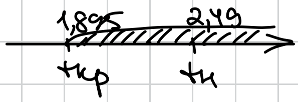
$$\sigma_B^2 = \frac{\sum (x_{Bi}-44{,}2)}{5-1} = \frac{(47-44{,}2)^2+(39-44{,}2)^2+(52-44{,}2)^2+(41-44{,}2)^2+(42-44{,}2)^2}{4} = \frac{(2{,}8)^2+(-5{,}2)^2+(7{,}8)^2+(-3{,}2)^2+(-2{,}2)^2}{4} = \frac{277}{10} = 27{,}7$$

$$s_p^2 = \frac{(n_A-1)\sigma_1^2+(n_B-1)\sigma_2^2}{n_A+n_B-2} = \frac{(4-1)\cdot 108{,}67+(5-1)\cdot 27{,}7}{4+5-2} = \frac{326{,}01+110{,}8}{7} \approx 62{,}4$$

$$t_H = \frac{\bar{x}_2-\bar{x}_1}{s_p\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}} = \frac{44{,}2-31}{\sqrt{62{,}4}\sqrt{\frac{1}{4}+\frac{1}{5}}} = \frac{11\sqrt{78}}{39} \approx 2{,}49 \sim t_{n_A+n_B-2}$$

$$t_{кр} = t_{1-\alpha,\,4+5-2} = t_{0{,}95,\,7} = 1{,}895$$

~~H_0~~ ⇒ имеют более высокие доходы



б) Используем кр. Вилкоксона

Доход	19	26	37	39	41	42	42	47	52
Группа	A	A	A	B	B	A	B	B	B
Ранг	1	2	3	4	5	6,5	6,5	8	9

$R_1 = 1+2+3+6{,}5 = 12{,}5$

$R_2 = 4+5+6{,}5+8+9 = 32{,}5$

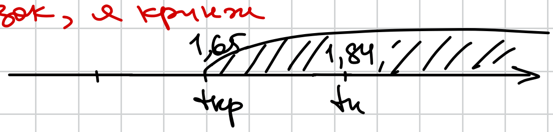
$$E(R_2) = \frac{n_2(n_1+n_2+1)}{2} = \frac{5(4+5+1)}{2} = 25$$

$D(R_2) = \frac{n_1 n_2 (n_1+n_2+1)}{12} = \frac{4 \cdot 5 (4+5+1)}{12} = \frac{5 \cdot 10}{3} \approx 16,67$

Тут должна быть формула для связок, я кринж

$t_н = \frac{R_2 - E(R_2)}{\sqrt{D(R_2)}} = \frac{32,5-25}{\sqrt{16,67}} \approx 1,84 \sim N(0,1)$

$t_{кр} = z_{1-\alpha} = z_{1-0,05} = 1,65$



Но ⇒ имеет более высокие доходы

Ответ: и в а), и в б) можно считать, что работники со средним специальном образованием имеют в среднем более высокие доходы, чем имеющие неполное среднее образование. Кроме Стьюдента и Вилкоксона тут мы можем использовать еще дисперсионный анализ.

Минус Стьюдента в том, что нам нужно равенство дисперсий

Плюс Вилкоксона в том, что он устойчив к выбросам и ненормальности распределения

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:

семинар 15-16 (прям эта, но без 6)

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) $\hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$ → вывод данной формулы можно посмотреть в семинаре 15-16

$\hat{a} = \bar{y} - \hat{b}\bar{x}$

$\bar{x} = \frac{1+2+3+4+5}{5} = 3$

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

$\sum (x_i - \bar{x})(y_i - \bar{y}) = (1-3)(6-4,4) + (2-3)(5-4,4) + (3-3)(4-4,4) + (4-3)(4-4,4) + (5-3)(3-4,4) = (-2) \cdot 1,6 + (-1) \cdot 0,6 + 0 \cdot (-0,4)$

$+ 1 \cdot (-0,4) + 2 \cdot (-1,4) = -3,2 - 0,6 + 0 - 0,4 - 2,8 = -7$

$\sum (x_i - \bar{x})^2 = (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 = 4+1+0+1+4 = 10$

$\hat{b} = \frac{-7}{10} = -0,7$ $\qquad \hat{a} = 4,4 - (-0,7) \cdot 3 = 4,4 + 2,1 = 6,5$

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

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

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

$\hat{y}_i = 6{,}5 - 0{,}7X$

$y_i = 6{,}5 - 0{,}7X + \varepsilon$

x_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

Сумма остатков: $0{,}04 + 0{,}01 + 0{,}16 + 0{,}09 + 0 = 0{,}3$

$\hat{\sigma}^2 = \frac{0{,}3}{5-1-1} = \frac{0{,}3}{3} = 0{,}1$

3) и 4) пункт они благополучно продублировали, не дописав $H_0: b = 0$. Предполагаю это по тому, что задачка 2 из семинара 15-16

$H_0: b = 0$

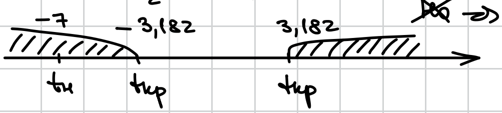
$H_1: b \neq 0$

$T = \frac{\hat{b} - b}{\hat{\sigma} \sum (x_i - \bar{x})} \sim t(n-p-1)$

$t_н = \frac{-0{,}7 - 0}{\sqrt{0{,}1} \cdot 10} = \frac{-0{,}7}{0{,}01 \cdot 10} = \frac{-0{,}7}{0{,}1} = -7$

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

$\cancel{H_0} \Rightarrow b \neq 0$ влияние фактора значимо



5) Точечная оценка

$X = 6$ $\hat{y} = 6{,}5 - 0{,}7X = 6{,}5 - 0{,}7 \cdot 6 = 6{,}5 - 4{,}2 = 2{,}3$

Интервальный прогноз

$\hat{y}(x_0) \pm t_{\frac{\alpha}{2}, n-2} \cdot SEpred$, где $SEpred = \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 + \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)} = \sqrt{0{,}1 \left(1 + \frac{1}{5} + \frac{(6-3)^2}{10}\right)} = \sqrt{0{,}1 \cdot 2{,}1} \approx 0{,}458$

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

6) $R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{0{,}3}{5{,}2} = \frac{49}{52} \approx 0{,}94$

$TSS = \sum (y_i - \bar{y})^2 = (1{,}6)^2 + (0{,}6)^2 + (-0{,}4)^2 + (-0{,}4)^2 + (-1{,}4)^2 = \frac{64}{25} + \frac{9}{25} + 2 \cdot \frac{4}{25} + \frac{49}{25} = \frac{26}{5} = 5{,}2$

$RSS = \sum (y_i - \hat{y})^2 = 0{,}3$

не забудьте что тут уже КВАДРАТ R^2

$R^2_{скорр} = 1 - (1 - R^2) \frac{n-1}{n-p-1} = 1 - (1 - 0{,}94) \left(\frac{5-1}{5-1-1}\right) = 0{,}92$

Ответ: 1) $\hat{a} = 6{,}5$ $\hat{b} = -0{,}7$ $\hat{y} = 6{,}5 - 0{,}7X$

2) $\hat{\sigma}^2 = 0{,}1$

3) Отвергли H_0 гипотезу. Влияние фактора значимо

4) я их не пон, но мне кажется они случайно повторили 3)

5) Точечная. 2,3 , интервальная $[0{,}84; 3{,}76]$

6) $R^2 = 0{,}94$, $R_{adj} = 0{,}92$

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.

задачка 1 со стр 141 моего файла (похожая)

там опечатки, аккуратно смотрите семинар 11 тоже

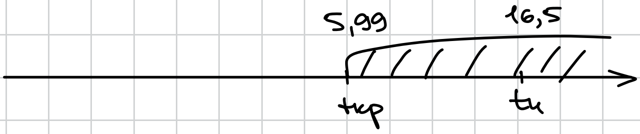
	+	−	±	
М	30	40	30	100
Ж	80	50	20	150
	110	90	50	250

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

$H_a: \exists (i,j): p_{ij} \neq p_{i\cdot} \cdot p_{\cdot j}$ (зависимы)

$\alpha = 0{,}05$

~~Do~~ => зависимы



1) Считаем хи-квадрат

(строки − 1)·(столбцы − 1) = (2−1)(3−1) = 1·2 = 2

$\chi^2_{1-0,05;2} = \chi^2_{0,95;2} = 5{,}99$ ← tкр

2) $t_н = \hat{\chi}^2 = n\left(\sum_{i=1}^{m}\sum_{j=1}^{k} \frac{n_{ij}^2}{n_i n_j} - 1\right)$

$$250\left(\frac{30^2}{110\cdot 100} + \frac{40^2}{90\cdot 100} + \frac{30^2}{50\cdot 100} + \frac{80^2}{110\cdot 150} + \frac{50^2}{90\cdot 150} + \frac{20^2}{50\cdot 150} - 1\right) \approx 250\cdot 0{,}066 = 16{,}5$$

$$p = \sqrt{\frac{t_н}{t_н + n}} = \sqrt{\frac{16{,}5}{16{,}5+250}} \approx 0{,}24 \in [0,1] \text{ зависимы}$$

$$C = \sqrt{\frac{t_н}{n\cdot \min\{(m-1),(k-1)\}}} = \sqrt{\frac{16{,}5}{250\cdot 1}} \approx 0{,}26$$

$C \in [0, 0{,}3)$ — связь слабая, но почти умеренная

Ответ: да, можно. Присутствует слабо-умеренная связь

если душнить слабая

Некоторых задач не было, поэтому давайте найдем те, которых не было

КОММЕНТАРИИ

В данном варианте отсутствуют (но в других вариантах могут быть) задачи на следующие темы:

1) Проверка гипотезы об однородности двух выборок против альтернативы масштаба;
2) Задача однофакторного дисперсионного анализа;
3) Задача на меры прогноза в номинальной шкале
4) Построение асимптотического доверительного интервала для коэффициента корреляции показателей, имеющих гауссовское распределение;
5) Задача о согласованности экспертной группы;
6) Построение оценок и проверка гипотез в моделях множественной линейной регрессии (в качестве дополнительной информации будут даны некоторые промежуточные вычисления);
7) Частные коэффициенты корреляции;
8) Критерий Чоу;
9) Фиктивные переменные.
Несмотря на то, что в экзаменационных задачах будут заданы выборки малых объёмов, разрешается (из-за отсутствия нужного набора таблиц) использовать асимптотический подход.

1) Проверка гипотезы об однородности двух выборок против альтернативы масштаба

(задача с 1 консультации, стр 153)

3. (5 баллов) Известно, что за последние годы акции A и акции B давали в среднем одинаковую доходность. За последние 8 месяцев доходность акций A составила: 108, 102, 95, 94, 97, 104, 105, 103 у.е., доходность акций B составила: 161, 56, 101, 51, 60, 150, 57, 156 у.е. Риски акций определяются среднеквадратическими отклонениями доходностей. Можно ли считать, опираясь на эти данные, что вложения в акции B являются более рискованными, чем вложения в акции A? Решите задачу, используя следующие предположения: а) все наблюдения имеют нормальное распределение; б) наблюдения имеют некоторые неизвестные непрерывные распределения. Полностью опишите процедуру проверки соответствующих гипотез. Какие ещё известные Вам критерии можно применить для решения данной задачи?

Фишер и Ансари – Брэдли ставит гипотезу через $\frac{\sigma_y}{\sigma_x}$

$X: 108, 102, 95, 94, 97, 104, 105, 103$, $n_x = 8$

$Y: 161, 56, 101, 51, 60, 150, 57, 156$, $n_y = 8$

X – выборка A

Y – выборка B

$H_0: \Delta = \frac{\sigma_y}{\sigma_x} = 1$

$H_a: \Delta > 1$

а) Все имеют норм распред

$$F_{n,m} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}{\frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x})^2} \sim \frac{\tilde{S}_y^2}{\tilde{S}_x^2} \sim F(n-1, m-1)$$

Используем критерий Фишера

помним что наблюдаемое ВСЕГДА >1

если $\frac{\sigma_y}{\sigma_x} < 1$, то $\frac{\tilde{S}_x^2}{\tilde{S}_y^2} \sim F(m-1, n-1)$ и тут

поменяли тут

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

$$\tilde{S}_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 25{,}714$$

$$\bar{x} = \frac{108+102+95+94+97+104+105+103}{8} = 101$$

$$\frac{1}{7}\sum_{i=1}^{8}(x_i-\bar{x})^2 = \frac{(108-101)^2+(102-101)^2+(95-101)^2+(94-101)^2+(97-101)^2+(104-101)^2+(105-101)^2+(103-101)^2}{7} \approx 25{,}714$$

$$\tilde{S}_y^2 = \frac{1}{n_y-1}\sum_{j=1}^{n_y}(y_i-\bar{y})^2 = 2448$$

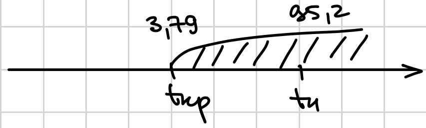
$$\bar{y} = \frac{161+56+101+51+60+150+57+156}{8} = 99$$

$$\frac{1}{7}\sum_{i=1}^{8}(y_i-\bar{y})^2 = \frac{(161-99)^2+(56-99)^2+(101-99)^2+(51-99)^2+(60-99)^2+(150-99)^2+(57-99)^2+(156-99)^2}{7} = \frac{17136}{7} = 2448$$

$\max(\tilde{S}_x^2, \tilde{S}_y^2) = \tilde{S}_y^2$

$t_н = \frac{\tilde{S}_y^2}{\tilde{S}_x^2} = \frac{2448}{25{,}714} \approx 95{,}2$ обращаем внимание что $t_н > 1$, для этого проверяем что больше S_x^2 или S_y^2

$t_{кр} = F(n-1, m-1) = F(7,7) = 3{,}79$



б) Используем критерий Ансари – Бредли

$X \sim F(t-m)$

$Y \sim F\left(\frac{t-m}{\Delta}\right)$

Группа 237 может и со связками использовать обычные формулы

$$A_{m,n} = \sum_{i=1}^{\overset{\text{кол-во } X}{m}} q_i = \begin{cases} R_i, & \text{если } R_i \le \frac{N+1}{2} \\ N+1-R_i, & \text{иначе} \end{cases}$$

$$A^* = \frac{A - E\{A\}}{\sqrt{D\{A\}}} \sim N(0,1)$$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
51	56	57	60	94	95	97	101	102	103	104	105	108	150	156	161
B	B	B	B	A	A	A	B	A	A	A	A	A	B	B	B

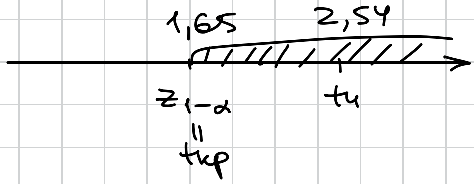
$$\frac{N+1}{2} = \frac{8+8+1}{2} = \frac{17}{2} = 8{,}5$$

$$A_ч = 5+6+7+(17-9)+(17-10)+(17-11)+(17-12)+(17-13) = 5+6+7+8+7+6+5+4 = 48$$

$$E_A = \frac{m(N+2)}{4} = \frac{8(8+8+2)}{4} = 2\cdot 18 = 36 \quad \text{т.к. } N = 8+8 = 16 \vdots 2$$

$$D_A = \frac{m\cdot n(N+2)(N-2)}{48(N-1)} = \frac{8\cdot 8(16-2)(16+2)}{48\cdot 15} = 22{,}4$$

$$A^*_ч = \frac{A - E\{A\}}{\sqrt{D\{A\}}} = \frac{48-36}{\sqrt{22{,}4}} \approx 2{,}54 \sim N(0,1) \quad \cancel{H_0} \Rightarrow \text{да, можно так считать}$$



$$z_{1-0{,}05} = z_{0{,}95} = 1{,}65$$

Ответ: и для а), и для б) да, можно так считать

2) Задача однофакторного дисперсионного анализа

стр 154

2. (5 баллов) Проведено исследование по оценке роли чистой мотивации (знания цели работы) на выполнение скучных монотонных операций. Исследовалась операция вытачивания металлической заготовки. 15 рабочих, имеющих одинаковую квалификацию и опыт работы, были случайным образом разделены на три группы. Группа A не имела информации о требуемой производительности. Группа B получила лишь общие сведения. Группа C имела точную информацию о задании. В таблице приведено число обработанных заготовок. Можно ли считать (по результатам этого эксперимента), что производительность труда растет с осведомленностью? 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.

A	40	35	38	43	44
B	38	40	47	44	41
C	48	40	45	43	49

б) >2 выборок + неизвестн. распр $\Rightarrow$ Джонкхиер

$H_0: \theta_1 = \ldots = \theta_k = \theta$ выборки однородны

$H_a: \theta_1 \le \theta_2 \le \ldots \le \theta_k$ выборки неоднородны

	n_j	
A	$n_1 = 5$	40 35 38 43 44
B	$n_2 = 5$	38 40 47 44 41
C	$n_3 = 5$	48 40 45 43 49

$$\varphi(y,z) = \begin{cases} 1, & y<z \\ 0{,}5, & y=z \\ 0, & y>z \end{cases} \qquad U_{\ell m} = \sum_{x\in \ell} \sum_{x \in m} \varphi(x,y)$$

$$J = \sum_{\ell < m} U_{\ell m}$$

A и B

$U_{em} = 3,5+5+4,5+2+1,5 = 16,5$

B и C

$U_{lm} = 5+4,5+2+3+4 = 18,5$

A и C

$U_{lm} = 4,5+5+5+3,5+3 = 21$

$J = 16,5+18,5+21 = 56$

$E(J) = \frac{1}{4}(N^2 - \sum w_j^2) = \frac{1}{4}((5\cdot 3)^2 - 5^2\cdot 3) = \frac{1}{4}(15^2 - 75) = 37,5$

$D(J) = \frac{1}{72}(N^2(2N+3) + \sum w_j^2(2w_j+3)) = \frac{1}{72}(15^2\cdot 33 - 3\cdot 25\cdot 13) = 89,58$

$\sigma_J = \sqrt{D(J)} = \sqrt{89,58} = 9,47$

$J^* = \frac{J - E(J)}{\sigma_J} = \frac{56 - 37,5}{9,47} \approx 1,95 \sim N(0,1)$

$t_{кр} = z_{0,95} = 1,65$ ~~H_0~~