1 Aprašomoji statistika

1.1 Variacinė eilutė

Statistinė eilutė, išdėstyta nuo mažiausos reikšmės iki didžiausios:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}.$$

1.2 Imties vidurkis

Populiacijos vidurkis (populiacijos didumas N):

$$\mu := \frac{x_1 + x_2 + \cdots + x_N}{N}.$$

Imties vidurkis (imties didumas n):

$$\bar{x} := \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Grupuotų duomenų vidurkiui skaičiuoti pasirenkami intervalų viduriniai taškai.

Imties vidurkio savybės:

- Jei iš kiekvieno duomes atimsime imties vidurkį, tai naujosios imties vidurkis bus 0.
- Kiekvieną duomenį padauginus iš C, naujasis vidurkis bus lygus senajam, padaugintam iš C.
- Prie kiekvieno duomens pridėjus tą pačią konstantą, vidurkis padidėja per tą konstantą.

1.3 Moda

Moda - dažniausiai duomenų aibėje pasikartojusi reikšmė. Jeigu tokių reikšmių yra kelios ir jos variacinėje eilutėje yra greta, tai moda yra lygi tų reikšmių vidurkiui. Duomenų aibė gali neturėti modos arba turėti daugiau, kaip vieną modą.

1.4 Mediana

Mediana yra skaičius, už kurį 50% variacinės eilutės reikšmių yra nedidesnės ir 50% yra nemažesnės. (Mediana dalija variacinę eilutę pusiau).

- Jeigu n nelyginis, tai mediana yra variacinės eilutės **reikšmė**, atitinkanti $(n+1)/2$ poziciją.
- Jeigu stebėjimų skaičius n lyginis, tai mediana yra variacinės eilutės **reikšmių**, atitinkančių pozicijas $(n/2)$ ir $(n/2)+1$, aritmetinis vidurkis.

1.5 Kvantiliai

Reikšmė, dalijanti variacinę eilutę į $100q$ ir $100(1-q)$ procentinių dalių, vadinama q-osios $(0 < q < 1)$ eilės kvantiliu.

Kvartiliai Q_1, Q_2, Q_3 dalija variacinę eilutę į keturias "maždaug" lygias dalis. Vienas iš kvartilių radimo būdų Q_1:

- Randamas indeksas $i = n/4$.

- Jeigu i nėra sveikasis skaičius, tai imama sveikoji jo dalis $[i]$. Ieškomasis kvartilis yra $[i]+1$ variacinės eilutės narys, t.y. $Q_1 = x_{([i]+1)}$.
- Jeigu i yra sveikasis skaičius, tai
$$Q_1 = \frac{x_{(i)} + x_{(i+1)}}{2}.$$

Vienas iš kvartilių radimo būdų Q_3:

- Randamas indeksas $i = 3n/4$.
- Jeigu i nėra sveikasis skaičius, tai imama sveikoji jo dalis $[i]$. Ieškomasis kvartilis yra $[i]+1$ variacinės eilutės narys, t.y. $Q_3 = x_{([i]+1)}$.
- Jeigu i yra sveikasis skaičius, tai
$$Q_3 = \frac{x_{(i)} + x_{(i+1)}}{2}.$$

1.6 Imties dispersija

Populiacijos dispersija:

$$\sigma^2 := \frac{(x_1-\mu)^2 + (x_2-\mu)^2 + \cdots + (x_N-\mu)^2}{N} = \frac{x_1^2 + \cdots + x_N^2}{N} - \mu^2.$$

Imties dispersija:

$$s^2 := \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n-1} = \frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n-1} - \frac{n}{n-1}(\bar{x})^2.$$

Imties dispersijos savybės:

- Kiekvieną duomenį padauginus iš C, naujoji dispersija bus lygi senajai, padaugintai iš C^2.
- Prie kiekvieno duomens pridėjus tą pačią konstantą, dispersija nepasikeičia.

1.7 Standartinis nuokrypis

Standartinis nuokrypis – tai šaknis iš dispersijos. Populiacijos: $\sigma := \sqrt{\sigma^2}$. Imties $s := \sqrt{s^2}$.

1.8 Kvartilių skirtumas

$IQR := Q_3 - Q_1$.

1.9 Kokybinės įvairovės indeksas

$$IQV := \frac{k[n^2 - (f_1^2 + f_2^2 + \cdots + f_k^2)]}{n^2(k-1)}.$$

Čia k – kategorijų skaičius, n – stebėjimų skaičius, f_j – $j$osios kategorijos stebėjimų skaičius ($j$osios kategorijos dažnis).

Duomenys – kategoriniai. Didesnis indeksas – didesnė įvairovė. Didžiausia reikšmė 1 (visų kategorijų atstovų imtyje po lygiai). Mažiausia reikšmė 0 (imtį sudaro vienos kategorijos atstovai).

1.10 Kitimo koeficientas

Populiacijos (σ – populiacijos standartinis nuokrypis, μ – populiacijos vidurkis):

$$CV := \frac{\sigma}{\mu}.$$

Imties (s – imties standartinis nuokrypis, $\bar{x}$ – imties vidurkis):

$$CV := \frac{s}{\bar{x}}.$$

1.11 Asimetrijos koeficientas

$$g_3 := \frac{m_3}{s^3}, \quad m_3 := \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^3,$$

s – standartinis nuokrypis. Histograma simetriška, kai $g_3 = 0$.

1.12 Eksceso koeficientas

$$g_4 := \frac{m_4}{s^4}, \quad m_4 := \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^4,$$

s – standartinis nuokrypis. Histograma *normaliai* lėkšta, kai $g_4 = 0$.

1.13 Išskirtys

Sąlygine išskirtimi vadinamas duomuo, priklausantis intervalui $[Q_1 - 3IQR; Q_1 - 1,5IQR)$ arba intervalui $(Q_3 + 1,5IQR; Q_3 + 3IQR]$. Išskirtimi vadinamas duomuo, mažesnis už $Q_1 - 3IQR$ arba didesnis už $Q_3 + 3IQR$. Čia IQR – kvartilių skirtumas.

1.14 Empirinė taisyklė

Jei duomenys normalūs (gausiški), tai

- apytiksliai 68% visų duomenų patenka į intervalą $(\bar{x} - s, \bar{x} + s)$
- apytiksliai 95% visų duomenų patenka į intervalą $(\bar{x} - 2s, \bar{x} + 2s)$
- beveik visi duomenys patenka į intervalą $(\bar{x} - 3s, \bar{x} + 3s)$.

1.15 Standartizuotosios reikšmės

$$z_k := \frac{x_k - \bar{x}}{s}.$$

Čia s – standartinis nuokrypis, $\bar{x}$ – imties vidurkis.

2 Įverčiai ir hipotezių tikrinimas

2.1 Pasikliautinieji intervalai

- Normalaus kintamojo $X \sim \mathcal{N}(\mu, \sigma^2)$ 95% pasikliautinis intervalas vidurkiui:

$$\bar{x} \pm t_{0.025}(n-1)\frac{s}{\sqrt{n}}.$$

  Čia s – standartinis nuokrypis, o $t_{0.025}(n-1)$ – Stjudento kriterijaus su $n-1$ laisvės laipsniu 0.025 lygmens kritinė reikšmė (imama iš lentelių).

- Normalaus kintamojo $X \sim \mathcal{N}(\mu, \sigma^2)$. 95% pasikliautinis intervalas dispersijai:

$$\hat{\sigma}_1^2 := \frac{s^2 \cdot (n-1)}{\chi^2_{0.025}(n-1)}, \qquad \hat{\sigma}_2^2 := \frac{s^2 \cdot (n-1)}{\chi^2_{0.975}(n-1)}.$$

  Čia s^2 – imties dispersija, o $\chi^2_{0.025}(n-1)$ – chi kvadrato kriterijaus su $n-1$ laisvės laipsniu 0.025 lygmens kritinė reikšmė (imama iš lentelių).

- Puasono kintamojo $X \sim \mathcal{P}(\lambda)$. 95% pasikliautinis intervalas parametrui λ:

$$\hat{\lambda}_1 := \frac{\chi^2_{0.975}(2S_n)}{2n}, \qquad \hat{\lambda}_2 := \frac{\chi^2_{0.025}(2S_n+2)}{2n}.$$

  Čia $S_n = x_1 + x_2 + \cdots x_n$, o $\chi^2_{0.975}(2S_n)$ – chi kvadrato kriterijaus su $2S_n$ laisvės laipsniu 0.975 lygmens kritinė reikšmė (imama iš lentelių).

- Proporcija p. 95% pasikliautinis intervalas proporcijai p:

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \qquad \hat{p} := \frac{\text{vienetų skaičius imtyje}}{n},$$

  t.y. $\hat{p}$ yra imties proporcija (respondentų, tenkinančių požymį, dalis imtyje. Pvz., kairiarankių dalis imtyje).

2.2 Stjudento kriterijus vienai imčiai

$X \sim \mathcal{N}(\mu, \sigma^2)$, μ ir σ^2 nežinomi. Statistinė hipotezė:

$$\begin{cases} H_0 : \mu = a, \\ H_1 : \mu \neq a \end{cases} \quad arba \quad \begin{cases} H_0 : \mu = a, \\ H_1 : \mu > a. \end{cases} \quad arba \quad \begin{cases} H_0 : \mu = a, \\ H_1 : \mu < a. \end{cases}$$

Statistika:

$$t = \frac{\bar{x} - a}{\sqrt{s^2/n}}.$$

Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$:

H_1	$\mu \neq a$	$\mu > a$	$\mu < a$
H_1 priimame, jei	$\lvert t\rvert > t_{0.025}(n-1)$	$t > t_{0.05}(n-1)$	$t < -t_{0.05}(n-1)$

Čia $t_{0.025}(n-1)$ – Stjudento kriterijaus su $n-1$ laisvės laipsniu 0.025 lygmens kritinė reikšmė (imama iš 3 lentelės).

2.3 Stjudento kriterijus dviems nepriklausomoms imtims

$X \sim \mathcal{N}(\mu_x, \sigma^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma^2)$, μ_x, μ_y ir σ^2 nežinomi. Imčių didumai n ir m. Statistinė hipotezė:

$$\begin{cases} H_0 : \mu_x = \mu_y, \\ H_1 : \mu_x \neq \mu_y, \end{cases} \quad arba \quad \begin{cases} H_0 : \mu_x = \mu_y, \\ H_1 : \mu_x > \mu_y, \end{cases} \quad arba \quad \begin{cases} H_0 : \mu_x = \mu_y, \\ H_1 : \mu_x < \mu_y. \end{cases}$$

Statistika:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_x^2 + (m-1)s_y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Čia s_x^2, s_y^2 yra imčių dispersijos. Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$:

H_1	$\mu_x \neq \mu_y$	$\mu_x > \mu_y$	$\mu_x < \mu_y$
H_1 priimame, jei	$\lvert t\rvert > t_{0.025}(n+m-2)$	$t > t_{0.05}(n+m-2)$	$t < -t_{0.05}(n+m-2)$

Čia $t_{0.025}(n+m-2)$ – Stjudento kriterijaus su $n+m-2$ laisvės laipsniu 0.025 lygmens kritinė reikšmė (imama iš 3 lentelės).

2.4 Porinis Stjudento kriterijus

$X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_x, \sigma_y^2)$, μ_x, μ_y ir σ_x^2, σ_y^2 nežinomi. Imčių didumas n. Statistinė hipotezė:

$$\begin{cases} H_0 : \mu_x = \mu_y, \\ H_1 : \mu_x \neq \mu_y, \end{cases} \quad arba \quad \begin{cases} H_0 : \mu_x = \mu_y, \\ H_1 : \mu_x > \mu_y, \end{cases} \quad arba \quad \begin{cases} H_0 : \mu_x = \mu_y, \\ H_1 : \mu_x < \mu_y. \end{cases}$$

Statistika. Randame porų skirtumus: $d_1 = x_1 - y_1, d_2 = x_2 - y_2, \ldots, d_n = x_n - y_n$. Suskaičiuojame vidurkį $\bar{d}$ ir dispersiją s_d^2. Suskaičiuojame

$$t = \frac{\bar{d}}{\sqrt{s_d^2/n}}.$$

Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$:

H_1	$\mu_x \neq \mu_y$	$\mu_x > \mu_y$	$\mu_x < \mu_y$
H_1 priimame, jei	$\lvert t\rvert > t_{0.025}(n-1)$	$t > t_{0.05}(n-1)$	$t < -t_{0.05}(n-1)$

Čia $t_{0.025}(n-1)$ – Stjudento kriterijaus su $n-1$ laisvės laipsniu 0.025 lygmens kritinė reikšmė (imama iš 3 lentelės).

2.5 Kriterijus dispersijos lygybei skaičiui tikrinti

$X \sim \mathcal{N}(\mu, \sigma^2)$, μ ir σ^2 nežinomi. Statistinė hipotezė:

$$\begin{cases} H_0 : \sigma^2 = a, \\ H_1 : \sigma^2 \neq a, \end{cases} \quad arba \quad \begin{cases} H_0 : \sigma^2 = a, \\ H_1 : \sigma^2 > a, \end{cases} \quad arba \quad \begin{cases} H_0 : \sigma^2 = a, \\ H_1 : \sigma^2 < a. \end{cases}$$

Statistika:

$$t = \frac{(n-1)s^2}{a}.$$

Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$:

H_1	$\sigma^2 \neq a$	$\sigma^2 > a$	$\sigma^2 < a$
H_1 priimame, jei	$t < \chi^2_{0.975}(n-1)$ arba $t > \chi^2_{0.025}(n-1)$	$t > \chi^2_{0.05}(n-1)$	$t < \chi^2_{0.95}(n-1)$

Čia $\chi^2_{0.975}(n-1)$ – chi kvadrato kriterijaus su $n-1$ laisvės laipsniu 0.975 lygmens kritinė reikšmė (imama iš 4 lentelės).

2.6 Kriterijus dviejų dispersijų lygybei tikrinti

$X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_x, \sigma_y^2)$, μ_x, μ_y ir σ_x^2, σ_y^2 nežinomi. Imčių didumai n ir m. Statistinė hipotezė (K kažkokia konstanta. Dispersijų lygybei tikrinti, imame $K = 1$):

$$\begin{cases} H_0 : \sigma_x^2 = K\sigma_y^2, \\ H_1 : \sigma_x^2 \neq K\sigma_y^2, \end{cases} \quad arba \quad \begin{cases} H_0 : \sigma_x^2 = K\sigma_y^2, \\ H_1 : \sigma_x^2 > K\sigma_y^2, \end{cases} \quad arba \quad \begin{cases} H_0 : \sigma_x^2 = K\sigma_y^2, \\ H_1 : \sigma_x^2 < K\sigma_y^2. \end{cases}$$

Statistika:

$$F = \frac{s_x^2}{Ks_y^2}.$$

Čia s_x^2, s_y^2 yra imčių dispersijos. Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$:

H_1	H_1 priimame, jei
$\sigma_x^2 \neq K\sigma_y^2$	$F < F_{0.975}(n-1, m-1)$ arba $F > F_{0.025}(n-1, m-1)$
$\sigma_x^2 > K\sigma_y^2$	$F > F_{0.05}(n-1, m-1)$
$\sigma_x^2 < K\sigma_y^2$	$F < F_{0.95}(n-1, m-1)$

Čia $F_{0.975}(n-1, m-1)$ Fišerio skirstinio su $n-1$ ir $m-1$ laisvės laipsniais 0.975 lygmens kritinė reikšmė (imama iš 5 lentelės) arba naudojamas ryšys

$$F_{0.975}(n, m) = \frac{1}{F_{0.025}(m, n)}.$$

2.7 Kriterijus hipotezei apie proporciją tikrinti

Dvireikšmių duomenų aibę sudaro nuliai ("nesėkmė") ir vienetai ("sėkmė"). Parametras p – vienetų dalis populiacijoje.

Statistinė hipotezė:

$$\begin{cases} H_0 : p = a, \\ H_1 : p \neq a, \end{cases} \quad arba \quad \begin{cases} H_0 : p = a, \\ H_1 : p > a, \end{cases} \quad arba \quad \begin{cases} H_0 : p = a, \\ H_1 : p < a. \end{cases}$$

Statistika:

$$z = \frac{\hat{p} - a}{\sqrt{a(1-a)/n}}.$$

Čia $\hat{p}$ – vienetų dalis imtyje, n – imties didumas. Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$:

H_1	$p \neq a$	$p > a$	$p < a$
H_1 priimame, jei	$\lvert z\rvert > 1.96$	$z > 1.64$	$z < -1.64$

2.8 Kriterijus dviejų proporcijų lygybei tikrinti

Dvi dvireikšmių duomenų aibes sudaro nuliai ("nesėkmė") ir vienetai ("sėkmė"). Parametras p_1 – vienetų dalis pirmojoje populiacijoje. Parametras p_2 – vienetų dalis antrojoje populiacijoje. Skaičius a rodo, kiek skiriasi p_1 nuo p_2. Lygybei tikrinti imame $a = 0$.

Statistinė hipotezė:

$$\begin{cases} H_0 : p_1 = p_2 + a, \\ H_1 : p_1 \neq p_2 + a, \end{cases} \quad arba \quad \begin{cases} H_0 : p_1 = p_2 + a, \\ H_1 : p_1 > p_2 + a, \end{cases} \quad arba \quad \begin{cases} H_0 : p_1 = p_2 + a, \\ H_1 : p_1 < p_2 + a. \end{cases}$$

Statistika:

$$z = \frac{\hat{p}_1 - \hat{p}_2 - a}{\sqrt{\hat{p}_1(1-\hat{p}_1)/n + \hat{p}_2(1-\hat{p}_2)/m}}\,.$$

Čia $\hat{p}_1$ – vienetų dalis pirmojoje imtyje, $\hat{p}_2$ – vienetų dalis antrojoje imtyje, n – pirmosios imties didumas, m – antrosios imties didumas. Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$:

H_1	$p_1 \neq p_2 + a$	$p_1 > p_2 + a$	$p_1 < p_2 + a$
H_1 priimame, jei	$\lvert z\rvert > 1.96$	$z > 1.64$	$z < -1.64$

2.9 Kriterijus koreliacijos lygybei nuliui tikrinti

$X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_x, \sigma_y^2)$, μ_x, μ_y ir σ_x^2, σ_y^2 nežinomi. Imčių didumas n (turime n porų (x_i, y_i)).

Statistinė hipotezė:

$$\begin{cases} H_0 : \varrho = 0, \\ H_1 : \varrho \neq 0, \end{cases} \quad arba \quad \begin{cases} H_0 : \varrho = 0, \\ H_1 : \varrho > 0, \end{cases} \quad arba \quad \begin{cases} H_0 : \varrho = 0, \\ H_1 : \varrho < 0. \end{cases}$$

Statistika:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}, \qquad t = r\sqrt{\frac{n-2}{1-r^2}}\,.$$

Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$:

H_1	$\varrho \neq 0$	$\varrho > 0$	$\varrho < 0$
H_1 priimame, jei	$\lvert t\rvert > t_{0.025}(n-2)$	$t > t_{0.05}(n-2)$	$t < -t_{0.05}(n-2)$

Čia $t_{0.025}(n-2)$ – Stjudento kriterijaus su $n-2$ laisvės laipsniu 0.025 lygmens kritinė reikšmė (imama iš 3 lentelės).

2.10 Dviejų koreliacijos koeficientų lygybė (nepriklausomos imtys)

Intervalinių normaliųjų duomenų poros $(x_{1a}, y_{1a}), (x_{2a}, y_{2a}), , (x_{na}, y_{na})$ (čia pirma grupė) ir (x_{1b}, y_{1b}), (x_{2b}, y_{2b}),Ě, (x_{mb}, y_{mb}) (čia antra grupė). n porų pirmoje ir m porų antrojoje grupėje.

Statistinė hipotezė:

$$\begin{cases} H_0 : \varrho_a = \varrho_b, \\ H_1 : \varrho_a \neq \varrho_b, \end{cases} \quad arba \quad \begin{cases} H_0 : \varrho_a = \varrho_b, \\ H_1 : \varrho_a > \varrho_b, \end{cases} \quad arba \quad \begin{cases} H_0 : \varrho_a = \varrho_b, \\ H_1 : \varrho_a < \varrho_b. \end{cases}$$

Statistika: Taikydami formulę

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}},$$

surandame r_a ir r_b. Tada surandame Fišerio transformacijas:

$$z_a = \frac{1}{2} \ln \frac{1 + r_a}{1 - r_a}, \qquad z_b = \frac{1}{2} \ln \frac{1 + r_b}{1 - r_b}.$$

Po to suskaičiuojame

$$z = \frac{z_a - z_b}{\sqrt{\frac{1}{n-3} + \frac{1}{m-3}}}.$$

Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$

H_1	$\varrho_a \neq \varrho_b$	$\varrho_a > \varrho_b$	$\varrho_a < \varrho_b$
H_1 priimame, jei	$\lvert z \rvert > 1.96$	$z > 1.64$	$z < -1.64$

2.11 Dviejų koreliacijos koeficientų lygybė (priklausomos imtys)

Intervalinių normaliųjų duomenų trejetai $(x_1, y_1, z_1), (x_2, y_2, z_2), , (x_n, y_n, z_n)$.

Statistinė hipotezė:

$$\begin{cases} H_0 : \varrho_{xy} = \varrho_{xz}, \\ H_1 : \varrho_{xy} \neq \varrho_{xz}, \end{cases} \quad arba \quad \begin{cases} H_0 : \varrho_{xy} = \varrho_{xz}, \\ H_1 : \varrho_{xy} > \varrho_{xz}, \end{cases} \quad arba \quad \begin{cases} H_0 : \varrho_{xy} = \varrho_{xz}, \\ H_1 : \varrho_{xy} < \varrho_{xz}. \end{cases}$$

Statistika: Taikydami formulę

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}},$$

surandame r_{xy}, r_{xz} ir r_{yz}. Tada

$$t = \frac{(r_{xy} - r_{xz})\sqrt{(n-3)(1 + r_{yz})}}{\sqrt{2(1 - r_{xy}^2 - r_{xz}^2 - r_{yz}^2 + 2r_{xy}r_{xz}r_{yz})}}.$$

Sprendimas, kai reikšmingumo lygmuo $\alpha = 0.05$

H_1	$\varrho_{xy} \neq \varrho_{xz}$	$\varrho_{xy} > \varrho_{xz}$	$\varrho_{xy} < \varrho_{xz}$
H_1 priimame, jei	$\lvert t \rvert > t_{0.025}(n-3)$	$\lvert t \rvert > t_{0.05}(n-3)$	$\lvert t \rvert < -t_{0.05}(n-3)$

Čia $t_{0.025}(n-3)$ – Stjudento kriterijaus su $n-3$ laisvės laipsniu 0.025 lygmens kritinė reikšmė (imama iš 3 lentelės).

3 Tikimybių teorija

3.1 Veiksmai su atsitiktiniais įvykiais

$$A \cup \emptyset = A, \qquad A \cup \Omega = \Omega, \qquad A \subset A \cup B, \qquad (A \cup B) \cap C = (A \cap C) \cup (B \cap C),$$
$$A \cap \emptyset = \emptyset, \qquad A \cap \Omega = A, \qquad A \cap B \subset A, \qquad A = (A \cap B) \cup (A \cap \overline{B}),$$
$$A \cup \overline{A} = \Omega, \qquad \overline{A \cup B} = \overline{A} \cap \overline{B}, \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

3.2 Kombinatorika

- Keliais skirtingais būdais galima n objektų išrikiuoti į eilę?
$$n! = 1 \cdot 2 \cdot 3 \cdots n.$$
- Keliais skirtingais būdais iš n objektų galima išrinkti k objektų (ir svarbi pakliuvimo į išrinktuosius eilė)?
$$n(n-1)(n-2)\cdots(n-k+1).$$
- Keliais skirtingais būdais iš n objektų galima išrinkti k objektų (ir nesvarbi pakliuvimo į išrinktuosius eilė)?
$$C_n^k = \binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

3.3 Paprasčiausia tikimybė

Dežėje a baltų, b juodų rutulių. Atsitiktinai paimame vieną rutulį. Kokia tikimybė, kad jis baltas?
$$P(baltas) = \frac{a}{a+b}.$$

3.4 Hipergeometrinė tikimybė

Dežėje a baltų, b juodų rutulių. Atsitiktinai paimame m. Kokia tikimybė, kad tarp paimtųjų yra k baltų rutulių?
$$P(k \text{ baltų iš } m) = \frac{\binom{a}{k}\binom{b}{m-k}}{\binom{a+b}{m}} = \frac{C_a^k C_b^{m-k}}{C_{a+b}^m}.$$

3.5 Tikimybės savybės

1) $P(\emptyset) = 0$.
2) $P(\Omega) = 1$.
3) $P(A \setminus B) = P(A) - P(A \cap B)$.
4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
5) $P(A \cup B) = P(A) + P(B)$, jeigu $A \cap B = \emptyset$.
6) $P(A) = P(A \cap B) + P(A \cap \overline{B})$.
7) $P(\overline{A}) = 1 - P(A)$.

3.6 Sąlyginė tikimybė

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Be to, jei $A \cap B = \emptyset$, tai

$$P(A \cup B|D) = P(A|D) + P(B|D).$$

3.7 Tikimybių sandaugos teorema

$$P(A \cap B) = P(B)P(A|B).$$

3.8 Nepriklausomi įvykiai

$$P(A \cap B) = P(A)P(B), \qquad P(A|B) = P(A).$$

3.9 Pilnosios tikimybės formulė

Tegul

1) $H_1 \cup H_2 \cup \cdots = \Omega$.

2) $H_m \cap H_k = \emptyset$, jei $m \neq k$.

Tada

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \ldots$$

3.10 Bajeso formulė

Tegul

1) $H_1 \cup H_2 \cup \cdots = \Omega$.

2) $H_m \cap H_k = \emptyset$, jei $m \neq k$.

Tada

$$P(H_j|A) = \frac{P(H_j)P(A|H_j)}{P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \ldots}$$

3.11 Bernulio schema

Atliekama n nepriklausomų identiškų eksperimentų. Vieno eksperimento sėkmės tikimybė lygi p. Tada

$$P(k \text{ iš } n) = \binom{n}{k} p^k (1-p)^{n-k} = C_n^k p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

3.12 Bernulio schemos apibendrinimas

Atliekama n nepriklausomų identiškų eksperimentų. Kiekvienas eksperimentas turi k galimų baigčių. Tų baigčių tikimybės yra $p_1, p_2, \ldots, p_k$, $p_1 + p_2 + \cdots + p_k = 1$. Ieškome tikimybės, kad po n bandymų bus m_1 pirmų baigčių, m_2 antrų baigčių, $\ldots$, m_k k-tų baigčių, ($n = m_1 + m_2 + \cdots + m_k$).

$$P(m_1, m_2, \ldots, m_k) = \frac{n!}{m_1! m_2! \cdots m_k!} p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}.$$

3.13 Atsitiktinių dydžių nepriklausomumas

- Jeigu X ir Y nepriklausomi, tai ir $f(X)$ ir $g(Y)$ nepriklausomi. Čia f, g bet kokios tolydžios funkcijos.
- C ir bet koks atisitiktinis dydis X nepriklausomi.

3.14 Absoliučiai tolydaus atsitiktinio dydžio tankis

Tankio funkcijos $p(x)$ savybės:

- $p(x) \geqslant 0$,
- $$\int_{-\infty}^{\infty} p(x)\ \mathrm{d}x = 1.$$

Ryšys su skirstiniu ir pasiskirstymo funkcija:

$$\begin{aligned} P(a \leqslant X \leqslant b) &= \int_a^b p(x)\ \mathrm{d}x, \\ F(x) = P(X \leqslant x) &= \int_{-\infty}^x p(t)\ \mathrm{d}t, \\ F'(x) &= p(x). \end{aligned}$$

3.15 Atsitiktinio dydžio vidurkis

Diskretaus atsitiktinio dydžio vidurkis:

X	x_1	x_2	x_3	$\dots$
P	p_1	p_2	p_3	$\dots$

$$\mathbf{E}X = x_1p_1 + x_2p_2 + x_3p_3 + \dots$$

$f(X)$ vidurkis: $\mathbf{E}f(X) = f(x_1)p_1 + f(x_2)p_2 + f(x_3)p_3 + \dots$

Pvz., $\mathbf{E}X^2 = x_1^2p_1 + x_2^2p_2 + x_3^2p_3 + \dots$

Tolydaus atsitiktinio dydžio vidurkis:

$$\mathbf{E}X = \int_{-\infty}^{\infty} xp(x)\ \mathrm{d}x.$$

Čia $p(x)$ – a.d. X tankis.

Vidurkio savybės:

- $\mathbf{E}X$ yra skaičius.
- $\mathbf{E}C = C$.
- $\mathbf{E}CX = C\mathbf{E}X$.
- $\mathbf{E}(X+Y) = \mathbf{E}X + \mathbf{E}Y$.
- Jeigu X ir Y nepriklausomi, tai $\mathbf{E}XY = \mathbf{E}X\mathbf{E}Y$.

3.16 Atsitiktinio dydžio dispersija

Apibrėžimas:

$$\mathbf{D}X = \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2.$$

Diskretaus atsitiktinio dydžio dispersija:

X	x_1	x_2	x_3	$\ldots$
P	p_1	p_2	p_3	$\ldots$

Tada

$$\mathbf{D}X = \left[x_1^2 p_1 + x_2^2 p_2 + x_3^2 p_3 \cdots\right] - \left(x_1 p_1 + x_2 p_2 + x_3 p_3 \cdots\right)^2.$$

Tolydaus atsitiktinio dydžio dispersija. Tarkime, kad $p(x)$ – a.d. X tankis. Tada

$$\mathbf{D}X = \int_{-\infty}^{\infty} x^2 p(x)\ \mathrm{d}x - \left(\int_{-\infty}^{\infty} x p(x)\ \mathrm{d}x\right)^2.$$

Dispersijos savybės:

- $\mathbf{D}X$ yra skaičius.
- $\mathbf{D}C = 0$.
- $\mathbf{D}CX = C^2\mathbf{D}X$.
- $\mathbf{D}(X+Y) = \mathbf{D}X + \mathbf{D}Y + 2\mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y)$.
- Jeigu X ir Y nepriklausomi, tai $\mathbf{D}(X+Y) = \mathbf{D}X + \mathbf{D}Y$.

3.17 Kovariacija

Atsitiktinių dydžių X ir Y kovariacija

$$\mathrm{Cov}(X,Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y) = \mathbf{E}XY - \mathbf{E}X\mathbf{E}Y.$$

Savybė:

$$\mathrm{Cov}(aX+b, cY+d) = ac\mathrm{Cov}(X,Y).$$

3.18 Koreliacija

Atsitiktinių dydžių X ir Y koreliacijos koeficientas

$$\varrho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbf{D}X\mathbf{D}Y}} = \frac{\mathbf{E}XY - \mathbf{E}X\mathbf{E}Y}{\sqrt{\mathbf{D}X\mathbf{D}Y}}.$$

3.19 Atsitiktinių dydžių pavyzdžiai

- Binominis a.d. (n kartų kartojame eksperimentą, kurio sėkmės tikimybė p; skaičiuojame kiek kartų eksperimentas pavyko): $X \sim \mathcal{B}(n,p)$.

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}=C_n^k p^k(1-p)^{n-k}, \quad k=0,1,\ldots,n.$$

  $\mathbf{E}X=np$, $\mathbf{D}X=np(1-p)$.

- Puasono a.d. (retų įvykių – razinų skč. bandelėje, korektūros klaidų lape, telefono skambučių per valandą – skaičius): $X \sim \mathcal{P}(\lambda)$.

$$P(X=k)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \qquad k=0,1,\ldots$$

  $\mathbf{E}X=\lambda$, $\mathbf{D}X=\lambda$.

- Geometrinis a.d. (Eksperimento sėkmės tikimybė p; kartojame kol pasiseka; X – bandymų iki pirmo pasisekimo skaičius).

$$P(X=k)=p(1-p)^{k-1}, \qquad k=1,2,\ldots$$

$$\mathbf{E}X=\frac{1}{p}, \qquad \mathbf{D}X=\frac{1-p}{p^2}.$$

- Hipergeometrinis a.d. (Tarp N objektų yra M žymėtų. Atrenkame n ir skaičiuojame žymėtus. X – žymėtų skaičius tarp atrinktųjų.) $\max(0,M+n-N)\le k\le \min(M,n)$.

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n},$$

$$\mathbf{E}X=\frac{nM}{N}, \qquad \mathbf{D}X=\frac{nM(N-M)(N-n)}{N^2(N-1)}.$$

- Normalusis a.d. $X\sim\mathcal{N}(\mu,\sigma^2)$. Tankis

$$\varphi(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big\{-\frac{(x-\mu)^2}{2\sigma^2}\Big\}, \qquad -\infty<x<\infty.$$

  $\mathbf{E}X=\mu$, $\mathbf{D}X=\sigma^2$.

3.20 Normaliojo skirstinio tikimybės

Normalusis a.d. $X\sim\mathcal{N}(\mu,\sigma^2)$, $a<b$, $\Phi(x)$ – standartinio normaliojo atsitiktinio dydžio pasiskirstymo funkcija. Tada

$$P(a<X<b)=\Phi\Big(\frac{b-\mu}{\sigma}\Big)-\Phi\Big(\frac{a-\mu}{\sigma}\Big),$$

$$P(X<b)=\Phi\Big(\frac{b-\mu}{\sigma}\Big), \qquad P(X>a)=1-\Phi\Big(\frac{a-\mu}{\sigma}\Big),$$

$$\Phi(-x) = 1 - \Phi(x).$$

Griežtos nelygybės ($<$), gali būtin pakeistos į negriežtas ($\leq$). Kartais praverčia tokia bendra tikimybės savybė:

$$P(a < |X| < b) = P(a < X < b) + P(-b < X < -a).$$

3.21 CRT taikymas apytiksliam skaičiavimui

Jeigu X yra nepriklausomų a.d. su baigtinėmis dispersijomis suma, tai

$$P(a < X < b) \approx \Phi\left(\frac{b - \mathbf{E}X}{\sqrt{\mathbf{D}X}}\right) - \Phi\left(\frac{a - \mathbf{E}X}{\sqrt{\mathbf{D}X}}\right),$$

$$P(X < b) \approx \Phi\left(\frac{b - \mathbf{E}X}{\sqrt{\mathbf{D}X}}\right), \qquad P(X > a) \approx 1 - \Phi\left(\frac{a - \mathbf{E}X}{\sqrt{\mathbf{D}X}}\right).$$