

Analýza a vizualizácia dát

18. marca 2024

generovanie

- ▶ rovnomerné rozdelenie
 - *RAND()*, *RANDARRAY()*
 - reálne čísla, celé čísla
 - pozor, prepisuje sa

cvičenie

- ▶ generujte realizácie z rovnomerného rozdelenia na intervale $(0, 1)$
- ▶ generujte realizácie z diskrétneho rovnomerného rozdelenia s nosičom $\{1, 2, \dots, 10\}$
- ▶ generujte realizácie z rovnomerného rozdelenia na intervale $(-5, -1)$
- ▶ vykreslite pomocou histogramu

generovanie

- ▶ diskrétno rozdelenie
- ▶ normálne rozdelenie
- ▶ *Data Analysis*
- ▶ *Seed*
- ▶ veta o inverznej transformácii

cvičenie

- ▶ generujte realizácie z diskrétného rozdelenia s nosičom $\{1, 2, \dots, 5\}$ s nerovnomernými pravdepodobnosťami
- ▶ generujte realizácie z normálneho rozdelenia s parametrami $\mu = 5, \sigma^2 = 4$
- ▶ vykreslite

odhad strednej hodnoty a disperzie

- ▶ výberový priemer, výberový rozptyl
- ▶ *AVERAGE()*, *VAR.S()*, *STDEV.S()*
- ▶ *Data Analysis*

interval spoľahlivosti pre strednú hodnotu

- ▶ obojstranný, jednostranný
- ▶ presný - $(1 - \alpha)100\%$ -ný (dáta z normálneho rozdelenia)
 - známa disperzia

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}$$

- neznáma disperzia

$$\bar{X} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(n - 1)$$

- ▶ *CONFIDENCE.NORM()*, *CONFIDENCE.T()*
- ▶ *Data Analysis* - iba pri neznámej disperzii

interval spoľahlivosti pre strednú hodnotu

- ▶ aproximatívny (dáta z nie normálneho rozdelenia)
- veľa dát (rádovo desiatky, cca nad 30)

$$\bar{X} \pm \frac{s}{\sqrt{n}} z_{\frac{\alpha}{2}}$$

$$\bar{X} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)$$

- ▶ *CONFIDENCE.T()*
- ▶ *Data Analysis*

interval spoľahlivosti pre disperziu

- ▶ presný (dáta z normálneho rozdelenia)

$$\left(\frac{s^2(n-1)}{\chi_{\frac{\alpha}{2}}^2(n-1)}, \frac{s^2(n-1)}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} \right)$$

cvičenie

- ▶ nájdite interval spoľahlivosti pre μ, σ^2 pre vygenerované dáta z normálneho rozdelenia s parametrami $\mu = 5, \sigma^2 = 4$
- ▶ porovnajte so skutočnými hodnotami

cvičenie

- ▶ nájdite interval spoľahlivosti pre μ pre vygenerované dáta z diskrétného rozdelenia s nosičom $\{1, 2, \dots, 5\}$ s nerovnomernými pravdepodobnosťami
- ▶ porovnajte so skutočnou hodnotou

binomický test

- ▶ problém: je táto minca vyvážená alebo je nevyvážená?
(dostane sa politická strana do parlamentu?)

$$H_0 : P(\{H\}) = p = \frac{1}{2} \quad \text{vs} \quad H_1 : p \neq \frac{1}{2}$$

- ▶ ako budeme riešiť?
 - hodíme mincou 100x, NP X - počet padnutých H
 - ak H_0 platí, tak $X|H_0 \sim \text{bin}(n = 100, p = 1/2)$
 - ak H_0 platí, tak očakávame +- 50 H
 - ak je počet H blízky 50, tak H_0 nezamietneme

cvičenie

- ▶ simulujte 100 hodov mincou s pravdepodobnosťou padnutia H $1/2$ a $1/3$, spočítajte deskriptívne štatistiky
- ▶ vykreslite pravdepodobnostné funkcie $\text{bin}(100; 1/2)$, $\text{bin}(100; 1/3)$, $\text{bin}(100; 0,9)$, $\text{bin}(100; 0,55)$ do jedného grafu

testovanie hypotéz

- ▶ formulovanie hypotézy ($p = 1/2$)
- ▶ nájdenie testovej štatistiky (potrebujeme poznať rozdelenie - počet H)
- ▶ nájdenie kritickej oblasti (oblasť, kam ak padne testová štatistika, tak zamietame H_0 - počet H ďaleko od 50)
- ▶ vyhodnotenie testu (zamietame/nezamietame H_0)

čo môže nastať

- ▶ H_0 platí, H_0 nezamietame
- ▶ H_0 platí, H_0 zamietame - chyba I. druhu
- ▶ H_0 neplatí, H_0 zamietame
- ▶ H_0 neplatí, H_0 nezamietame - chyba II. druhu

cvičenie

- ▶ simulujte 100 hodov mincou s pravdepodobnosťou padnutia H $1/3$
- ▶ nájdite pre tieto dáta hodnotu (realizáciu) testovacej štatistiky
- ▶ nájdite kritickú oblasť tohto testu
- ▶ vyhodnoňte test