Statystyka

Matematyczna

Zaliczenie:

Kolokwium: po zajęciach / na ostatnich

1.1. X_1, X_2, X_3, X_4, $X_i \sim N(1,1)$

$P(|X|>6)$, $X = X_1 + X_2 + X_3 + X_4$

$\mu = 1$, $\sigma = 1$

$E(X) = E(X_1) + E(X_2) + E(X_3) + E(X_4)$

$D^2(X) = D^2(X_1) + \ldots = 4$

$D(X) = 2$

$X \sim N(4,2)$

$Z \sim N(0,1)$

$P(|X|>6) = 1 - P(|X| \leq 6) =$

$1 - P(-6 \leq X \leq 6) = 1 - P\left(\frac{-6-4}{2} \leq \frac{X-4}{2} \leq \frac{6-4}{2}\right)$

$= 1 - P(-5 \leq Z \leq 1) = 1 - [\Phi(1) - \Phi(-5)]$ Z

$\Phi(-5) = 1 - \Phi(5) = 0$

$\dots = 1 - \Phi(1) = 1 - 0{,}8413 = 0{,}1587$

1.2. $X \sim N(1, \sqrt{2}), Y \sim N(4, \sqrt{7})$

$P(X+Y > 0)$

$U = N(1+4, \sqrt{2} + \sqrt{7}) = N(5, 3)$

$U = \alpha_1 X_1 + \alpha_2 X_2$

$P(U>0) = 1 - P(U<0) = 1 - P(Z < -\frac{5}{3}) =$

$= 1 - \Phi(-\frac{5}{3}) = 1 - 1 + \Phi(\frac{5}{3}) = \Phi(1{,}67) =$

$= 0{,}9525$

1.2 c) $X \sim N(1, \sqrt{2}), Y \sim N(4, \sqrt{7})$

$P(3X + 4Y > 20)$

$$U = 3X + 4Y \sim N(3\cdot 1 + 4\cdot 4, \sqrt{3^2\cdot 2 + 4^2\cdot 7}) =$$

$$= N(19, \sqrt{130}) \approx N(19, 11{,}4)$$

$$Z = \frac{U - 19}{11{,}4} \sim N(0,1)$$

$$P(U > 20) = 1 - P(U \leq 20) = 1 - P\left(Z \leq \frac{20-19}{11{,}4}\right) =$$

$$= 1 - P\left(\frac{1}{11{,}4}\right) = 1 - \Phi(0{,}09) = 0{,}5349 =$$

$$= 0{,}4641$$

CENTRALNE TWIERDZENIE GRANICZNE

PRÓBA > 30

1. 5. $X_1, X_2 \ldots X_{180}$ są niezależne i mają identyczne rozkłady o dystrybuancie:

$$F(x) = \begin{cases} 0 & \text{dla } x \le 0 \\ x^2 & \text{dla } 0 < x \le 1 \\ 1 & \text{dla } x > 1 \end{cases}$$

Oblicz $P(Y < 115)$, $Y = \sum_{1}^{180} X_n$

CTG $\quad Z_n = \dfrac{Y - 180\mu}{\sigma\sqrt{180}}$

$$\lim_{n\to\infty} F_n(z) = \lim_{n\to\infty} P(Z_n < z) = \Phi(z)$$

$$P\left(a < \frac{\sum_{1}^{180} x_n}{\sigma\sqrt{n}} < b\right) \approx \Phi(b) - \Phi(a)$$

$$f(x) = \begin{cases} 2x & \text{dla } 0 < x < 1 \\ 0 & \text{dla pozost. } x. \end{cases}$$

verte →

$$E(x) = \int_0^1 2x^2 = \frac{2}{3}$$

$$D^2(x) = E(x^2) - [E(x)]^2$$

$$E(x^2) = \int_0^1 2x^3 dx = \frac{1}{2}$$

$$D^2(x) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

$$\mu = E(x) = \frac{2}{3}, \quad \sigma = \sqrt{\frac{1}{18}} = \frac{\sqrt{2}}{6}$$

$$P\left(\frac{Y - 180 \cdot \frac{2}{3}}{\frac{\sqrt{2}}{6} \cdot \sqrt{180}} < \frac{115 - 180 \cdot \frac{2}{3}}{\frac{\sqrt{2}}{6} \cdot \sqrt{180}}\right) =$$

$$P\left(Z < \frac{-5}{\sqrt{10}}\right) = P(Z < -1,58) =$$

$$= \Phi(-1,58) = 1 - \Phi(1,58) = 1 - 0,9429 =$$

$$= 0,0571$$

1.8. POPRAWKA NA CIĄGŁOŚĆ

$X_1 \dots X_{30}$ niezalezne, z tym samym rozkładem Poissona o wartości $2/3$. Oblicz:

a) $P(15 < X_1 + X_2 + \dots < 22)$

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0,1,2\dots$$

$$E(X_i) = \frac{2}{3}$$

λ - parametr rozkładu

$$E(X) = \lambda, \quad D^2(X) = \lambda$$

Poprawka na ciągłość — poprawka ze skokowego na normalny

$$P(15 < X < 22) = P(15{,}5 < Y < 22{,}5)$$

$\lambda = \frac{2}{3}$, $\mu = \frac{2}{3}$, $n = 30$, $n\mu = 20$,

$\sigma\sqrt{n} = \sqrt{\frac{2}{3}}\sqrt{30} = \sqrt{20} \approx 4{,}47$

$P(15 < Y \leqslant 22) \overset{\text{Pop. na ciągłość}}{=} P(15{,}5 < Y < 22{,}5) =$

$= P\left(\frac{15{,}5-20}{4{,}47} < Z < \frac{22{,}5-20}{4{,}47}\right) =$

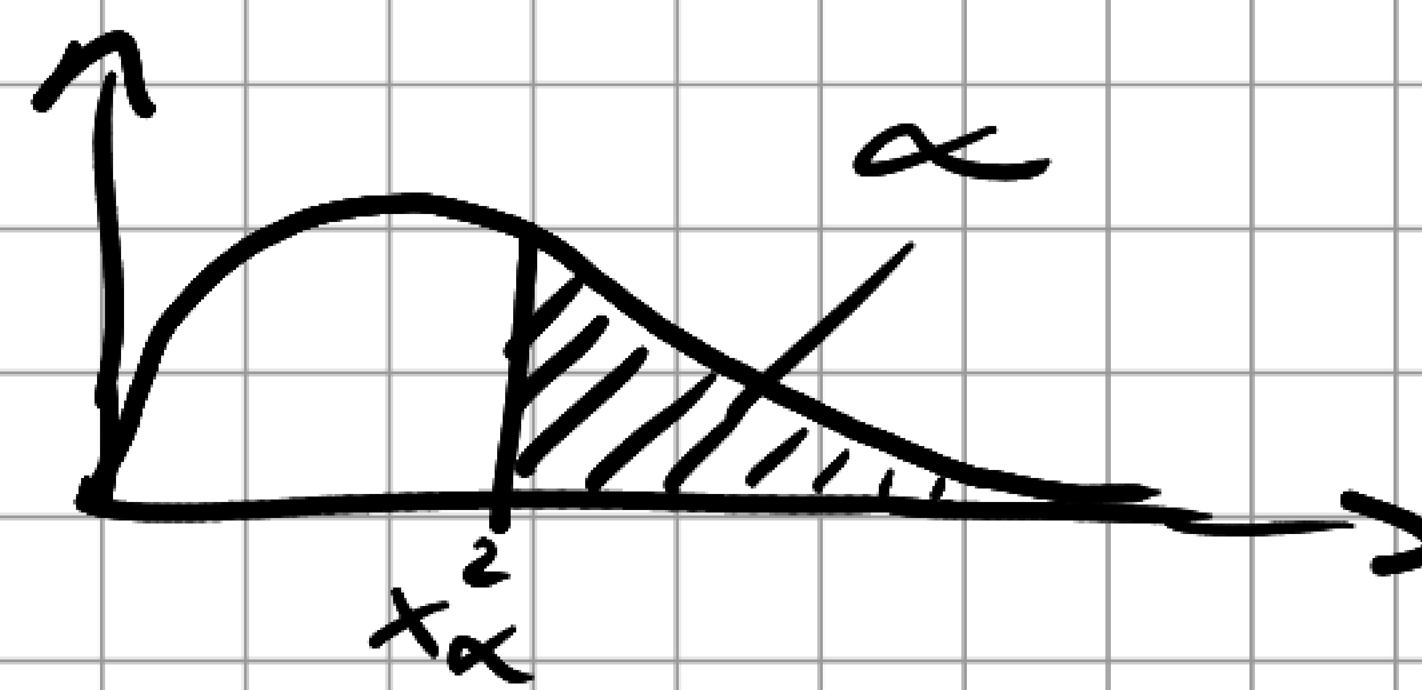
$= P(-1{,}01 < Z < 0{,}56) = \Phi(0{,}56) - \Phi(-1{,}01) =$

$= \Phi(0{,}56) - 1 + \Phi(1{,}01)$

1.12. $P(9{,}54 \le Y_{22} < c) = 0{,}09$

$Y_{22} - \chi^2(22)$, $c = ?$

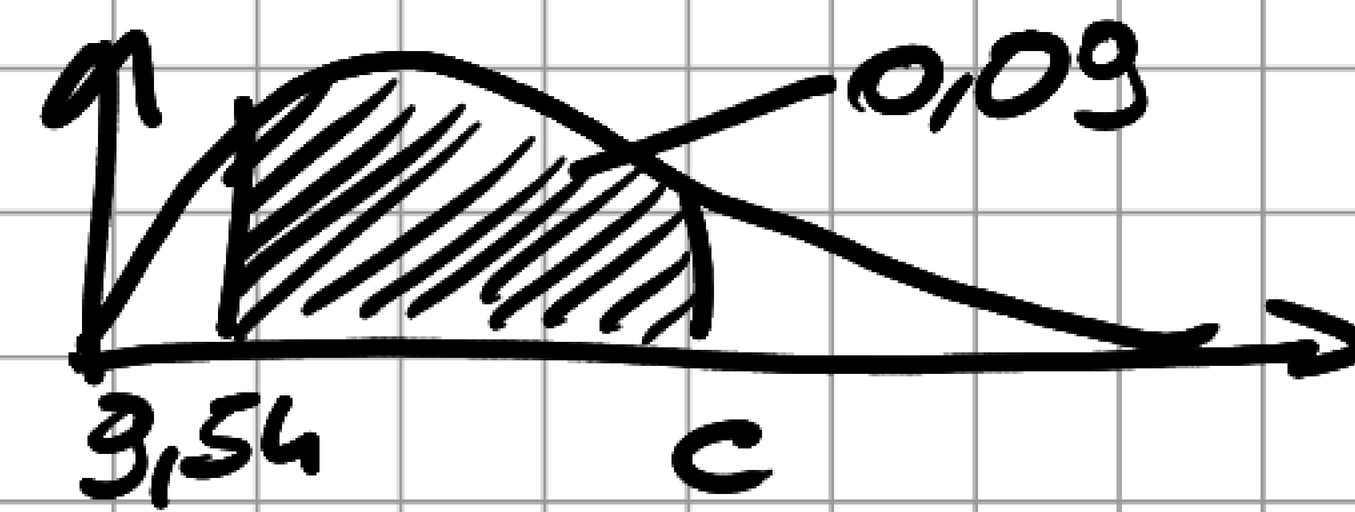
$\chi^2_\alpha : P(Y \geq \chi^2_\alpha) = \alpha$



$P(9{,}54 \le Y_{22} \le c) =$

$= P(Y_{22} \geq 9{,}54) - P(Y_{22} \geq c) =$

$= 0{,}09.$



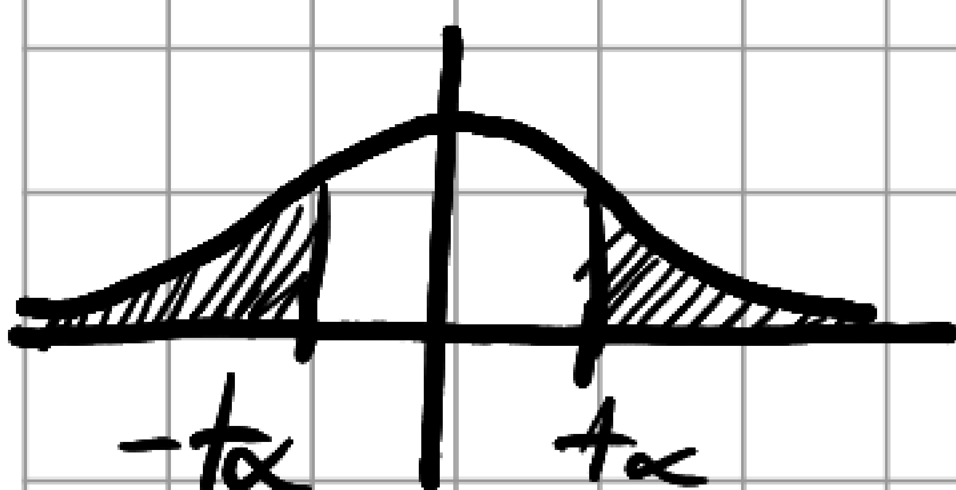
$P(Y_{22} \geq c) = P(Y_{22} \geq 9{,}54) - 0{,}09 = 0{,}39 - 0{,}09 = 0{,}3$

$c = 14{,}04$

Chi^2 tablica

Rozkład t-studenta

1.14. $c = ?$ $P(-c \le P_{22} \le c) = 0{,}98$



$P(|T| \ge t_\alpha) = \alpha$

$P \le -t_\alpha \vee P \ge t_\alpha$

$P(-c \le P_{22} \le c) = P(|P_{22}| \le c) =$

$= 1 - P(|P_{22}| \ge c)$

$$P(|P_{22}| \ge \underset{\substack{\uparrow \\ t_\alpha}}{c}) = \underset{\substack{\uparrow \\ \alpha}}{0{,}02}$$

z t. studenta

	0.02
22	2,51

2.8 Obliczanie wartości z próby

B $-2, 7, -3, 5, 6, 4$

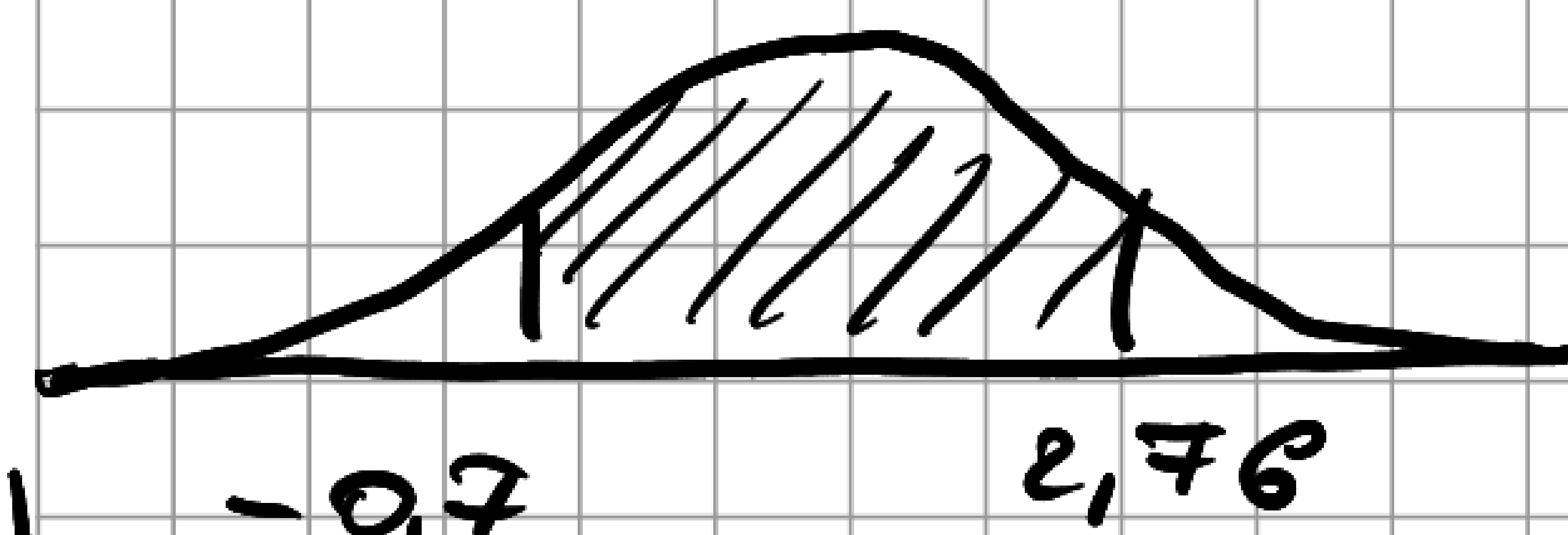
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{17}{6}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left[\sum x_i^2 - n\overline{x}^2 + n\overline{x}^2\right] =$$

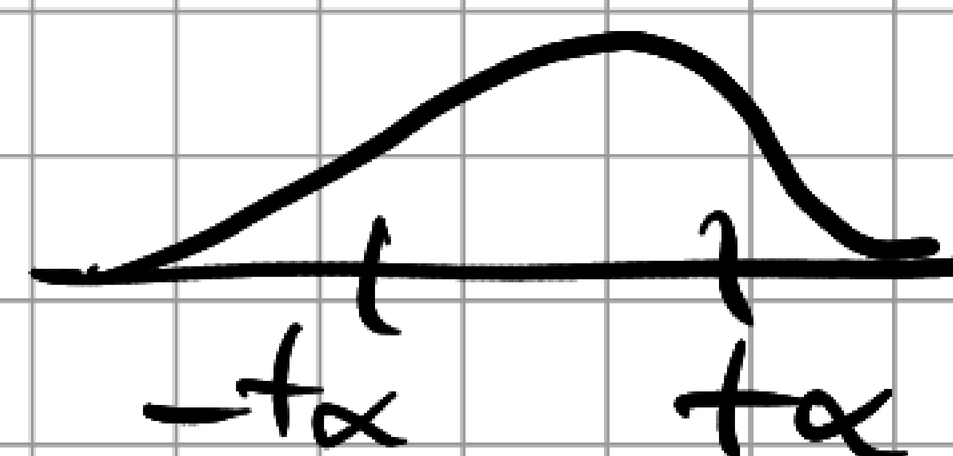
$$= \frac{1}{n-1}\sum x_i^2 - \frac{n}{n-1}\overline{X}^2 = \dots$$

$1.\ 13.\ 6)\ P(-0{,}7 < P_{10} < 2{,}76)$

t - studenta



$t(\alpha):\ P(|P| \geq t\alpha) = \alpha$



$$P(|P| \geq t\alpha) = 1 - P(|P| < t\alpha) =$$

$$= 1 - P(-t\alpha < P < t\alpha) = \alpha$$

$$P(-t\alpha < P < t\alpha) = 1 - \alpha$$

$$\hookrightarrow 1 - (P_{10} < -0{,}7) - P(P_{10} > 2{,}76) =$$

$$= -\tfrac{1}{2}(|P_{10}| \geq 0{,}7) - \tfrac{1}{2}(|P_{10}| \geq 2{,}76) =$$

$$= 1 - \tfrac{1}{2} \cdot 0{,}5 - \tfrac{1}{2} \cdot 0{,}02 = 0{,}74$$

1.15. a) oszacuj $P(\rho_{12} > 2{,}9)$

$$P(\rho_{12} > 2{,}9) = \frac{1}{2} P(|\rho_{12}| > 2{,}9)$$

$$0{,}01 < P(\rho_{12} \geq 2{,}90) < 0{,}02$$

$$0{,}005 < P(\rho_{12} > 2{,}9) < 0{,}01$$

Estymator wariancji $E(s) = \sigma^2$

2.1 L1 $X_1, X_2, X_3 \dots X_n$

a) $E(X_i^2) = \mu^2 + \sigma^2$

$D^2(X_i) = E(X_i^2) - [E(X_i)]^2$

$\sigma^2 = E(X_i^2) - \mu^2 \Rightarrow E(X_i^2 = \mu^2 + \sigma^2$

$\overline{X} : E(\overline{x}) = \mu$

$E(\overline{x}) = \frac{1}{n} \cdot n \cdot \mu = \mu$

jeśli zmienne są niezależne, to wariancja sumy jest sumą wariancji

$D^2(x) = \left(\frac{1}{n}\right)^2 \cdot n \cdot \sigma^2 = \frac{\sigma}{n}$

$D^2(\overline{x}) = E(\overline{x}^2) - [E(\overline{x})]^2$

2.2. Udowodnić, że $E(s^2) = \sigma^2$.

$E(X_i) = \mu$, $D^2(x) = \sigma$

$E\left(\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{x})^2\right) =$

$= E\left(\frac{1}{n-1}\left(\sum_{k=1}^{n} x_k^2\right) - \frac{1}{n-1}(\overline{x})^2\right) =$

$= \frac{1}{n-1} E\left(\sum_{k=1}^{n}(X_k^2) + \ldots \sum(X_n^2)\right) - nE(\overline{x}^2) =$

$= \frac{1}{n-1}\left[n(\mu^2 + \sigma^2) - n(\mu^2 + \frac{1}{n}\sigma^2)\right] =$

$= \frac{1}{n-1}(n\mu^2 + \sigma^2 n - n\mu^2 - \sigma^2) =$

$= \sigma^2$

Szereg rozdzielczy

$\overline{X}$, s, s^2

klasy	liczebność
⟨0, 4)	17
⟨4, 8)	23
⟨8, 12)	15
⟨12, 16)	11
⟨16, 20)	8
⟨20, 24)	6
	80

0-24, 6 przedziałów

$x_1^{10} = 2$, $x_2^{10} = 6$, $x_3^{10} = 10$, $x_4^{10} = 14$

$x_5^{10} = 18$, $x_6^{10} = 22$

Bierzemy uśrednione wartości n razy.

$\overline{x} = \frac{1}{80}(2\cdot 4 + 4\cdot 6 + 4\cdot 10 + 4\cdot 14 \ldots)$

$= 9,4$

$\longrightarrow$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{l} n_i (x_i^\circ - \overline{x})^2 =$$

$$= \frac{1}{n-1} \sum_{i=1}^{l} n_i (x_i^\circ)^2 - \frac{n}{n-1} \overline{x}^2 =$$

$$= \frac{1}{79} \cdot 10\,048 - \frac{80}{79} \cdot 3{,}4^2 = 37{,}71.$$

Przedziały ufności

3.2. $X \sim N(\mu, 4{,}45)$

X ; $\bar{x} = 30{,}82$

a) wyznaczyć 95% przedział ufności

przedział ufności dla wartości przeciętnej cechy X przy założeniu, że a) $X = 36$, b) 64, c) 100

ru

$$\mu\left(\bar{X} - \frac{z\sigma}{\sqrt{n}} ; \bar{x} + \frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right)$$

$1 - \alpha = 0{,}05$, $\frac{\alpha}{2} = 0{,}025$

$1{,}96$

$\phi(z_{0{,}025}) = 1 - 0{,}025 = 0{,}075$

PU dla $n = 36$

$$\mu\left(30,82 - \frac{1,96 \cdot 4,45}{\sqrt{36}};\ 30,82 + \frac{1,96 \cdot 4,45}{\sqrt{36}}\right)$$

$$\mu\,(-\,9,37;\ 32,27)$$

dla $n = 64$

$$(29,73;\ 31,91)$$

Im więcej informacji tym mniejszy przedział nieufności.

Patrzymy jak na prawdopodobieństwa

$P(- —$

3.4. $X \sim N(\mu, \sigma)$

Ponieważ nieznamy σ to PU2:

$$PU2: \left(\bar{x} - \frac{t_\alpha \cdot S}{\sqrt{n}} ; \bar{x} + \frac{t_\alpha \cdot S}{\sqrt{n}}\right)$$

$$\bar{x} = \frac{1}{26}(285 + 339 + \ldots + 249) = 345,5$$

$$S^2 = \frac{1}{26-1}(285^2 + 339^2 + \ldots + 249^2) - \frac{26}{26-1} \cdot (345,5)^2$$

$$S^2 = 7341,06 \qquad S = 85,68$$

a) $\alpha = 0,05$; $t_\alpha = 2,06$

$$PU: \left(345,5 - \frac{2,06 \cdot 85,68}{\sqrt{26}} ; 345,5 + \frac{2,06 \cdot 85,68}{\sqrt{26}}\right)$$

$$PU: (310,89 ; 380,11)$$

$$P(310,89 < \mu < 380,11) = 0,95$$

b) $\alpha = 0{,}01 \qquad t_\alpha = 2{,}76$

$$PU: \left(345{,}5 - \frac{2{,}79 \cdot 85{,}68}{\sqrt{26}};\ 345{,}5 + \frac{2{,}79 \cdot 85{,}68}{\sqrt{26}}\right)$$

$$PU: (298{,}62;\ 392{,}38) \Rightarrow 0{,}99$$

Model PU2 używany zazwyczaj,
t-Studenta nie używamy dla dużych
prób.

NATĘŚCIE

$\phi(z_\alpha) = 1-\alpha$

3.7. $n = 60$

a) $1-\alpha = 0{,}95$ b) $1-\alpha = 0{,}99$

Duża próba bez danych: PU3

$$PU: \left(\overline{x} - \frac{z_{\frac{\alpha}{2}} \cdot S}{\sqrt{n}} ; \overline{x} + \frac{z_{\frac{\alpha}{2}} \cdot S}{\sqrt{n}}\right)$$

$$\overline{x} = \frac{1}{60} \sum_{k=1}^{l} n_k \cdot x_k^0$$

$x_1^0 = 7$

$x_2^0 = 12$

$x_3^0 = 17$

$x_4^0 = 22$

$x_5^0 = 27$

$x_6^0 = 32$

$x_7^0 = 37$

$$\overline{x} = \frac{1}{60}(7 \cdot 2 + \ldots + 37 \cdot 4) = 23{,}33$$

$$S^2 = \frac{1}{n-1} \sum_{k=1}^{l} n_k (x_k^0)^2 - \frac{n}{n-1} \overline{x}^2$$

$$S^2 = \frac{1}{59}(2 \cdot 7^2 + \ldots + 4 \cdot 37^2) - \frac{60}{59} 23{,}33^2 = 49{,}2$$

$$S = 7{,}01$$

a) $\alpha = 0{,}05$ $z_{0,025} = 1{,}96$

$$\phi(z_{0,025}) = 1 - 0{,}025 = 0{,}975$$

$$PU: \left(23{,}33 - \frac{1{,}96 \cdot 7{,}01}{\sqrt{60}} ; 23{,}33 + \frac{1{,}96 \cdot 7{,}01}{\sqrt{60}}\right)$$

PU: — // — ; — // —

b) $\alpha = 0{,}01$, $\frac{\alpha}{2} = 0{,}005$ $z_{0,005} = 2{,}58$

$$PU: \left(23{,}33 - \frac{2{,}58 - 7{,}01}{\sqrt{60}} ; 23{,}33 + \ldots\right)$$

— (| — ; — // —

$\sigma = \sigma$

3.10. $n = 20 \quad s^2 = 25$

PU4 dla σ^2: $\left(\frac{(n-1) \cdot s^2}{\chi^2_{\frac{\alpha}{2}}} ; \frac{n-1 \cdot s^2}{\chi^2_{1-\frac{\alpha}{2}}} \right)$

a) $1-\alpha = 0{,}9 \rightarrow \alpha = 0{,}1$

$\frac{\alpha}{2} = 0{,}05 \quad 1-\frac{\alpha}{2} = 0{,}95$

$\chi^2_{\frac{\alpha}{2}} = 30{,}14 \quad \chi^2_{1-\frac{\alpha}{2}} = 10{,}12$

PU dla $\sigma = \left(\frac{19 \cdot 25}{30{,}14} ; \frac{19 \cdot 25}{10{,}12} \right) =$

$= (15{,}76 ; 46{,}94)$

$P(15{,}76 < \sigma < 46{,}94) = 0{,}9$

PU dla σ: $P(3{,}97 < \sigma < 6{,}85) = 0{,}9$

b) $1 - 0{,}95 = 0{,}05 = \alpha \quad \frac{\alpha}{2} = 0{,}025 \quad 1-\frac{\alpha}{2} = 0{,}975$

$\chi^2_{\frac{\alpha}{2}} = 32{,}85 \quad \chi^2_{1-\frac{\alpha}{2}} = 8{,}91$

PU dla $\sigma^2 = \left(\frac{19 \cdot 25}{32{,}85} ; \frac{19 \cdot 25}{8{,}91} \right) =$

$= (14{,}46 ; 53{,}31)$

3.13. $n=50 \quad 1-\alpha=0{,}98$

PU$^{(5)}$ dla $\sigma^2 = \left(\left(\frac{s}{1+z_{\frac{\alpha}{2}}/\sqrt{2n}}\right)^2 \text{ i } \left(\frac{s}{1-\ldots}\right)^2 \right)$

$\overline{x} = \frac{1}{50}(0\cdot 6 + 1\cdot 10 + \ldots 7\cdot 1) = 2{,}56$

$s^2 = \frac{1}{49}(6\cdot 0^2 + \ldots) - \frac{50}{49}\cdot 2{,}59^2$

$s^2 = 3{,}07$

$\alpha = 0{,}02 \quad \frac{\alpha}{2} = 0{,}01$

$z_{\frac{\alpha}{2}} = 2{,}33$

PU: $\left(\frac{3{,}07}{\left(1+\frac{2{,}33}{\sqrt{2\cdot 50}}\right)^2} \text{ i } \frac{3{,}07}{(1-\ldots)^2} \right)$

4.2. $X \sim N(\mu, 10)$, $H_0 : \mu = 95$, $H_1 : \mu > 95$

$\alpha = 0,05$, $\overline{X} = 100$; (a) $n = 9$, (b) $n = 16$

5 KROKÓW

I: $H_0 : \mu = 95$, $H_1 : \mu > 95$

Udowadniamy zawsze H_1, H_0 pomocnicze

II $\alpha = 0,05$

III TH1: St. testowa $U = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} \sim N(0,1)$

Grupy hipotez ABC do H_1

z tabeli 4.3. Obszar kryt: $K = \langle k; \infty)$, $k = z_{\alpha}$

TEST 1° Jeżeli $u \in K$, to H_0 należy odrzucić i przyjąć H_1.

2° Jeżeli $u \notin K$, to nie ma podstaw do H_0.

IV $u = \frac{100 - 95}{10} \cdot \sqrt{9} = 0,5 \cdot 3 = 1,5$; $k = z_{0,005} = 1,64$

$K = (1,64; +\infty)$

V $u \notin K$, nie ma podstaw do odrzucenia H_0.

b)

I, II, III tak samo

IV $u = \frac{100-95}{10}\sqrt{16} = 0{,}5 \cdot 4 = 2$

$K = \langle 1{,}64; +\infty)$

V $u \in K \Rightarrow$, odrzucamy H_0, przyjmujemy H_1

Niezależność zmiennych losowych

(X,Y) - zmienna dwuwymiarowa
skokowa
lub przeliczalna liczba wartości

$\Omega_X = \{x_1, x_2, \ldots\}$, $\Omega_Y = \{y_1, y_2, y_3 \ldots\}$

Rozkład prawa? $P(X=x_i, Y=y_j)$ $i=1,2,3,\ldots$ $j=1,2,3,\ldots$

$$P(X=x_i, Y=y_i) \geq 0, \quad \sum_j \sum_i P(X=x_i, Y=y_i) = 1$$

Funkcja łącznego rozkładu prawdopodobieństwa

$$P_{X,Y}(x,y) = P(X=x, Y=y)$$

$$P_{X,Y}(x,y) = \begin{cases} p_{ij}, & X = x_i, Y = y_j \\ 0 & \text{dla poz } x, y \end{cases}$$

Prawdopodobieństwo brzegowe

$\Omega_X = \{x_1, x_2, x_3 \ldots x_r\}$, $\Omega_Y = \{y_1, y_2, \ldots, y_s\}$

$$\begin{cases} P_X(x_i) = P_{X,Y}(x_i, y_1) + \ldots + P_{X,Y}(x_i, y_s) = \\ \qquad p_{i1} + p_{i2} + \ldots p_{is} = p_{i\bullet} \\ \text{Dla } y \text{ jest } p_{\bullet i} \end{cases}$$

X \ Y	y_1	…	y_s	P_X
x_1	p_{11}	…	p_{1s}	
⋮	⋮		⋮	
x_r	p_{r1}	…	p_{rs}	
P_Y				1

Niezależność zmiennych losowych

X i Y nazywamy niezależnymi, jeśli

$$P(X=A, Y=B) = P(X=A) \cdot P(X=B)$$

dla podzbiorów $A, B \in \mathbb{R}$

Kryterium niezależności:

zmienne losowe X, Y są niezależne iff

$$P_{X,Y}(X,Y) = P_X(X) \cdot P_Y(Y)$$

Przykład

(X, Y)

$$P_{X,Y}(x,y) = \begin{cases} \frac{1}{18} xy & \text{dla } x=1,2,3 \quad y=1,2 \\ 0 & \text{dla innych} \end{cases}$$

$X \backslash Y$	1	2	P_X
1	$1/18$	$1/9$	$1/6$
2	$1/9$	$2/9$	$1/3$
3	$1/6$	$1/3$	$1/2$
P_Y	$1/3$	$2/3$	1

$p_{11} = 1/18$, $p_{1\cdot} \cdot p_{\cdot 1} = 1/18$

$p_{12} = 1/9$, $p_{1\cdot} \, p_{\cdot 2} = 1/9$

....

X, Y są niezależne

Próba testu niezależności

(X_i, Y_j), $i = 1,2,3 \dots r$ $j = 1,2,3,\dots s$

m_{ij} – liczebności rzeczywiste, x_i, y_i występuje w próbie m_{ij} razy.

$m_{ij} \geq 8$

$X \backslash Y$	y_1	$\dots$	y_s	$m_{i\cdot}$
x_1	m_{11}	$\dots$	m_{1s}	$m_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_r	m_{r1}	$\dots$	m_{rs}	$m_{r\cdot}$
$m_{\cdot j}$	$m_{\cdot 1}$	$\dots$	$m_{\cdot s}$	m_{rs}

Hipotezy

H_0 - zmienne są niezależne

H_1 - zmienne są zależne

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{ij} - n p_{ij})^2}{n p_{ij}}$$

wartości n_{ij} są znane

wartość brzegowa - $\frac{n_{\cdot 1}}{n_{rs}}$ dla $X=1$

dla 2 wartościowej mnożymy

χ^2 ma rozkład $\chi^2 = (r-1)(s-1)$, s^2

$K = \langle k, \infty)$, $k = \chi^2_\alpha$

7.6.

x \ y	y_1	y_2	y_3
x_1	20	40	60
x_2	45	20	15

$\alpha = 0{,}05$

I. H_0 - x, y są niezależne

H_1 - x, y istnieje zależność

II. $\alpha = 0{,}05$

III. St. test. $\chi^2 = \sum_i^2 \sum_j^3 \left(\frac{N_{ij} - np_{ij}}{np_{ij}} \right)^2$

$\chi^2 = (2-1)(3-1) \Rightarrow 2$ stopnie swobody

$K = \langle K, \infty)$, $\alpha = \chi^2_{0{,}05}$

1° $\chi^2 \in K$ istnieje zależność

2° $\chi^2 \notin K$ nie istnieje zależność

IV

X \ Y	y_1	y_2	y_3	$n_{j\cdot}$
x_1	20	40	60	120
x_2	45	20	15	80
$n_{\cdot j}$	65	60	75	200

X \ Y	y_1	y_2	y_3	y_i
x_1	0,195	0,18	0,225	$\frac{120}{200}=0,6$
x_2	0,13	0,12	0,15	$\frac{80}{200}=0,4$
$p_{\cdot j}$	0,325	0,3	0,375	1

$$\chi^2 = \left(\frac{20-200\cdot 0,195}{200\cdot 0,195}\right)^2 + \left(\frac{40-200\cdot 0,18}{200\cdot 0,18}\right)^2 +$$

$$+ \ldots = 36.752$$

$$\chi^2_{0,05} = 5,99 \quad , \quad K = \langle 5,99, \infty)$$

V χ^2 należy do K, odrzucamy H_0.

Istnieje zależność stanu zdrowia od metody leczenia.

Regresja

Na zaliczeniu!

5.2 $(3,2),(5,3),(6,4),(8,6),(9,5),(11,8)$

$$y = b_0 + b_1 x \qquad b_1 = \frac{S_{xy}}{S_{xx}}, \quad b_1 = \bar{y} - b_1 \bar{x}$$

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum_i x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2$$

$$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum y_i\right)$$

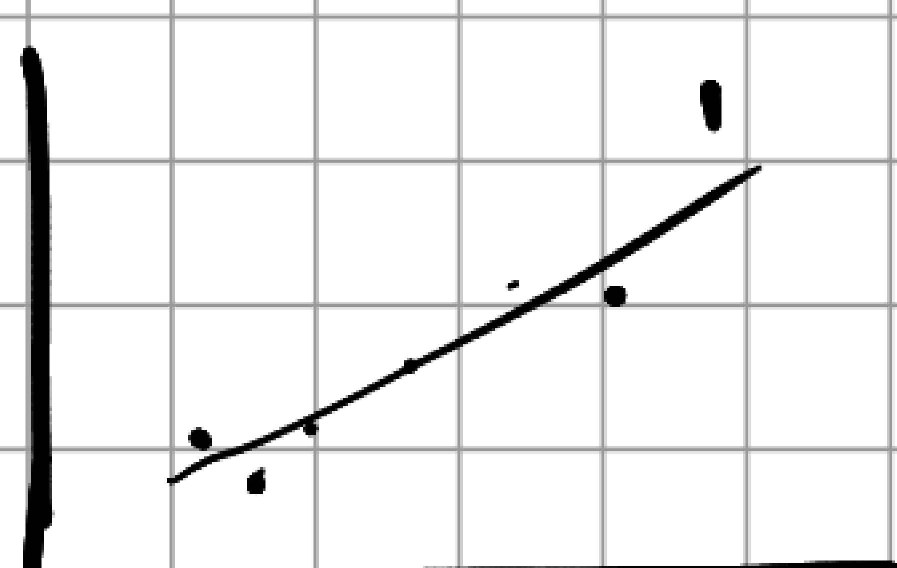
$$\sum_{i=1}^{6} x_i = 42 \quad \sum_{i=1}^{6} y_i = 28, \quad \sum_{i=1}^{6} x_i^2 = 336, \quad \sum_{i=1}^{6} y_i x_i = 226$$

$$\bar{x} = \frac{1}{6}\cdot 42 = 7 \quad \bar{y} = \frac{1}{6}\cdot 28 = \frac{14}{3}, \quad S_{xx} = 336 - \frac{1}{6}42^2 = 42$$

$$S_{xy} = 226 - \frac{1}{6}\cdot 42 \cdot 28 = 30$$

$$b_1 = \frac{30}{42} = 0{,}71 \qquad b_0 = \frac{14}{3} - \frac{30}{42}\cdot 7 = -0{,}33$$

$$y = -0{,}33 + 0{,}71x$$



← graf konieczny

→ zakładając, że dane punkty są

c) obserwacjami zmiennej losowej (X, Y)

wyznaczyć prostą regresji zmiennych Y wzg.? X

z definicji MNK jest $y = -0{,}33 + 0{,}71x$

d) $r^2 = \frac{S^2_{xy}}{S_{xx} \cdot S_{yy}}$ $\qquad S_{yy} = \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2 = 23{,}33$

$r^2 = \frac{30^2}{42 \cdot 23{,}33} = 0{,}92$

92% rozproszenia Y da się wyjaśnić regresją liniową z X.

e) wyjaśnić czy regresja liniowa jest właściwym modelem.

$\alpha = 0{,}05$

$y = \beta_0 + \beta_1 x \quad , \quad \beta_1 \neq 0$

I $H_0 - \beta_1 = 0$

$H_1 - \beta_1 \neq 0$

II $\alpha = 0{,}05$

III $T = \frac{B_1}{S_e / \sqrt{S_{xx}}}$ $\quad S_e = \frac{S_{yy} - b_1 S_{xy}}{n-2}$

T_{n-2}

$K = (-\infty, k), (k, \infty)$

test 1° $t \in K$ jest

2° $t \notin K$ nie jest

IV $S_e \approx 0{,}71$, $t = 6{,}48$, $K = (-\infty, -2{,}78) \ldots$

V $t \in K$, Reg. lin. jest.

5.7

x	1	4	4	7	9	10	12	14	17	22
y	21	31	37	44	58	65	60	73	84	91

$\sum x_i = 100$ $\qquad \sum x_i y_i = 6945$

$\sum x_i^2 = 1376$ $\qquad \bar{x} = 10$

$\sum y_i = 564$ $\qquad \bar{y} = 56{,}4$

$\sum y_i^2 = 36562$

$$S_{xx} = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 376$$

$$S_{xy} = \sum x_i y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum y_i\right) = 1305$$

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{1305}{376} = 3{,}47$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x} = 56{,}4 - 3{,}47 \cdot 10 = 21{,}7$$

$$y = 21{,}7 + 3{,}47x$$

verte $\Longrightarrow$

1) Współczynnik złożoności

$$r^2 = \frac{S_{xy}^2}{S_{xx} \cdot S_{yy}} = \frac{13{,}05^2}{376 \cdot 4752{,}4} \approx 0{,}95$$

2) 95% rozproszenia pochodzi od liniowej regresji z X.

3) istotność regresji liniowej

$$\mu_{y/x} = \beta_0 + \beta_1 x, \quad \beta_1 \neq 0$$

I. $H_1 . \ \beta_1 = 0$

$H_2 . \ \beta_2 \neq 0$

II $\alpha = 0{,}05$

III St. test. $\frac{\beta_1}{S_2 / \sqrt{S_{xx}}}$, $K = \langle -\infty, k), (k, \infty\rangle$

$k = t_\alpha$

Jeśli u jest zawarte w K, to regresja liniowa jest istotna.

Jeśli u nie jest... verte →

IV $\quad S_e^1 = \dfrac{S_{yy} - b_1 S_{xy}}{n-2\ (?)} =$

$= \dfrac{4752{,}4 - 3{,}47 \cdot 1305}{8} \approx 28, \ S_e = 5{,}29$

$t = \dfrac{3{,}47}{5{,}29}\sqrt{376} \approx 12{,}72$

$t_{0{,}05} = 2{,}31$

V regresja liniowa jest istotna.

Prognoza

$\hat{y}_{x=4} = 21{,}7 + 3{,}47 \cdot 14 = 70{,}28$

$\hat{y} = 70$

verte →

przedział ufności prognozowania

PU dla $\hat{q}$: $\left(B_0 + B_1 x_0 - t_\alpha \cdot S_e \sqrt{1 - \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}\right)$

...

6.2

x	79	95	81	86	87	94	59
y	85	97	78	76	84	84	67

$R = \frac{S_{xy}}{\sqrt{S_{xx}} \cdot \sqrt{S_{yy}}}$, obliczamy to samo co ostatnio

$r = \frac{728}{\sqrt{1108{,}86}\sqrt{652}} = 0{,}86$

$0{,}5 < r < 0{,}9$

Istotność statystyczna korelacji ! TEST

I. $H_0: \rho = 0$, $H_1: \rho \neq 0$

II $\alpha = 0.05$

III $T = R \dfrac{\sqrt{n-1}}{\sqrt{1-R^2}}$

$K = (-\infty, -t_\alpha) \cup (t_\alpha, \infty)$

TEST 1° Jeżeli $t \in K$, to odrzucamy H_0, skorelowane liniowo.

IV $t = 0{,}86 \cdot \dfrac{\sqrt{5}}{\sqrt{1-0.86}} = 3{,}77$

$t_\alpha = 2{,}57$

$K = (-\infty, -2{,}57) \cup \langle 2{,}57, \infty)$

V $t \in K$, co oznacza, że są skorelowane liniowo.