Do zadań na hipotezy i estymację przedziałową ważne jest wypisanie danych!

strony (1)

POPULACJA	PRÓBA	
m	$\overline{x}$	← średnia
$б^2$	s^2	← wariancja
$б$	s	← odchylenie standardowe
$N(m,б)$ - jeżeli jest powiedziane, że dana cecha ma **rozkład normalny**, to zawsze dotyczy to populacji	n - liczebność próby, m - ilość obserwacji spośród próby	

Jeżeli w zadaniu nie ma nic na temat losowania próby to dane będą przyjmowane dla populacji, jeżeli jest wspomniane o próbie bądź losowaniu to dane są dla próby.

Populacja - wszyscy np. cała Polska

Próba - wylosowane obserwacje np. 1000 osób

Zad. 1

$m_0 = 100$ sztuk (norma, która może, ale nie musi być spełniona, potrzebna do hipotez)

POPULACJA

$N(m,б)$ - populacja ma rozkład normalny

PRÓBA

$n = 15$ osób

$\overline{x} = 94$ sztuk

współczynnik zmienności: $\boxed{V = \frac{s}{\overline{x}} \cdot 100\%}$

$V = 9\%$

można wyliczyć s

$9\% = \frac{s}{94} \cdot 100\%$

$s = \frac{9\% \cdot 94}{100\%} = 8,46$ sztuk

Formułujemy hipotezy (zawsze są dwie) ②

H_0: $m = \overset{m_0}{100}$ (hipoteza zerowa: średnia wydajność dla populacji wynosi 100, w H_0 zawsze musi być znak „=" niezależnie od treści zadania)

H_1: $m < \overset{m_0}{100}$ (hipoteza alternatywna: średnia wydajność jest niższa od 100, w H_1 występują znaki „>", „<", „≠" zależnie od treści zadania)

Zastosowanie testu istotności (hipotezy są dla średniej i nie ma cech jakościowych np. kolor) stosujemy PARAMETRYCZNE TESTY ISTOTNOŚCI dla średniej:

Model II (patrzymy na dane)

Populacja ma rozkład $N(m, \sigma)$, σ-nieznane, $n < 30$

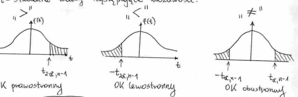
$$t_{obl.} = \frac{\bar{x} - m_0}{s} \cdot \sqrt{n-1}$$

$$t_{obl.} = \frac{94-100}{8{,}46} \cdot \sqrt{15-1} = \frac{-6}{8{,}46} \cdot \sqrt{14} \approx -2{,}653$$

Następnie odczytujemy z tablic wartość krytyczną t (test t-Studenta) dla α (poziom istotności) i $(n-1)$ stopni swobody i skonstruować obszar krytyczny.

W zadaniu mamy kilka α, więc będziemy czytać ~~dla~~ parę razy.

Obszar krytyczny (OK) ~~mamy~~ zależy od znaku w H_A. Dla testu t-Studenta mamy następujące możliwości:

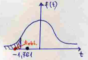


podwajamy α przy czytaniu z tablic

W zadaniu jest „<", zatem mamy OK lewostronny i będziemy czytać $t_{2\alpha, n-1}$ (minus się dopisuje)

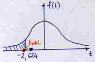
$\alpha = 0{,}05$

$t_{2\cdot 0{,}05;\,15-1} = t_{0{,}1;\,14} = 1{,}761$



$\alpha = 0{,}01$

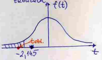
$t_{2\cdot 0{,}01;\,15-1} = t_{0{,}02;\,14} = 2{,}624$



$\alpha = 0{,}02$ (3)

$t_{2\cdot 0{,}02;\,15-1} = t_{0{,}04;\,14} = 2{,}145$

czytam dla $\alpha = 0{,}05$ bo $\alpha = 0{,}04$ nie ma w tablicach



Teraz szukamy, gdzie znajduje się $t_{obl.} = -2{,}653$.

Jeżeli $t_{obl.}$ należy do obszaru krytycznego to odrzucamy H_0 i przyjmujemy H_1, OK – ▨.

Jeżeli $t_{obl.}$ <u>nie</u> należy do OK to H_0 przyjmujemy i odrzucamy H_1.

- dla $\alpha = 0{,}05$ $t_{obl.} \in OK$ zatem H_1 jest prawdziwa
- dla $\alpha = 0{,}01$ $t_{obl.} \in OK$ zatem H_1 jest prawdziwa
- dla $\alpha = 0{,}02$ $t_{obl.} \in OK$ zatem H_1 jest prawdziwa

czyli w każdym przypadku średnia jest niższa od 100.

Odp. b i e

Zad. 2

Funkcja $f(x)$ jest gęstością zmiennej losowej jeżeli:

$$\boxed{\int_{-\infty}^{+\infty} f(x)\,dx = 1}$$

$$f(x) = \begin{cases} 0, \text{ dla } x \leqslant 6 \rightarrow x \in (-\infty, 6\rangle \\ \frac{1}{14}x, \text{ dla } 6 < x \leqslant a \rightarrow x \in (6, a\rangle \\ 0, \text{ dla } x > a \rightarrow x \in (a, +\infty) \end{cases}$$

Funkcja $f(x)$ jest w „trzech kawałkach" więc całkę też musimy podzielić na trzy części:

przy czym $\int_a^b 0\,dx = 0$

$$\underbrace{\int_{-\infty}^{6} 0\,dx}_{0} + \int_6^a \frac{1}{14}x\,dx + \underbrace{\int_a^{+\infty} 0\,dx}_{0} = 1$$

$0 + \int_6^a \frac{1}{14} x\,dx + 0 = 1$ (4)

$\int_6^a \frac{1}{14} x\,dx = 1$

$\frac{1}{14}\int_6^a x\,dx = 1$, bo $\int a \cdot x^n dx = a \cdot \int x^n dx$ (stałą wyciągamy przed całkę)

$\frac{1}{14} \cdot \left[\frac{1}{2}x^2\right]_6^a = 1$, bo $\int_a^b x^n dx \underset{x=x^1}{=} \left[\frac{1}{n+1}x^{n+1}\right]_a^b$

$\frac{1}{14} \cdot \left(\frac{1}{2}a^2 - \frac{1}{2}\cdot 6^2\right) = 1$, bo $[F(x)]_6^a = F(a) - F(b)$

$\frac{1}{14} \cdot \left(\frac{1}{2}a^2 - 18\right) = 1 \mid \cdot 14$

$\frac{1}{2}a^2 - 18 = 14$

$\frac{1}{2}a^2 = 14 + 18$

$\frac{1}{2}a^2 = 32 \mid \cdot 2$

$a^2 = 64 \Rightarrow a = \sqrt{64} \Rightarrow a = -8 \vee a = 8$, ale przedziały były od 6 więc $\underline{a = 8}$

Odp. 6

Zad. 3

POPULACJA	PRÓBA
	$n = 120$ osób
	$m = 45$ osób (45 osób spośród próby zgadza się na budowę elektrowni atomowej)
	$p = \frac{m}{n} = 0{,}375$ (częstość - już dana) ↓ wskaźnik struktury = odsetek = procent

Podana też jest statystyka $u_{obl.} = 1{,}793$ (więc nie trzeba szukać wzoru)

$p_0 = 30\% = 0{,}3$ (założony sprzeciw i wcale nie musi taki być, potrzebny do $H_{0,1}$)

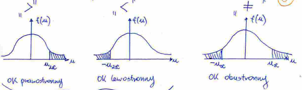
Formułujemy hipotezy:

$H_0: p = \overset{p_0}{0{,}3}$

$H_1: p > 0{,}3$

Podano już $u_{obl} = 1{,}793$, zatem określamy OK (obszar krytyczny)
u - rozkład normalny

Obszar krytyczny w przypadku rozkładu normalnego wygląda następująco (też zależy od H_1): (5)

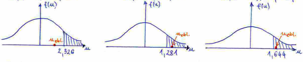


podwajamy α przy czytaniu z tablic

W zadaniu jest „>", zatem mamy OK prawostronny i czytamy z tablic $u_{2\alpha}$ (dla kilku α)

$\alpha = 0{,}01$ $\quad$ $\alpha = 0{,}1$ $\quad$ $\alpha = 0{,}05$

$u_{2\cdot 0{,}01} = u_{0{,}02} = 2{,}326$ $\quad$ $u_{2\cdot 0{,}1} = u_{0{,}2} = 1{,}281$ $\quad$ $u_{2\cdot 0{,}05} = u_{0{,}1} = 1{,}644$



Szukamy, gdzie jest $u_{obl.} = 1{,}793$

Jeżeli $u_{obl.}$ należy do OK to H_0 odrzucamy i przyjmujemy H_1.

Jeżeli $u_{obl.}$ nie należy do OK to H_0 przyjmujemy (odrzucamy H_1)

- dla $\alpha = 0{,}01$ $u_{obl.} \notin OK \rightarrow$ przyjmujemy H_0 (sprzeciw wynosi 30%)
- dla $\alpha = 0{,}1$ $u_{obl.} \in OK \rightarrow$ przyjmujemy H_1 (sprzeciw jest większy od 30%)
- dla $\alpha = 0{,}05$ $u_{obl.} \in OK \rightarrow$ przyjmujemy H_1 (sprzeciw jest większy od 30%)

Odp. c, e

Zad. 4

x_i	1	2	3	4	5
$P\{X \leqslant x_i\}$	0,1	z	0,4	0,9	1

Jeżeli podane jest $P\{X \leqslant x_i\}$ to wartości te są dystrybuantą (skumulowane prawdopodobieństwo), zwłaszcza jeżeli dodamy do siebie te wartości to otrzymamy więcej niż 1, po tym też można to poznać.

6

x_i	p_i sk.	p_i
1	0,1	0,1 - pierwsze zawsze takie same
2	z	z - 0,1
3	0,4	0,4 - z
4	0,9	0,9 - 0,4 = 0,5
5	1	1 - 0,9 = 0,1
Σ	X	1

wartości w dystrybuancie zawsze rosną i kończą się na 1

$1-(0,1+0,5+0,1) = 1-0,7 = 0,3$

najprościej

$\underset{\text{przed } z}{z = 0,4} - \underset{\text{przed } z}{0,1} = 0,3$

Odp. a

Zad. 5

$X \sim N(\underset{m}{15}, \underset{6}{3})$

$P(X > x_0) = 0,8414$

Trzeba wystandaryzować zmienną $X \sim (15,3)$ do rozkładu o parametrach $U \sim N(0,1)$ za pomocą wzoru:

$$\boxed{X \sim N(m, 6) \xrightarrow{U = \frac{X-m}{6}} U \sim N(0,1)}$$

$P\left(\frac{X-m}{6} > \frac{x_0 - m}{6}\right) = 0,8414$

$P\left(U > \frac{x_0 - 15}{3}\right) = 0,8414$

z własności prawdopodobieństwa:

$P(U > a) = 1 - F(a)$, zatem:

$1 - F\left(\frac{x_0-15}{3}\right) = 0,8414$

$-F\left(\frac{x_0-15}{3}\right) = 0,8414 - 1$

$-F\left(\frac{x_0-15}{3}\right) = -0,1586 \quad | \cdot (-1)$

$F\left(\frac{x_0-15}{3}\right) = 0,1586$

Szukamy w tablicach dystrybuanty rozkładu normalnego wartości 0,1586 (w środku)

$$\frac{x_0-15}{3} = -1 \quad | \cdot 3$$

$$x_0-15=-3$$

$$x_0=-3+15$$

$$x_0=12$$

Odp. b

$\boxed{\text{Zad.6}}$

ile należy wylosować $\longrightarrow$ MINIMALNA LICZEBNOŚĆ PRÓBY

$p=10\%=0{,}1$ - spodziewany odsetek

$q=1-p=1-0{,}1=0{,}9$

$d=2\%=0{,}02$ - błąd szacunku

$\alpha=0{,}1$

Minimalna liczebność próby dla odsetka:

$$\boxed{n=\frac{u_\alpha^2\cdot p\cdot q}{d^2}}$$

$$n=\frac{u^2_{0,1}\cdot 0{,}1\cdot 0{,}9}{(0{,}02)^2}$$ gdzie $u_{0,1}=1{,}644$ odczytane z tablic rozkładu normalnego

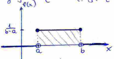
$$n=\frac{(1{,}644)^2\cdot 0{,}1\cdot 0{,}9}{0{,}0004}=608{,}12\approx 609$$ (wynik zawsze zaokrąglamy w górę)

Odp. a

$\boxed{\text{Zad.7}}$

Rozkład jednostajny (prostokątny) charakteryzuje się następującą postacią: na przedziale $\langle a,b\rangle$

$$f(x)=\begin{cases}0, & \text{dla } x<a\\ \frac{1}{b-a}, & \text{dla } a\leqslant x\leqslant b\\ 0, & \text{dla } x>b\end{cases}$$



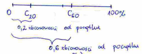
$P=1$

$\langle \overset{a}{3},\overset{b}{18}\rangle$

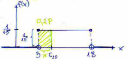
$$f(x)=\begin{cases}0, & \text{dla } x<3\\ \frac{1}{18-3}, & \text{dla } 3\leqslant x\leqslant 18\\ 0, & \text{dla } x>18\end{cases} \Leftrightarrow \begin{cases}0, & \text{dla } x<3\\ \frac{1}{15}, & \text{dla } 3\leqslant x\leqslant 18\\ 0, & \text{dla } x>18\end{cases}$$

I sposób (korzystając z wykresu i pola prostokąta)

Szukamy $C_{60}-C_{20}$ (centyle dzielą zbiorowość na 100 części)



Centyl 20 to 20% pola prostokąta licząc od początku

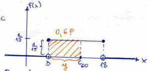


$P = \frac{1}{15}\cdot x$ $\qquad C_{20} = 3 + x$ (z rys.)

$0{,}2 = \frac{1}{15}\cdot x \,|\cdot 15$ $\qquad C_{20} = 3+3 = \underline{6}$

$x = 3$

Centyl 60 to 60% pola prostokąta licząc od początku



$P = \frac{1}{15}\cdot y$ $\qquad C_{60} = 3 + y$

$0{,}6 = \frac{1}{15}\cdot y \,|\cdot 15$ $\qquad C_{60} = 3+9 = \underline{12}$

$y = 9$

$$C_{60} - C_{20} = 12 - 6 = 0$$

II sposób (korzystając z całek)

nie bierzemy pod uwagę wartości 0 z funkcji, bo i tak dają zero.

$$\int_3^{C_{20}} \frac{1}{15}\,dx = 0{,}2$$

$\frac{1}{15}\int_3^{C_{20}} dx = 0{,}2$, bo $a\int x^n dx = \int a x^n dx$ (wyciągamy stałą)

$\frac{1}{15}\left[x\right]_3^{C_{20}} = 0{,}2$, bo $\int_a^b dx = \left[x\right]_a^b$

$\frac{1}{15}(C_{20}-3)=0{,}2 \,|\cdot 15$, bo $[f(x)]_e^b = f(b)-f(e)$ 9

$C_{20}-3=3$

$C_{20}=3+3=6$

$\int_3^{C_{60}} \frac{1}{15}dx = 0{,}6$

$\frac{1}{15}\int_3^{C_{60}} dx = 0{,}6 \Leftrightarrow \frac{1}{15}\cdot[x]_3^{C_{60}} = 0{,}6 \,|\cdot 15 \Leftrightarrow [x]_3^{C_{60}} = 9 \Leftrightarrow$

$\Leftrightarrow C_{60}-3=9 \Leftrightarrow C_{60}=9+3=12$

$C_{60}-C_{20}=12-6=6$

Odp. e

$\boxed{\text{Zad. 8}}$

tempo zmian średnie $\boxed{\bar{i} = \sqrt[n-1]{\frac{y_n}{y_0}}}$ n - ilość okresów
y_n - wartość z ostatniego okresu
y_0 - wartość z pierwszego okresu

$\boxed{\bar{T} = (\bar{i}-1)\cdot 100\%}$

$n = 5$ kwartałów

$y_5 = 15{,}3$

$y_0 = 13{,}0$

$\bar{i} = \sqrt[5-1]{\frac{15{,}3}{13{,}0}} = \sqrt[4]{\frac{15{,}3}{13{,}0}} \approx 1{,}0416$

$\bar{T} = (1{,}0416-1)\cdot 100\% = 0{,}0416\cdot 100\% = \underline{4{,}16\%}$

Odp. c

$\boxed{\text{Zad. 9}}$

PRZEDZIAŁ UFNOŚCI - ESTYMACJA

POPULACJA	PRÓBA
	$n = 100$ osób $m = 36$ osób (36 osób spośród próby mieszka na stałe w Białymstoku)

przedział ufności: [26,592% ; 45,408%]

[0,26592 ; 0,45408] → zamiana % na ułamki

przedział ufności był wyznaczony dla ~~odsetka~~ = wskaźnika struktury,

zatem szukamy we wzorach wzoru na przedział ufności dla wskaźnika struktury (p):

$$P\left(\underbrace{\frac{m}{n}-u_\alpha\cdot\sqrt{\frac{\frac{m}{n}\left(1-\frac{m}{n}\right)}{n}}}_{0,26592}<p<\underbrace{\frac{m}{n}+u_\alpha\cdot\sqrt{\frac{\frac{m}{n}\left(1-\frac{m}{n}\right)}{n}}}_{0,45408}\right)=\underbrace{1-\alpha}_{\text{współczynnik ufności}}$$

Wybieramy do-wolną końcówkę przedziału, bo i tak były szacowane dla tego samego współczynnika ufności:

np.

$$\frac{m}{n}-u_\alpha\cdot\sqrt{\frac{\frac{m}{n}\left(1-\frac{m}{n}\right)}{n}}=0,26592$$

$$\frac{m}{n}=\frac{36}{100}=0,36$$

$$0,36-u_\alpha\cdot\sqrt{\frac{0,36(1-0,36)}{100}}=0,26592$$

$$0,36-u_\alpha\cdot\sqrt{\frac{0,36\cdot0,64}{100}}=0,26592$$

$$0,36-u_\alpha\cdot0,048=0,26592$$

$$-u_\alpha\cdot0,048=0,26592-0,36$$

$$-u_\alpha\cdot0,048=-0,09408 \quad |:(-0,048)$$

$u_\alpha = 1,96$ – szukamy w tablicach rozkładu normalnego 1,96 (w środku) i odczytujemy α

$\alpha = 0,05$, bo $u_{0,05}=1,96$

współczynnik ufności $1-\alpha = 1-0,05 = 0,95$

Odp. d

Zad. 10

$\bar{x} = 10$ szt./h (średnia)

$d_o = 12$ sztuk/h (wartość najczęściej występująca to dominanta - do)

$s = 4$ szt./h (odchylenie standardowe)

Będziemy analizować kolejne stwierdzenia licząc odpowiednie wielkości.

• zmienność wydajności
↓
współczynnik zmienności: $\boxed{V = \frac{s}{\bar{x}} \cdot 100\%}$

$V = \frac{4}{10} \cdot 100\% = 40\% \neq 33\%$

• asymetria → aby określić jej kierunek i siłę oraz dysponując $\bar{x}$, s, do korzystamy ze wzoru:

$\boxed{A = \frac{\bar{x} - do}{s}}$

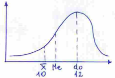
$A = \frac{10-12}{4} = \frac{-2}{4} = -0,5$

Znak wyniku określa kierunek asymetrii: $\begin{cases} + \leftarrow \text{prawostronna} \\ - \leftarrow \text{lewostronna} \end{cases}$

Sama liczba oznacza siłę:
0 – 0,4 – asymetria słaba
0,4 – 0,7 – asymetria niezbyt silna
0,7 – 1 – asymetria silna
} to skala orientacyjna

• Jeżeli wiemy, jaką asymetrią charakteryzuje się rozkład, to można określić ~~ich~~ (Me, do, $\bar{x}$) ~~są~~ ułożenie tych miar

Asymetria lewostronna:

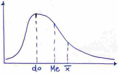


$\bar{x} < Me < do$

Odp. b, c

Zad. 11 ?

asymetria prawostronna



$do < Me < \bar{x}$

Tutaj podpasowuję dane do asymetrii lewostronnej – mogą być dowolne

Średnia z danych indywidualnych np. 2, 4, 4, 4, 4, 5, 6, 7, 9 12

$$\bar{x} = \frac{2+4+4+4+4+5+6+7+9}{9} = \frac{45}{9} = 5 \quad do = 4 \quad do < \bar{x}$$

tworzę z tych danych szereg rozdzielczy:

$\langle x_{0i}, x_{1i})$	n_i	$\mathring{x}_i$ - środek klas	$\mathring{x}_i \cdot n_i$
2-4	1	3	3
4-6	5	5	25
6-8	2	7	14
8-10	1	9	9
Σ	9	$\times$	51

$$\mathring{x}_i = \frac{\text{początek + koniec}}{2}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} \mathring{x}_i \cdot n_i}{N} = \frac{51}{9} \approx 5{,}67$$

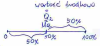
Zmieniłam też wielkość danych indywidualnych tak, aby zbliżały się do początku przedziału, końca, ale tak aby szereg rozdzielczy był taki sam, za każdym razem średnia była z danych indywidualnych niższa, ale trzeba by było to sprawdzić w różnych przykładach.

Także nie jestem pewna: !

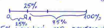
Odp c lub d

Zad. 12

$Me = 54{,}7\ m^2$ (połowa zbiorowości - mediana)

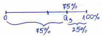


$Q_1 = 41{,}1\ m^2$ (kwartyl pierwszy inaczej dolny)



25% mieszkań ma powierzchnię niższą od $41{,}1\ m^2$, 75% ma wyższą

$Q_3 = 67{,}9\ m^2$ (kwartyl trzeci inaczej górny)



75% mieszkań ma powierzchnię niższą od $67{,}9\ m^2$, 25% ma wyższą

- zróżnicowanie = zmienność

pozycyjny współczynnik zmienności (jeżeli dysponujemy Q_1, Me, Q_3)

$$\boxed{V_p = \frac{Q}{Me} \cdot 100\%} \quad \text{gdzie} \quad \boxed{Q = \frac{Q_3 - Q_1}{2}}$$

$V_p \Rightarrow \quad Q = \frac{67,9-41,1}{2} = \frac{26,8}{2} = 13,4$

$$V_p = \frac{13,4}{54,7} \cdot 100\% \approx 24,5\%$$

- współczynnik asymetrii gdy dysponujemy (Q_1, Q_3, Me)

$$\boxed{A = \frac{Q_3 + Q_1 - 2 \cdot Me}{2 \cdot Q}}$$

$$A = \frac{67,9 + 41,1 - 2 \cdot 54,7}{2 \cdot 13,4} = \frac{-0,4}{26,8} \approx -0,015 \approx \text{około } 0$$

- zakres wartości <u>typowych</u> – typowy obszar zmienności, jeżeli mamy Q_1, Me, Q_3 to:

$$\boxed{Me - Q < x_{typ.} < Me + Q}$$

$$54,7 - 13,4 < x_{typ.} < 54,7 + 13,4$$

$$41,3 < x_{typ.} < 68,1$$

Odp. **a, b**

Zad. 13

wartość produkcji $\Rightarrow W_{2003} = 1,5$ mld zł

indeks dynamiki zatrudnienia $\Rightarrow i_{\frac{2004}{2000}} = 0,6 \quad (60\%)$

$i_{\frac{2003}{2000}} = 0,8 \quad (80\%)$

$W_{2004} = ?$

Produkcja nie ma nic wspólnego z wielkością zatrudnienia, mowa o bezpośrednim związku, na wartość produkcji bezpośredni wpływ ma jedynie cena i ilość. Musielibyśmy mieć podane indeksy wartości produkcji:

$$i_{\frac{04}{03}} = \frac{\text{wartość produkcji 2004}}{\text{wartość produkcji 2003}}$$

jak widać nie ma ich, zatem za mało danych

Odp. **d**

$\boxed{\text{Zad. 14}}$ 14

przed zmianami

ilość dzieci x_i	ilość n_i	$x_i \cdot n_i$
0	x tyle samo	0
2	x	$2x$
$\sum$ (suma)	$2x = N$	$2x$

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i \cdot n_i}{N} = \frac{2x}{2x} = 1$$

po zmianach

ilość dzieci x_i	ilość n_i	$x_i \cdot n_i$
0	x	0
1	2	2
2	x	$2x$
$\sum$	$2x+2 = N$	$2x+2$

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i \cdot n_i}{N} = \frac{2x+2}{2x+2} = 1$$

Odp. C