Programowanie narzędzi analitycznych – Z02

1 Statystyka opisowa

Wariancją z próby losowej $X1, X2, ..., Xn$ nazywamy liczbę

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 \tag{1}$$

Nieobciążony estymator wariancji określamy jako

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 \tag{2}$$

Wskaźnik asymetrii to

$$g_1 = \frac{M_3}{\hat{S}^3} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^3}{\hat{S}^3} \tag{3}$$

Współczynnik koncentracji (skupienia, kurtozy) to

$$g_1 = \frac{M_4}{\hat{S}^4} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^4}{\hat{S}^4} \tag{4}$$

Odchylenie standardowe

$$\hat{S} = \sqrt{\hat{S}^2} \tag{5}$$

Odchylenie przeciętne

$$d = \frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}| \tag{6}$$

Współczynnik zmienności

$$V = \frac{\hat{S}}{\bar{X}}100\% \tag{7}$$

WSKAŹNIKI ASYMETRII

Współczynnik asymetrii (klasyczny)

$$A = \frac{M_3}{\hat{S}^3} \tag{8}$$

gdzie M_3 jest trzecim momentem centralnym równym dla danych surowych

$$M_3 = \frac{1}{n}\sum_{i=1}^{k} n_i(X_i - \bar{X})^3 \tag{9}$$

Pozycyjny miernik asymetrii

$$A_2 = \frac{Q_3 - 2Med + Q_1}{Q_3 - Q_1} \tag{10}$$

Współczynnik skośności

$$A_1 = \frac{\bar{X} - Mo}{S} \tag{11}$$

Zadanie 1.1
Sprawdzić, którą wariancję wylicza się w R poleceniem `var`.

Zadanie 1.2 (Zadanie 1.34 z [1])
Liczba koni (w mln szt.) w Polsce w kolejnych latach od 1947 do 1974 wynosiła: 2,0; 2,3; 2,7; 2,8; 2,9; 2,7; 2,6; 2,6; 2,5; 2,6; 2,7; 2,8; 2,8; 2,7; 2,7; 2,6; 2,6; 2,6; 2,6; 2,6; 2,7; 2,6; 2,6; 2,5; 2,4; 2,4; 2,3. Obliczyć: średnią liczbę koni w tym okresie, medianę, odchylenie standardowe współczynnik zmienności.

Zadanie 1.3 (Zadanie 1.1 z [2])
W grupie 25 studentów zbadano oceny pracy kontrolnej ze statystyki. Otrzymano wyniki: 3.5; 4; 3; 3.5; 4.5; 3; 3; 3; 2.5; 4; 2; 3.5; 2; 2.5; 3.5; 4; 5; 2.5; 3; 2; 5; 4; 2; 3; 3. Wyznacz podstawowe miary położenia i rozproszenia (wartość średnią, medianę, wariancję próbkową oraz pozycyjny miernik asymetrii).

2 Przedziały ufności

Model I
$X_1, X_2, \ldots, X_n$ i.i.d. z rozkładu $N(\mu, \sigma^2)$, $\mu \in R$ nieznane, $\sigma > 0$ znane. Przedział

$$\left[\bar{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

jest przedziałem ufności na poziomie ufności $1 - \frac{\alpha}{2}$ dla parametru μ, gdzie $u_{1-\frac{\alpha}{2}}$ jest kwantylem rzędu $1-\alpha$ z rozkładu $N(0,1)$

Model II
$X_1, X_2, \ldots, X_n$ i.i.d. z rozkładu $N(\mu, \sigma^2)$, $\mu \in R$ nieznane, $\sigma > 0$ nieznane. Przedział

$$\left[\bar{X} - t_{1-\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}; \bar{X} + t_{1-\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}\right]$$

jest przedziałem ufności na poziomie ufności $1 - \alpha$ dla parametru μ, gdzie $t_{1-\frac{\alpha}{2},n-1}$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$ z rozkładu t-Studenta z $n-1$ stopniami swobody

Model III
$X_1, X_2, \ldots, X_n$ i.i.d. z rozkładu $N(\mu, \sigma^2)$, $\sigma > 0$ nieznane. Przedział

$$\left[\frac{S^2(n-1)}{\chi^2(1-\frac{\alpha}{2}, n-1)}; \frac{S^2(n-1)}{\chi^2(\frac{\alpha}{2}, n-1)}\right]$$

jest przedziałem ufności na poziomie ufności $1 - \frac{\alpha}{2}$ dla wariancji, gdzie $\chi^2(1-\frac{\alpha}{2}, n-1)$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ z rozkładu chi-kwadrat z $n-1$ stopniami swobody, $\chi^2(\frac{\alpha}{2}, n-1)$ jest kwantylem rzędu $\frac{\alpha}{2}$ z rozkładu chi-kwadrat z $n-1$ stopniami swobody

Model IV
$X_1, X_2, \ldots, X_n$ i.i.d. z dowolnego rozkładu o skończonej wartości oczekiwanej i wariancji, zakładamy, że n jest duże ($n > 50$). $N(\mu, \sigma^2)$, $\sigma > 0$ nieznane. Przedział

$$\left[\bar{X} - u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}; \bar{X} + u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$$

przybliżonym przedziałem ufności dla parametru μ na poziomie ufności $1-\alpha$, gdzie $u_{1-\frac{\alpha}{2}}$ jest kwantylem rzędu $1-\alpha$ z rozkładu $N(0,1)$

Model V
Wykonujemy n niezależnych doświadczeń typu sukces-porażka, n jest duże. Obserwowana zmienna losowa X ma rozkład dwupunktowy

$$P(X=1) = p, P(X=0) = 1-p$$

$p \in (0,1)$ to prawdopodobieństwo sukcesu w pojedynczym doświadczeniu, nieznany parametr nazywany też wskaźnikiem struktury. Przedział

$$\left[\hat{p} - u_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}; \hat{p} + u_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right]$$

przybliżonym przedziałem ufności dla parametru p na poziomie ufności $1-\alpha$, gdzie $u_{1-\frac{\alpha}{2}}$ jest kwantylem rzędu $1-\alpha$ z rozkładu $N(0,1)$

Model VI
Rozważmy dwie niezależne próbki $X_1, X_2, \ldots, X_k \sim N(\mu_X, \sigma_X^2)$, $Y_1, Y_2, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$. Wówczas zachodzi: $P(f_1 \frac{S_Y^2}{S_X^2} \leq \frac{\sigma_Y^2}{\sigma_X^2} \leq f_2 \frac{S_Y^2}{S_X^2}) = 1-\alpha$, gdzie f_1, f_2 to odpowiednio kwantyle rzędu $\frac{\alpha}{2}$ i $1-\frac{\alpha}{2}$ dla rozkładu F-Snedecora z $(k-1)$ i $(m-1)$ stopniami swobody. Wówczas

$$\left[f_1 \frac{S_Y^2}{S_X^2}; f_2 \frac{S_Y^2}{S_X^2}\right]$$

jest przedziałem ufności na poziomie ufności $1-\alpha$ dla ilorazu wariancji.

Zadanie 3.1 (Zadanie 6.1 z [2])
Firma telekomunikacyjna chce oszacować średnią długość rozmów zamiejscowych w soboty i niedziele na podstawie 20 elementowej próby losowej, dla której średnia wynosi 14,5 i odchylenie standardowe 5,6. Zakładając, że czas rozmowy ma rozkład normalny wyznaczyć realizacji przedziału ufności dla wartości oczekiwanej czasu rozmowy na poziomie ufności 95%. Jak zmieni się długość przedziału ufności gdy poziom ufności wzrośnie.

Zadanie 3.2 (Zadanie 6.3 z [2])
Właściciel kantoru wymiany walut na lotnisku chce wyestymować średnią wielkość gotówki potrzebną do wymiany nocą franków na dolary. Z doświadczenia właściciel wie, że wielkość popytu na dolary ma rozkład normalny. Obserwując popyt przez 10 dni właściciel otrzymał wyniki: $\bar{X} = 24.4$, i $S = 3.5$. Podać realizację przedziału ufności dla wartości oczekiwanej popytu na poziomie ufności 0.9.

Zadanie 3.3 (Zadanie 6.6 z [2])
Procent gospodarstw rolnych, których właściciele przekroczyli 60 lat oznaczmy przez $100\theta\%$. W celu oszacowania parametru θ, pobrano spośród gospodarstw 400 elementową próbkę losową i okazało się, że 144 właścicieli przekroczyło 60 lat. Zbudować przedział ufności dla parametru θ na poziomie ufności 0,95.

Zadanie 3.4 (Zadanie 6.7 z [2])
Analityk chce oszacować procent rynku mikrokomputerów opanowany przez IBM. Próba losowa złożona z 590 spółek używających mikrokomputery dała rezultat, że 500 spółek miało komputery IBM. Podać 95% przedział ufności dla procentu rynku opanowanego przez IBM.

3 Testowanie hipotez statystycznych - porównanie z normą

Boratyńska A., *Wykłady ze statystyki matematycznej dla II roku WNE*:

Model I
$X_1, X2, \ldots, X_n$ - próba losowa z rozkładu normalnego $N(\mu, \sigma^2)$, gdzie σ jest ZNANE.
Hipoteza zerowa: $H_0 : \mu = \mu_0$
Statystyka testowa: $U = \frac{\bar{X}-\mu_0}{\sigma}\sqrt{n}$ przy prawdziwej H_0 ma rozkład $N(0, 1)$.
Alternatywa:
$H_1 : \mu \neq \mu_0 \Rightarrow K_1 = \left(-\infty, -u(1-\frac{\alpha}{2})\right) \cup \left(u(1-\frac{\alpha}{2}), +\infty\right)$
$H_2 : \mu > \mu_0 \Rightarrow K_2 = (u(1-\alpha), +\infty)$
$H_3 : \mu < \mu_0 \Rightarrow K_3 = (-\infty, -u(1-\alpha))$

Model II
$X_1, X2, \ldots, X_n$ - próba losowa z rozkładu normalnego $N(\mu, \sigma^2)$, gdzie σ jest NIEZNANE.
Hipoteza zerowa: $H_0 : \mu = \mu_0$
$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ jest estymatorem σ^2
Statystyka testowa: $T = \frac{\bar{X}-\mu_0}{S}\sqrt{n}$ przy prawdziwej H_0 ma rozkład t-Studenta z (n-1) stopniami swobody.
Alternatywa:
$H_1 : \mu \neq \mu_0 \Rightarrow K_1 = \left(-\infty, -t(1-\frac{\alpha}{2})\right) \cup \left(t(1-\frac{\alpha}{2}), +\infty\right)$
$H_2 : \mu > \mu_0 \Rightarrow K_2 = (t(1-\alpha), +\infty)$
$H_3 : \mu < \mu_0 \Rightarrow K_3 = (-\infty, -t(1-\alpha))$

Model III
$X_1, X2, \ldots, X_n$ - próba losowa z próba losowa z rozkładu normalnego $N(\mu, \sigma^2)$, gdzie σ jest NIEZNANE.
Hipoteza zerowa: $H_0 : \sigma^2 = \sigma_0^2$
Statystyka testowa: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ przy prawdziwej H_0 ma rozkład chi-kwadrat z (n-1) stopniami swobody.
Alternatywa:
$H_1 : \sigma^2 \neq \sigma_0^2 \Rightarrow K_1 = \left(0, \chi_2^{n-1}(1-\frac{\alpha}{2})\right) \cup \left(\chi_2^{n-1}(\frac{\alpha}{2}, +\infty)\right)$
$H_2 : \sigma^2 > \sigma_0^2 \Rightarrow K_2 = (\chi_2^{n-1}(1-\alpha), +\infty)$
$H_3 : \sigma^2 < \sigma_0^2 \Rightarrow K_3 = (0, \chi_2^{n-1}(\alpha))$

Model IV
$X_1, X2, \ldots, X_n$ - próba losowa z próba losowa z rozkładu o nieznanej wartości oczekiwanej $E(X_i) = \mu$ i skończonej, ale nieznanej wariancji. Zakładamy, że n jest duże ($n \geq 100$).
Hipoteza zerowa: $H_0 : \mu = \mu_0$
Statystyka testowa: $U = \frac{\bar{X}-\mu_0}{S}\sqrt{n}$ przy prawdziwej H_0 ma asymptotyczny rozkład $N(0, 1)$
Alternatywa:
$H_1 : \mu \neq \mu_0 \Rightarrow K_1 = \left(-\infty, -u(1-\frac{\alpha}{2})\right) \cup \left(u(1-\frac{\alpha}{2}), +\infty\right)$
$H_2 : \mu > \mu_0 \Rightarrow K_2 = (u(1-\alpha), +\infty)$
$H_3 : \mu < \mu_0 \Rightarrow K_3 = (-\infty, -u(1-\alpha))$

Model V
Wykonujemy n niezależnych doświadczeń typu sukces-porażka, zakładamy, że n jest duże.
Obserwowana zmienna Y ma rozkład dwupunktowy $P(Y=1)=p, P(Y=0)=1-p$.
$p \in (0,1)$ - prawdopodobieństwo sukcesu w pojedynczym doświadczeniu, nieznany parametr nazywany też wskaźnikiem struktury.
Niech X oznacza liczbę sukcesów w n doświadczeniach.
Hipoteza zerowa: $H_0 : p = p_0$
$\hat{p} = \frac{X}{n}$ - estymator punktowy parametru p. Statystyka testowa: $U_* = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ przy prawdziwej H_0 ma asymptotyczny rozkład $N(0,1)$.
Alternatywa:
$H_1 : p \neq p_0 \Rightarrow K_1 = \left(-\infty, -u(1-\frac{\alpha}{2})\right) \cup \left(u(1-\frac{\alpha}{2}), +\infty\right)$
$H_2 : p > p_0 \Rightarrow K_2 = (u(1-\alpha), +\infty)$
$H_3 : p < p_0 \Rightarrow K_3 = (-\infty, -u(1-\alpha))$

Zadanie 4.1 (Zadanie 9.1 z [2])
Dzienne zużycie wody w pewnej fabryce podlega wahaniom losowym o rozkładzie normalnym z odchyleniem standardowym równym 50. Na podstawie 25 dni roku stwierdzono, że średnie dzienne zużycie wody wynosi 1025. Zweryfikować hipotezę, że wartość oczekiwana μ dziennego zużycia wody wynosi 1000 przy alternatywie, że $\mu > 1000$ na poziomie istotności 0.02. Podać postać obszaru krytycznego dla odpowiedniego testu.

Zadanie 4.2 (Zadanie 3.46 z [2])
Wylosowano niezależnie 12 indywidualnych gospodarstw rolnych w pewnej wsi i otrzymano dla nich następujące wielkości uzyskanych plonów owsa (w q/ha): 23.3, 22.1, 21.8, 19.9, 23.7, 22.3, 22.6, 21.5, 21.9, 22.8, 23.0, 22.2. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę, że wartość przeciętna plonu owsa w całej wsi wynosi 22.6 q/ha, jeśli alternatywną jest hipoteza, że wartość przeciętna plonu owsa jest wyższa niż w roku ubiegłym, w którym wynosiła 22.6 q/ha.

Zadanie 4.3 (Zadanie 3.49 z [2])
Zbadano 10 kawałków stali ze względu na na granicę plastyczności (w kG/cm^2) i otrzymano następujące wyniki: 3570, 3700, 3650, 3590, 3720, 3710, 3550, 3720, 3580, 3630. Zakładając, że granica plastyczności stali ma rozkład normalny, zweryfikować na poziomie istotności $\alpha = 0.05$ hipotezę H, że wartość przeciętna granicy plastyczności jest równa 3600, jeśli hipotezą alternatywną jest hipoteza $K : \mu \neq 3600$.

Zadanie 4.4 (Zadanie 3.52 z [2])
Maszyna mieszająca nawóz jest tak nastawiona, aby w każdych 100 kg nawozu było 10 kg azotanu. Zbadano dziesięć 100-kilogramowych worków. Procentowa zawartość azotanu była następująca: 9, 12, 11, 10, 11, 9, 11, 12, 9, 10. Czy na poziomie istotności $\alpha = 0.05$ można uważać za słuszną hipotezę, że wartość przeciętna zawartości azotanu w worku jest równa 10%, jeśli hipotezą alternatywną jest hipoteza, że ta wartość przeciętna jest wyższa niż 10%?

Zadanie 4.5 (Zadanie 3.53 z [2])
Wykonano 15 pomiarów czasu likwidowania zrywów na przędzarce obrączkowej, otrzymując (w s): 4.5, 3.6, 6.0, 6.4, 7.9, 6.9, 6.1, 7.4, 9.0, 4.3, 6.1, 8.2, 4.9, 7.5, 5.8. Zakładając, że rozkład czasu likwidacji zrywu jest normalny, na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę H, że wariancja czasu likwidacji zrywów jest równa 2, jeśli hipotezą alternatywną jest hipoteza $K : \sigma^2 > 2$.

4 Porównanie dwóch i więcej populacji

Zadanie 5.1 (Zadanie 3.58 z [1])
Dla porównania regularności uzyskiwanych wyników sportowych dwóch zawodników (skok w dal) w pewnym okresie wylosowano 12 wyników skoków dla pierwszego zawodnika oraz 9 wyników drugiego, otrzymując rezultaty (w m): dla pierwszego zawodnika: 7.60, 7.81, 8.01, 7.95, 7.15, 8.06, 7.90, 7.91, 7.56, 7.62, 7.85, 8.02; dla drugiego zawodnika: 7.50, 7.90, 8.00, 7.17, 7.28, 7.35, 7.73, 7.20, 7.98. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę o jednakowej regularności uzyskiwanych wyników dla obydwu zawodników (tzn. hipotezę, że wariancje rezultatów obydwu zawodników są równe), wobec hipotezy alternatywnej, że regularność pierwszego zawodnika jest wyższa.

Zadanie 5.2 (Zadanie 3.63 z [1])
Dwóm grupom robotników zlecono wykonanie tej samej pracy z tym jednak, że robotnicy grupy pierwszej przeszli wcześniej odpowiednie przeszkolenie. Zaobserwowana wydajność pracy w pierwszej grupie kształtowała się następująco (w szt/h): 18.6, 17.9, 18.1, 17.0, 18.7, 18.3, podczas gdy w drugiej grupie zaobserwowano następujące wydajności: 17.3, 17.6, 17.1, 16.0, 17.8. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę H, że średnia wydajność pracy nie zależy od uprzedniego przeszkolenia, jeśli alternatywą jest hipoteza, że średnia wydajność pracy robotników jest wyższa.

Zadanie 5.3 (Zadanie 3.64 z [1])
Spośród uczniów pewnego liceum wylosowano 15 z klas pierwszych oraz 12 z klas drugich i obliczono średnią ocen uzyskanych w semestrze dla każdego z uczniów. Otrzymano rezultaty: dla uczniów klas pierwszych: 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 uczniów klas drugich: 3.10, 3.38, 4.06, 3.60, 3.81, 4.50, 4.00, 3.25, 4.11, 4.85, 2.80, 4.00. Zakładając, że średnie wyniki ocen mają rozkłady normalne, zweryfikować na poziomie istotności $\alpha = 0.05$ hipotezę, że wartości przeciętne ocen uzyskiwanych przez uczniów klas pierwszej i drugiej są jednakowe, wobec hipotezy alternatywnej, że wartość przeciętna ocen uzyskiwanych przez uczniów klasy drugiej jest większa.

Zadanie 5.4 (Zadanie 3.77 z [1])
Z produkcji dzianin trzech rodzajów: wiskozy, poliamidu i anilany pobrano wycinki o wielkości $1m^2$ i wyznaczono ich masy (w g).
wiskoza: 122.4, 118.0, 120.0, 116.0, 120.8;
poliamid: 73.6, 73.4, 79.4, 73.9;
anilana: 254.7, 243.2, 248.6, 236.0, 245.6.
Zakładając, że rozkłady mas wycinków ($1m^2$) dzianin w każdym z trzech rodzajów surowca są normalne, na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę, że wariancje tych rozkładów są jednakowe.

Zadanie 5.5 (Zadanie 3.78 z [1])
Poddano badaniu na zginanie trzy różne rodzaje prętów stalowych i otrzymano rezultaty (w liczbie cykli zginających potrzebnych do złamania pręta):
19, 16, 22, 20, 23, 18, 16;
24, 21, 18, 24, 35, 33, 15;
54, 74, 43, 47, 60, 67, 52.
Przyjmując, że rozkład liczby cykli potrzebny do złamania pręta jest rozkładem logarytmiczno-normalnym na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę, że wariancje liczby cykli tych rodzajów prętów są jednakowe, wobec hipotezy alternatywnej, że wariancje te są różne.

Oznaczenia: m – średnia, σ – odchylenie standardowe

Dwie próby (1) i (2) – Testy istotności dla dwóch średnich $H_0 : m_1 = m_2$ $\qquad S_p^2 = (n_1 S_1^2 + n_2 S_2^2)/(n_1 + n_2 - 2)$

				Obszar krytyczny		
Model	Założenia	Stat. test.	Jej rozkład	$H_1 : m_1 \neq m_2$	$H_1 : m_1 > m_2$	$H_1 : m_1 < m_2$
A	Rozkład normalny nie znamy m, znamy σ	$U = (\overline{X}_1 - \overline{X}_2)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$N(0,1)$	$\|U\| > u_{1-\frac{\alpha}{2}}$	$U > u_{1-\alpha}$	$U < -u_{1-\alpha}$
B	Rozkład normalny nie znamy m ani σ	$T = (\overline{X}_1 - \overline{X}_2)/\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}$	$t(n_1 + n_2 - 2)$	$\|T\| > t_{1-\frac{\alpha}{2}}$	$T > t_{1-\alpha}$	$T < -t_{1-\alpha}$
C	Dowolny rozkł. dużo obs. nie znamy m ani σ	$U = (\overline{X}_1 - \overline{X}_2)/\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$N(0,1)$	$\|U\| > u_{1-\frac{\alpha}{2}}$	$U > u_{1-\alpha}$	$U < -u_{1-\alpha}$

Testy istotności dla frakcji – Model A: $H_0 : p = p_0$ $\quad \hat{p} = \frac{X}{n}$ $\qquad$ – Model B: $H_0 : p_1 = p_2$ $\quad \hat{p} = (X_1 + X_2)/(n_1 + n_2)$

Model	Założenia	Stat. test.	Jej rozkład	$H_1 : p \neq p_0$	$H_1 : p > p_0$	$H_1 : p < p_0$
A	Rozkład zero-jedynkowy	$U = (\hat{p} - p_0)/\sqrt{\frac{p_0(1-p_0)}{n}}$	$N(0,1)$	$\|U\| > u_{1-\frac{\alpha}{2}}$	$U > u_{1-\alpha}$	$U < -u_{1-\alpha}$
Model	Założenia	Stat. test.	Jej rozkład	$H_1 : p_1 \neq p_2$	$H_1 : p_1 > p_2$	$H_1 : p_1 < p_2$
B	Rozkład zero-jedynkowy	$U = (p_1 - p_2)/\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$	$N(0,1)$	$\|U\| > u_{1-\frac{\alpha}{2}}$	$U > u_{1-\alpha}$	$U < -u_{1-\alpha}$

Test Bartletta
Test Bartletta jest używany do zweryfikowania hipotezy zerowej o tym, że wariancje we wszystkich k populacjach są sobie równe przeciwko hipotezie alternatywnej, że przynajmniej jedna różni się od pozostałych.

Jeśli dysponujemy k próbkami o licznościach n_i oraz wariancjach w próbkach S_i^2 wówczas statystyka testowa w teście Bartletta ma postać:

$$\chi^2_{test} = \frac{(N-k)\ln(S_p^2) - \sum_{i=1}^{k}(n_i-1)\ln(S_i^2)}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}(\frac{1}{n_i-1}) - \frac{1}{N-k}\right)},$$

gdzie $N = \sum_{i=1}^{k} n_i$ oraz $S_p^2 = \frac{1}{N-k}\sum_i (n_i-1)S_i^2$.

Statystyka testowa ma asymptotyczny rozkład χ^2 z $k-1$ stopniami swobody. Hipoteza zerowa jest odrzucana jeśli $\chi^2_{test} > \chi^2_{k-1,\alpha}$ (gdzie $\chi^2_{k-1,1-\alpha}$ jest wartością krytyczną z rozkładu χ^2_{k-1}).

5 Bibliografia

[1] Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M., *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*, Tom I i II, Wydawnictwo Naukowe PWN, Warszawa 2010.

[2] Boratyńska A., *Zadania na ćwiczenia ze statystyki matematycznej.*

