Zadania z wykładu 9

Krystian Baran 145000

11 maja 2021

Spis treści

1 Zadanie 1 (TG 6.32) **3**

2 Zadanie 2 **4**
2.1 a) 4
2.2 b) 5

3 Zadanie 3 **6**

4 Zadanie 6 **8**

5 Zadanie 7 **10**
5.1 a) 10
5.2 b) 11

1 Zadanie 1 (TG 6.32)

Zbadano wzrost 13 mężczyzn oraz 12 kobiet w pewnym ośrodku sportowym. Dane:
M: 171, 176, 179, 189, 176, 182, 173, 179, 184, 186, 189, 167, 177,
K: 161, 162, 163, 162, 166, 164, 168, 165, 168, 157, 161, 172.
Zakładając, że w obu populacjach rozkład wzrostu jest normalny, czy można powiedzieć, że mężczyźni charakteryzują się większą zmiennością wzrostu? Przyjąć poziom istotności 0,1.

Wyznaczono na początku średnią i wariancję nieobciążoną z podanych prób zgodnie ze wzorami poniżej:

$$\overline{X} = \frac{\sum x_i}{n}$$
$$S^2 = \frac{\sum (x_i - \overline{X})^2}{n-1}$$

Otrzymano następujące wartości:

	$\overline{X}$	S^2	n
M	179.076923	45.74359	13
K	164.083333	16.083333	12

Pytanie czy mężczyźni charakteryzują się większą zmiennością wzrostu można przedstawić jako hipotezę alternatywną wraz z hipotezą zerową następująco:

H_0	$\sigma_M^2 \leqslant \sigma_K^2$
H_1	$\sigma_M^2 > \sigma_K^2$

Ponieważ zakładamy że rozkłady są typu normalnego możemy zastosować test F Snedecora. Poziom istotności $\alpha = 0.1$.

$$F = \frac{max\{S_M^2, S_K^2\}}{min\{S_M^2, S_K^2\}} = \frac{45.74359}{16.083333} \approx 2.844161 = F_0$$

Statystyka ta ma rozkład statystki F Snedeora ze stopniami swobody licznika i mianownika odjąć jeden. W tym przypadku:

$$F(13-1, 12-1) = F(12, 11)$$

Obliczymy teraz *p value* zgodnie ze wzorem:

$$\text{p-value} = 1 - F_{F(12,11)}(F_0) \stackrel{R}{=} 1 - pf(2.844161, 12, 11) \approx 0.04689163$$

Ponieważ *p-value* jest mniejsze od przyjętego poziomu istotności odrzucamy hipotezę zerową i wnioskujemy że wariancją wzrostu mężczyzn jest znacznie większa niż wariancja wzrostu kobiet.

2 Zadanie 2

Spośród absolwentów pewnej uczelni wylosowano 15 osób z jednego wydziału oraz 12 osób z drugiego wydziału i obliczono średnią ocen ze studiów dla każdego absolwenta. Otrzymano następujące wyniki
dla pierwszego wydziału: 3.71, 4.28, 2.95, 3.20, 3.38, 4.05, 4.07, 4.98, 3.20, 3.43, 3.09, 4.50, 3.12, 3.68, 3.90,
dla drugiego wydziału: 3.10, 3.38, 4.06, 3.60, 3.81, 4.50, 4.00, 3.25, 4.11, 4.85, 2.80, 4.00.
Na poziomie istotności $\alpha = 0,05$ zweryfikować następujące hipotezy:

a) wariancje średnich ocen dla obydwu wydziałów są równe,

b) różnica wartości oczekiwanych ocen uzyskiwanych przez studentów obydwu wydziałów wynosi 0.

Wyznaczono na początku średnią i wariancję nieobciążoną z podanych prób zgodnie ze wzorami poniżej:

$$\overline{X} = \frac{\sum x_i}{n}$$

$$S^2 = \frac{\sum(x_i - \overline{X})^2}{n-1}$$

Otrzymano następujące wartości:

	$\overline{X}$	S^2	n
W1	3.702667	0.347207	15
W2	3.788333	0.349415	12

2.1 a)

Hipoteza że wariancję są sobie równe jest hipotezą zerową, zatem można wyznaczyć hipotezę alternatywną następująco:

H_0	$\sigma^2_{W1} = \sigma^2_{W2}$
H_1	$\sigma^2_{W1} \neq \sigma^2_{W2}$

Do oceny wariancji zastosujemy test F Snedecora wyrażony następująco:

$$F = \frac{max\{S^2_{W1}, S^2_{W2}\}}{min\{S^2_{W1}, S^2_{W2}\}} = \frac{0.349415}{0.347207} \approx 1.006359 = F_0$$

Gdzie statystyka F ma rozkład statystyki F Snedecora ze stopniami swobody licznika i mianownika obniżone o jeden, czyli:

$$F(12-1, 15-1) = F(11, 14)$$

Obliczymy teraz *p-value* zgodnie ze wzorem:

$$\text{p-value} = 2 \cdot min\{F_{F(11,14)}(F_0), 1 - F_{F(11,14)}(F_0)\}$$
$$\boxed{\begin{array}{c} F_{F(11,14)}(F_0) \stackrel{R}{=} pf(1.006359, 11, 14) \approx 0.513539 \\ 1 - F_{F(11,14)}(F_0) \stackrel{R}{=} 1 - pf(1.006359, 11, 14) \approx 0.486461 \end{array}}$$
$$= 2 \cdot 0.486461 = 0.972922$$

Ponieważ *p-value* jest większe od przyjętego $\alpha = 0.05$ nie możemy odrzucić hipotezę zerową, zatem jest możliwe że wariancje średnich ocen obydwu wydziałów są sobie równe.

2.2 b)

Podana hipoteza jest hipotezą zerową, zatem można wyznaczyć hipotezę alternatywną:

H_0	$m_{W1} - m_{W2} = 0$
H_1	$m_{W1} - m_{W2} \neq 0$

Ponieważ test przeprowadzony w poprzednim podpunkcie wykazało że wariancje obu populacji są sobie równe, możemy, do tego testu zastosować następującą statystykę:

$$t = \frac{(\overline{X}_{W1} - \overline{X}_{W2}) - m_0}{\sqrt{\frac{(n_{W1}-1)S_{W1}^2 + (n_{W2}-1)S_{W2}^2}{n_{W1}+n_{W2}-2} \frac{n_{W1}+n_{W2}}{n_{W1} \cdot n_{W2}}}} \sim t(n_{W1} + n_{W2} - 2)$$

Wtedy t_0 będzie równe:

$$t_0 = \frac{3.702667 - 3.788333}{\sqrt{\frac{14 \cdot 0.347207 + 11 \cdot 0.349415}{25} \frac{27}{15 \cdot 12}}} \approx -0.374854$$

Obliczymy teraz *p-value* zgodnie ze wzorem:

$$\text{p-value} = 2 \cdot min\{F_{t(25)}(t_0), 1 - F_{t(25)}(t_0)\}$$
$$\boxed{\begin{array}{c} F_{t(25)}(t_0) \stackrel{R}{=} pt(-0.374854, 25) \approx 0.3554651 \\ 1 - F_{t(25)}(t_0) \stackrel{R}{=} 1 - pt(-0.374854, 25) \approx 0.6445349 \end{array}}$$
$$= 2 \cdot 0.3554651 = 0.7109302$$

Widzimy że *p-value* jest większe od przyjętego $\alpha = 0.05$, zatem nie możemy odrzucić hipotezę zerową. Możemy powiedzieć że wartości oczekiwane średnich ocen z obu wydziałów są sobie równe.

3 Zadanie 3

Badano opony samochodowe typu 11.00-20/14PR/ produkowane przez dwóch producentów, które zostały wycofane z eksploatacji. Spośród zbadanych 1582 opon producenta A, 1250 opon wycofano z powodu zużycia bieżnika, a spośród 589 zbadanych opon producenta B, wycofano z powodu tego defektu 421 sztuk. Na poziomie istotności $\alpha = 0,01$ zweryfikować hipotezę, że frakcje opon wycofanych z eksploatacji na skutek zużycia się bieżnika są jednakowe dla obydwu producentów.

Wyrażona hipoteza jest hipotezą zerową, jako alternatywną wybierzemy że frakcje opon wycofanych z eksploatacji dla producenta A jest większa niż ta z producenta B, czyli:

H_0	$p_A - p_B = 0$
H_1	$p_A - p_B > 0$

Wyznaczymy wskaźnik struktury dla obu prób dzieląc liczebność wycofanych opon przez całkowitą ich ilość:

	$\overline{P}_k$	n
A	0.790139	1582
B	0.714771	589

Dla obu prób sprawdzono warunek:

$$\overline{p}_k \mp 3 \cdot \sqrt{\frac{\overline{p}_k(1-\overline{p}_k)}{n_k}}$$

$$A : 0.759425, 0.820853$$

$$B : 0.658957, 0.770585$$

Wszystkie te liczby należą do przedziału $(0,1)$, zatem warunek jest spełniony. Ponieważ liczebność obu populacji jest wystarczająco duża można zastosować następującą statystykę:

$$Z = \frac{\overline{P}_A - \overline{P}_B}{\sqrt{\overline{P}(1-\overline{P})(\frac{1}{n_A} + \frac{1}{n_B})}}$$

Gdzie $\overline{P}$ jest iloczynem sumy wyróżnionych elementów oby prób przez sumę całkowitych elementów oby prób:

$$\overline{P} = \frac{k_A + k_B}{n_A + n_B} = \frac{1671}{2171} \approx 0.769691$$

Statystyka ta ma rozkład statystki $\sim N(0,1)$. Obliczymy teraz Z_0 podstawiając znane wartości:

$$Z_0 = \frac{0.790139 - 0.714771}{\sqrt{0.769691(1-0.769691)(\frac{1}{1582} + \frac{1}{589})}} \approx 3.708551$$

Obliczymy teraz *p-value* zgodnie ze wzorem:

$$\text{p-value} = 1 - \Phi(3.708551) \overset{R}{=} 1 - pnorm(3.708551, 0, 1) \approx 0.0001042243$$

Liczba ta jest zdecydowanie mniejsza niż przyjęty $\alpha = 0.01$, zatem odrzucamy hipotezę zerową i wnioskujemy że frakcja opon wycofanych z eksploatacji na skutek zużycie się biernika producenta A jest większa niż ta producenta B.

4 Zadanie 6

W badaniu granicy plastyczności pewnego gatunku stali otrzymano następujące wyniki dla 15 kawałków tej stali
(wyniki w kg/cm^2): 3520, 3680, 3640, 3840, 3500, 3610, 3720, 3640, 3600, 3650, 3750, 3590, 3600, 3550, 3700.
Natomiast po dodatkowym procesie uszlachetniającym, mającym zwiększyć wytrzymałość tej stali, otrzymano dla tych samych próbek odpowiednio następujące wyniki badania granicy plastyczności:
3580, 3700, 3680, 3800, 3550, 3700, 3730, 3720, 3670, 3710, 3810, 3660, 3700, 3640, 3670.
Na poziomie istotności $\alpha = 0,05$ sprawdzić, czy granica plastyczności stali po dodatkowym procesie uszlachetniającym zwiększyła się.

Niech wyniki przed procesem będą X_1 a wyniki po procesie będą $X2$. Zbadamy $X_1 - X_2$, obliczymy średnią i wariancję zgodnie ze wzorami podanymi w poprzednich zadaniach.

Lp	$X_1 - X_2$
1	-60
2	-20
3	-40
4	40
5	-50
6	-90
7	-10
8	-80
9	-70
10	-60
11	-60
12	-70
13	-100
14	-90
15	30
SUM	-730
$\overline{X_1 - X_2}$	-48.666667
S^2	1769.52381
S	42.065708

Przedstawioną hipotezę można wyrazić następująco wraz z hipotezą zerową:

H_0	$m_{X_1} - m_{X_2} \geqslant 0$
H_1	$m_{X_1} - m_{X_2} < 0$

Zakładając że wyniki te mają rozkład normalny można zastosować następującą statystykę dla rozkładów sparowanych:

$$t = \frac{\overline{X_1 - X_2} - m_0}{\frac{S_{\overline{X_1 - X_2}}}{\sqrt{n}}} = \frac{-48.666667}{\frac{42.065708}{\sqrt{15}}} \approx -4.480733 = t_0$$

Statystyka ta ma rozkład statystyki $\sim t(n-1) = t(14)$. Obliczymy teraz *p-value* zgodnie ze wzorami:

$$\text{p-value} = F_{t(14)}(t_0) \stackrel{R}{=} pt(-4.480733, 14) \approx 0.0002589772$$

Wartość ta jest mniejsza od przyjętego $\alpha = 0.05$, zatem odrzucamy hipotezę zerową i wnioskujemy że granica plastyczności stali po dodatkowym procesie uszlachetniającym zwiększyła się.

5 Zadanie 7

Zmierzono czasy (w godzinach) usuwania awarii dla dwóch brygad remontowych. Dla pierwszej otrzymano czasy: 12, 13, 18, 25, 42, 19, 22, 35 a dla drugiej brygady: 23, 30, 27, 17, 21, 33, 31. Na poziomie istotności 0,05 zweryfikować hipotezy:

a) przeciętne czasy usuwania awarii dla obydwu brygad są równe,

b) wariancje czasów usuwania awarii dla obydwu brygad są równe.

Obliczono średnią i wariancję dla obu brygad i otrzymano następujące wyniki:

	$\overline{X}$	S^2	n
B1	23.25	110.214286	8
B2	26	34.333333	7

5.1 a)

Hipotezę tą można wyznaczyć wraz z hipotezą alternatywną w następujący sposób:

H_0	$m_{B_1} - m_{B_2} = 0$
H_1	$m_{B_1} - m_{B_2} \neq 0$

Jeżeli dane mają rozkład normalny możemy zastosować następującą statystykę:

$$t = \frac{(\overline{X}_{B_1} - \overline{X}_{B_2}) - m_0}{\sqrt{\frac{S_{B_1}^2}{n_{B_1}} + \frac{S_{B_2}^2}{n_{B_2}}}} \sim t(v)$$

Gdzie v jest wyrażony następującym wzorem:

$$\begin{aligned} v &= \frac{\left(\frac{S_{B_1}^2}{n_{B_1}} + \frac{S_{B_2}^2}{n_{B_2}}\right)^2}{\frac{1}{n_{B_1}-1}\left(\frac{S_{B_1}^2}{n_{B_1}}\right)^2 + \frac{1}{n_{B_2}-1}\left(\frac{S_{B_2}^2}{n_{B_2}}\right)^2} \\ &= \frac{\left(\frac{110.214286}{8} + \frac{34.333333}{7}\right)^2}{\frac{1}{7}\left(\frac{110.214286}{8}\right)^2 + \frac{1}{6}\left(\frac{34.333333}{7}\right)^2} \\ &\approx 11.213324 \end{aligned}$$

Obliczymy wartość t_0 podstawiając znane wartości:

$$t_0 = \frac{23.25 - 26}{\sqrt{\frac{110.214286}{8} + \frac{34.333333}{7}}} \approx -0.636248$$

Wyznaczymy teraz *p-value* zgodnie ze wzorem:

$$\text{p-value} = 2 \cdot min\{F_{t(11.213324)}(t_0), 1 - F_{t(11.213324)}(t_0)\}$$

$$\boxed{\begin{array}{c} F_{t(11.213324)}(t_0) \overset{R}{=} pt(-0.636248, 11.213324) \approx 0.2686926 \\ 1 - F_{t(11.213324)}(t_0) \overset{R}{=} 1 - pt(-0.636248, 11.213324) \approx 0.7313074 \end{array}}$$

$$= 2 \cdot 0.2686926 = 0.5373852$$

Wartość ta jest większa niż przyjęty $\alpha = 0.05$ zatem nie możemy odrzucić hipotezę zerową. Wnioskujemy że przeciętne czasy usuwania awarii dla obydwu brygad są równe.

5.2 b)

Hipotezę tą można wyznaczyć w następujący sposób wraz z hipotezą alternatywną:

H_0	$\sigma^2_{B_1} = \sigma^2_{B_2}$
H_1	$\sigma^2_{B_1} \neq \sigma^2_{B_2}$

Ponieważ badana jest wariancja zastosujemy statystykę F Snedecora wyrażona następująco:

$$F = \frac{max\{S^2_{B_1}, S^2_{B_2}\}}{min\{S^2_{B_1}, S^2_{B_2}\}} = \frac{110.214286}{34.333333} \approx 3.210125 = F_0$$

Gdzie statystyka F ma rozkład statystyki F Snedecora ze stopniami swobody licznika i mianownika obniżone o jeden, czyli:

$$F(8-1, 7-1) = F(7,6)$$

Obliczymy teraz *p-value* zgodnie ze wzorem:

$$\text{p-value} = 2 \cdot min\{F_{F(7,6)}(F_0), 1 - F_{F(7,6)}(F_0)\}$$

$$\boxed{\begin{array}{c} F_{F(7,6)}(F_0) \overset{R}{=} pf(3.210125, 7, 6) \approx 0.9117178 \\ 1 - F_{F(7,6)}(F_0) \overset{R}{=} 1 - pf(3.210125, 7, 6) \approx 0.08828215 \end{array}}$$

$$= 2 \cdot 0.08828215 = 0.1765643$$

Liczba ta jest większa od przyjętego $\alpha = 0.05$ zatem nie możemy odrzucić hipotezę zerową. Wnioskujemy że wariancje czasów usuwania awarii dla obydwu brygad są sobie równe.