Kaczmarek Wiktoria, Aktywność 09.

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.
Odp.: Wariancja wzrostu mężczyzn jest istotnie większa niż kobiet.

Na początek wyznaczamy średnia i wariancję dla podanych grup za pomocą wzorów:

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

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

M: $\bar{X} = 179{,}1 \quad S^2 = 45{,}74$

K: $\bar{X} = 164{,}1 \quad S^2 = 16{,}1$

Hipoteza alternatywna $\sigma_M^2 > \sigma_K^2$

Korzystamy z testu F

```
> x=c(171,176,179,189,176,182,173,179,184,186,189,167,177)
> var(x)
[1] 45.74359
> mean(x)
[1] 179.0769
> sqrt(var(x))
[1] 6.763401
> y=c(161,162,163,162,166,164,168,165,168,157,161,172)
> mean(y)
[1] 164.0833
> var(y)
[1] 16.08333
> var.test(x,y,alternative="greater")

        F test to compare two variances

data:  x and y
F = 2.8442, num df = 12, denom df = 11, p-value = 0.04689
alternative hypothesis: true ratio of variances is greater than 1
95 percent confidence interval:
 1.020301      Inf
sample estimates:
ratio of variances
          2.844161
```

Następnie obliczamy p-value lub odczytujemy z testu:

$$p\ value = 1 - F_{F(13-1=12;12-1=11)} = 1 - F_{F(12,11)} =$$

```
> 1-pf(2.844161,12,11)
[1] 0.04689163
```

P value < 0.1 więc wnioskujemy, że wariancja wzrostu mężczyzn jest znacznie większa od wariancji wzrostu kobiet.

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.

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

Po określeniu hipotez wyznaczymy wskaźnik struktury:

$$P_{k_A} = \frac{1250}{1582} =^R 0.7901391$$

$$P_{k_B} = \frac{421}{589} =^R 0.7147708$$

Korzystamy z:

7	Dla $k = 1, 2$ $X_k \sim B(p_k)$, $\bar{P}_k = \frac{1}{n_k} K_i$, $K_i = \sum_{i=1}^{n_k} X_{k,i}$ $\bar{p}_k \mp 3 \cdot \sqrt{\frac{\bar{p}_k(1-\bar{p}_k)}{n_k}} \subset (0,1)$	$p_1 - p_2 = d_0$	$\begin{cases} p_1 - p_2 \neq d_0 \\ p_1 - p_2 > d_0 \\ p_1 - p_2 < d_0 \end{cases}$	Jeśli $d_0 = 0$, to $Z = \frac{\bar{P}_1 - \bar{P}_2}{\sqrt{\bar{P}(1-\bar{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $\bar{P} = \frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1} X_{1,i} + \sum_{i=1}^{n_2} X_{2,i}\right)$ Jeśli $d_0 \neq 0$, to $Z = \frac{(\bar{P}_1 - \bar{P}_2) - d_0}{\sqrt{\frac{\bar{P}_1(1-\bar{P}_1)}{n_1}+\frac{\bar{P}_2(1-\bar{P}_2)}{n_2}}}$	asymptot. $\mathcal{N}(0,1)$	$\begin{cases} (-\infty, z_{\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}, \infty) \\ (z_{1-\alpha}, \infty) \\ (-\infty, z_\alpha) \end{cases}$

Na początek sprawdzamy warunek:

$$\overline{p_k} \mp 3\sqrt{\frac{\overline{p_k}(1-\overline{p_k})}{n}} = \frac{1250}{1582} \mp 3\sqrt{\frac{\frac{1250}{1582}\left(1-\frac{1250}{1582}\right)}{1582}}$$

$$\overline{p_k} \mp 3\sqrt{\frac{\overline{p_k}(1-\overline{p_k})}{n}} = \frac{421}{589} \mp 3\sqrt{\frac{\frac{421}{589}\left(1-\frac{421}{589}\right)}{589}}$$

```
> p<-1250/1582
> q<-3*sqrt(p*(1-p)/1582)
> p-q
[1] 0.7594251
> p+q
[1] 0.820853
> p<-421/589
> q<-3*sqrt(p*(1-p)/589)
> p-q
[1] 0.6589567
> p+q
[1] 0.7705849
```

A: 0 < {0.7594251; 0.820853} < 1 -> Warunek jest spełniony

B: 0 < {0.6589567; 0.770585} < 1 -> Warunek jest spełniony

Następnie musimy skorzystać z wzorów z tabelki.

Obliczamy:

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

$$Z = \frac{\overline{P_A} - \overline{P_B}}{\sqrt{\bar{P}(1-\bar{P})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

Statystyka Z~N(0,1)

$$Z_0 = \frac{0.790139 - 0.714771}{\sqrt{0.769691(1-0.769691)\left(\frac{1}{1582} + \frac{1}{589}\right)}} = 3{,}708567$$

```
> p<-1671/2171
> p
[1] 0.7696914
> p_a<-1250/1582
> p_b<-421/589
> z<-(p_a-p_b)/sqrt(p*(1-p)*(1/1582+1/589))
> z
[1] 3.708567
```

Następnie możemy obliczyć p value:

$$p\ value = 1 - \phi(3{,}708567) = 0{,}000104179$$

```
> 1-pnorm(z,0,1)
[1] 0.0001042179
```

P value << 0.01 stąd odrzucamy H0 i wnioskujemy, że frakcja opon wycofanych z eksploatacji na skutek zużycia biernika firmy A jest większa niż firmy B.

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

a) granica plastyczności stali po dodatkowym procesie uszlachetniającym zwiększyła się;

b) wariancja plastyczności stali po dodatkowym procesie uszlachetniającym zmniejszyła się.

Dokonujemy podstawowej analizy danych

przed	po	przed-po	(c-sr)^2
3520	3580	-60	128,4444
3680	3700	-20	821,7778
3640	3680	-40	75,11111
3840	3800	40	7861,778
3500	3550	-50	1,777778
3610	3700	-90	1708,444
3720	3730	-10	1495,111
3640	3720	-80	981,7778
3600	3670	-70	455,1111
3650	3710	-60	128,4444
3750	3810	-60	128,4444
3590	3660	-70	455,1111
3600	3700	-100	2635,111
3550	3640	-90	1708,444
3700	3670	30	6188,444
	suma	-730	
	średnia	-48,6667	
	wariancja	1769,524	
	odchylenie standardowe	42,06571	

a)

$\begin{cases} m_1 - m_2 = m_0 \\ m_1 - m_2 \le m_0 \\ m_1 - m_2 \ge m_0 \end{cases}$	$\begin{cases} m_1 - m_2 \ne m_0 \\ m_1 - m_2 > m_0 \\ m_1 - m_2 < m_0 \end{cases}$	$X_1 \sim \mathcal{N}(m_1, \sigma_1)$, $X_2 \sim \mathcal{N}(m_2, \sigma_2)$, σ_1, σ_2 nieznane, $n_1 = n_2 = n$	$\dfrac{\overline{X_1 - X_2} - m_0}{\frac{S_{\overline{X_1 - X_2}}}{\sqrt{n}}}$	$t(n-1)$	$\begin{cases} (-\infty, t_{\frac{\alpha}{2};n-1}) \cup (t_{1-\frac{\alpha}{2};n-1}, \infty) \\ (t_{1-\alpha;n-1}, \infty) \\ (-\infty, t_{\alpha;n-1}) \end{cases}$

H_0	$m_{x_1} - m_{x_2} \ge 0$
H_1	$m_{x_1} - m_{x_2} < 0$

Zakładamy, że wyniki mają rozkład normalny.

$$t = \frac{\overline{X_1 - X_2} - m_0}{\frac{S}{\sqrt{n}}}$$

$$t_0 = \frac{-48{,}667}{\frac{42{,}06571}{\sqrt{15}}} = -4{,}480733$$

Mamy rozkład zbliżony do t(n-1) czyli t(14).

Obliczamy p value:

```
> pt(-4.480733,14)
[1] 0.0002589772
```

p value = 0.0002589772

P value << 0.05 więc odrzucamy H0 i wnioskujemy, że granicza plastyczności stali po dodatkowym procesie zwiększyła się.

b)

H_0	$\sigma_{x_1}^2 - \sigma_{x_2}^2 \geq 0$
H_1	$\sigma_{x_1}^2 - \sigma_{x_2^2} < 0$

6	$\begin{cases}\sigma_1^2 = \sigma_2^2 \\ \sigma_1^2 \leq \sigma_2^2 \\ \sigma_1^2 \geq \sigma_2^2\end{cases}$	$\begin{cases}\sigma_1^2 \neq \sigma_2^2 \\ \sigma_1^2 > \sigma_2^2 \\ \sigma_1^2 < \sigma_2^2\end{cases}$	$X_1 \sim \mathcal{N}(m_1, \sigma_1)$ $X_2 \sim \mathcal{N}(m_2, \sigma_2)$ n_1, n_2 dowolne	$F = \frac{\max\{S_1^2, S_2^2\}}{\min\{S_1^2, S_2^2\}}$	$F(n_l - 1, n_m - 1)$ gdzie n_l, n_m liczebności licznika i mianownika	$\begin{cases}(F_{1-\frac{\alpha}{2}; n_l-1; n_m-1}, \infty) \\ (F_{1-\alpha; n_l-1; n_m-1}, \infty) \\ (F_{1-\alpha; n_m-1; n_l-1}, \infty)\end{cases}$

$$F_0 = \frac{67037,140}{4688,571} = 14,298$$

Następnie obliczamy przedział krytyczny:

```
> qf(0.95,14,14)
[1] 2.483726
```

$$(2,483726;\ \infty)$$

F0 należy do przedziału krytycznego, więc odrzucam H0 i wnioskujemy, że wariancja plastyczności stali po procesie uszlachetniania zmniejszyła się.

9. Wygenerować próby o liczebnościach 80 i 60 według rozkładów $\mathcal{N}(900; 50)$ i $\mathcal{N}(1000; 60)$ i na ich podstawie przeprowadzić test, że wartości oczekiwane różnią się o 50.

H_0	$m_{x_1} - m_{x_2} = 50$
H_1	$m_{x_1} - m_{x_2} \neq 50$

```
> x<-rnorm(80,900,50)
> sort(x)
 [1] 778.1253 799.2921 818.1077 825.0299 836.3916 840.4220 840.5588 840.8815 843.4571 847.5834 848.2139 855.9194 856.3703 856.4712
[15] 856.7895 857.5792 860.6759 869.0173 869.3348 878.2251 878.6567 879.0543 882.8738 883.4117 883.7300 884.7737 885.3520 890.7767
[29] 893.1621 893.6860 896.8981 899.0562 901.1276 901.8316 901.8437 903.1367 905.3861 908.0902 909.5924 909.7417 910.2074 912.1631
[43] 915.2465 915.7067 916.1193 916.8839 919.8486 920.1471 922.4571 925.2743 925.3733 927.8124 929.1802 929.9498 930.0189 930.0530
[57] 930.2960 934.8490 935.2502 937.5662 940.8860 942.2664 943.4990 943.9034 945.5227 947.0923 947.8192 953.4316 955.2303 955.3092
[71] 955.8616 961.9662 965.2559 969.9649 972.1674 974.7933 983.2231 984.3188 984.9063 986.9058
```

```
> y<-rnorm(60,1000,60)
> sort(y)
 [1]  827.1204  900.0272  925.8774  926.3792  927.5018  929.4909  935.8579  937.7152  940.9285  943.0775  943.2636  956.0388
[13]  965.0402  965.0928  965.5702  970.0062  977.2888  978.7376  978.7462  979.8611  981.0423  984.1150  984.8269  990.9601
[25]  992.4881  992.5056  994.0732  994.1031  996.0001 1000.1698 1001.1779 1004.0934 1004.1807 1013.1102 1021.1427 1023.6002
[37] 1026.6850 1031.5570 1039.1424 1039.9530 1041.5022 1041.6245 1042.2536 1045.3721 1047.3336 1050.7753 1051.3852 1054.8238
[49] 1058.7603 1068.1905 1073.5585 1074.3594 1088.4712 1098.6607 1100.1673 1105.3385 1106.0253 1111.0023 1120.0783 1125.9166
```

Podstawowe parametry:

```
> var(x)
[1] 2092.542
> mean(x)
[1] 905.8669
> var(y)
[1] 3723.452
> mean(y)
[1] 1008.236
```

Korzystamy z:

3	$\begin{cases} m_1 - m_2 = m_0 \\ m_1 - m_2 \leq m_0 \\ m_1 - m_2 \geq m_0 \end{cases}$	$\begin{cases} m_1 - m_2 \neq m_0 \\ m_1 - m_2 > m_0 \\ m_1 - m_2 < m_0 \end{cases}$	$X_1 \sim \mathcal{N}(m_1, \sigma_1)$, $X_2 \sim \mathcal{N}(m_2, \sigma_2)$, σ_1, σ_2 nieznane oraz różne	(Cochrana – Coxa) $\frac{(\bar{X}_1 - \bar{X}_2) - m_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$t(v)$	$\begin{cases} (-\infty, t_{\frac{\alpha}{2};v}) \cup (t_{1-\frac{\alpha}{2};v}, \infty) \\ (t_{1-\alpha;v}, \infty) \\ (-\infty, t_{\alpha;v}) \end{cases}$

$$t_0 = \frac{(905{,}8669 - 1008{,}236) - 50}{\sqrt{\frac{50}{80} + \frac{60}{60}}}$$

```
> ((var(x)-var(y))-50)/sqrt(50/80+60/60)
[1] -1318.615
```

$$t_0 = -1318{,}615$$

$$\alpha = 0.05$$

Obliczamy przedział krytyczny:

```
> qnorm(0.25,0,1)
[1] -0.6744898
> qnorm(0.75,0,1)
[1] 0.6744898
```

$$z_{\frac{\alpha}{2}} = -0.6744898, \qquad z_{1-\frac{\alpha}{2}} = 0.6744898$$

$$t_0 \in (-\infty; -0.6744898) \cup (0.6744898; \infty)$$

Wnioskujemy, że wartości oczekiwane różnią się o 50, bo wartość t0 należy do obszaru krytycznego.