

MA 0008 – teorie psti

cvičení 01: znak a průměr

Cvičení může probíhat před první přednáškou

Pst a stat je tak rozsáhlý obor, jako např. algebra nebo analýza

Proto v jedné knize nikdy nenajdete všechno. Přehled literatury:

Týden 1-2: Hindls: Statistika v ekonomii (2018)

Budíková, Králová, Maroš: Průvodce základními stat.met.(2009)

Týden 3-9: Fajmon, Hlavičková: Matematika 3 (2003)

Týden 10-12: Fajmon, Kolářek: Pst, stat a operační výzkum (2005)

Doplněk: Filip Danielka: Pst a stat v jazyku R (BP 2019)

Crawley: The R Book (2013, 2nd edition)

Podmínky absolvování předmětu:

- 3. ročník ...PST1: zápočet ... prověrka-a v 5.týdnu
kolokvium ... otázky probrané v týd. 1-6
- 2. ročník, MA0008: prověrka-a ... 5.týden semestru
prověrka-b ... 10.týden semestru
c-ústní část zkoušky ... zkouškové obd.

Rozdělení oborů tohoto předmětu (podle Budíková, Maroš, Králová 2009)

○ A) popisná statistika (týdny 1-2) ... descriptive statistics

Zpřehledňuje informace ukryté v souborech naměřených dat pomocí tabulek, grafů, funkcí, číselných hodnot.

○ B) teorie psti (týdny 3-8) ... probability theory

Zabývá se popisem náhodnosti v experimentech, kdy za stejných vstupních podmínek nastávají různé výsledky.

○ C) úsudková statistika (týdny 9-12) ... statistical inference

Buduje metody pro analýzu dat, kdy informace (data) získané z náhodného výběru jedinců zobecňuje na celou populaci; její součástí je teorie odhadu, testování statistických hypotéz, statistická predikce = předpověď.

A) Popisná statistika

Statistická jednotka ... elementární jednotka podrobená statistickému zpracování (např. zaměstnanec, student, firma, věc (např. doklad), událost (např. nehoda))

Jedna zpracovávaná jednotka má zpravidla několik statistických znaků

Statistický znak = statistická proměnná ... konkrétní vlastnost statistické jednotky (např. věk zaměstnance, mzda zaměstnance, nejvyšší dosažené vzdělání, atd.)

Statistické znaky = proměnné můžeme rozdělit na několik typů:

○ 1) kvantitativní znak (quantity = množství)

Lze vyjádřit číselně (např. počet členů domácnosti, spotřeba elektřiny...)

Kvantitativní znaky často dále dělíme na

○ **diskrétní znaky** ... nabývají oddělených číselných hodnot (\mathbb{N} , \mathbb{Z}^+)

např. počet členů domácnosti, počet výrobků

spojité znaky ... nabývají hodnot z intervalu

např. spotřeba elektřiny, doba čekání na příchod zákazníka k obslužné jednotce

Statistické znaky = proměnné můžeme rozdělit na několik typů:

○2) kvalitativní = nominální = kategoriální znak

Vyjadřuje se slovně (nomen = jméno, název) nebo subjektivním číslem

Kvalitativní znaky někdy dělíme na

alternativní ... nabývají dvou hodnot (např. ano-ne, muž-žena)

množné ... nabývají >2 hodnot (vzdělání: ZŠ, SŠ, VŠ, Ph.D.)

ordinální ... vyjádřené subjektivní stupnicí (např. míra spokojenosti s výrobkem vyjádřená číslem ze stupnice 1=hodně až 5=vůbec ne), známka ve škole)

Odišnost mezi kvantit. a kvalit. znaky:

- I kvalitativní znaky lze vyjádřit číslem a zpracovat počítačem:
 - alternativní** ... např. ano =1, ne=0
 - množné** ... např. ZŠ=1, SŠ=2, VŠ=3, Ph.D.=4)
 - ordinální** ... už je vyjádřeno číslem na stupnici 1 až 5
- Ovšem tyto číselné hodnoty jsou subjektivní, např. 2 minus 1 je číselně stejný rozdíl jako 3 minus 2, přesto nelze říci, že existuje stejný odstup mezi „rozhodně ano“ a „spíše ano“ jako mezi „nevím“ a „spíše ano“ (nebo ve škole nelze říci, že rozdíl mezi známkou 1 a 2 je hodnotově stejný jako rozdíl mezi známkou 4 a 5)

Jak získáme data? Měřením nebo dotazníkem

- **Základní soubor** ... soubor všech jednotek, na kterých má smysl sledovat určité znaky = proměnné. Zpravidla je velmi obsáhlý, někdy nekonečný, tj. změřit všechny jednotky je často nákladné nebo neproveditelné

A proto provádíme tzv.

- **Výběrové šetření** ... pro získání informací ze základního souboru vybereme jenom několik jednotek, měřením či dotazníkem získáme tzv. **výběrový soubor**

Jak získáme data? Měřením nebo dotazníkem

Tento výběrový soubor by měl být získán z **reprezentativního vzorku** populace = z takové množiny vybraných jedinců, který je věrnou zmenšeninou populace, má tedy stejné vlastnosti, výběr nestrání žádnému jednotlivci nebo skupině, tvoří ho jednotky pro základní soubor typické

(získat takový nestranný vzorek může být dost náročné)

Máme tedy výběrový soubor a chceme spočítat průměr jeho hodnot:

1. Aritmetický průměr hodnot: $\bar{x} = \frac{1}{n} \cdot \sum_{i=0}^n x_i$
2. Geometrický průměr hodnot: $\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$
3. Harmonický průměr hodnot: $\bar{x}_H = \frac{n}{\sum_{i=0}^n \frac{1}{x_i}}$

V jazyku R:

slova za znakem křížku jsou poznámky, které už do příkazu v jazyku R nepište (ani ten křížek ne, jen po napsání příkazu stiskněte ENTER)

Vsuvka ... jazyk R ... Hinds, str.35, př. 1.7

- `v<- c(4, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 11, 12, 12, 12, 12, 12, 12, 12, 12)`
přičadí výběrový soubor do vektoru „v“ (pozn.: c=column ... sloupcový vektor)
- `summary(v)` # najde průměr, medián, min, max, dolní (1.) kvartil, horní (3.) kvartil souboru naměřených hodnot
- `table(v)` # spočítá četnosti hodnot v souboru „v“

1. Aritmetický průměr = běžný průměr libovolných reálných hodnot

Aritmetický průměr hodnot: $\bar{x} = \frac{1}{n} \cdot \sum_{t=0}^n x_t$

Vážený aritmetický průměr hodnot, když máme k dispozici četnosti

jednotlivých hodnot: $\bar{x} = \frac{1}{\sum n_i} \cdot \sum_{t=0}^n n_t x_t$

➤ `mean(v)` # spočte průměr hodnot ve vektoru „v“

(pozor, záleží na velikosti písmen, funkce Mean neprovede nic)

příklad

Hinds, str.32, příklad 1.5 ... a) průměr, b) vážený průměr

2. Geometrický průměr ... vyjadřuje např. průměrný roční nárůst produkce za n let

- a) Příklad: Hodnoty růstu v procentech byly za posledních 5 let 101,3 %, 108,5 %, 100,6 %, 98,7 %, 102,3 %. Vypočtete průměrný roční nárůst za dané pětileté období:

- b) Hinds, str. 35, př. 1.8. ... geom. průměr, vážený geom. průměr

V jazyce R:

- `geometric <- function (x) exp(mean(log(x)))` # definice funkce
- `geometric(v)` # spočte průměr geometrický v souboru měření „v“

Důležitá poznámka k jazyku R:

Pokud v R nadefinujeme funkce `geometric(v)`, `harmonic(v)`, dále např. výpočet rozptylu (viz cvičení 2) pomocí funkce `rozptyl(v)` a dáme při vypínání prostředí R volbu uložit „workspace“ YES, tak při dalším spuštění je obsah všech proměnných, včetně všech funkcí, znovu nastaven, tj. lze pokračovat v rozdělané práci, používat nadefinované funkce, apod.

(při uložení několika různých workspace pod různými jmény je možné nahrát proměnné a funkce různých projektů podle názvu daného workspace)

3. Harmonický průměr = převrácená hodnota průměru převrácených hodnot:

Harmonický průměr hodnot: $\bar{x}_H = \frac{1}{\frac{1}{n} \sum_{i=0}^n \frac{1}{x_i}} = \frac{n}{\sum_{i=0}^n \frac{1}{x_i}}$

◦ V jazyce R:

- `harmonic <- function(x) 1/mean(1/x)` # **nadefinuje funkci**
- `harmonic(v)` # **vypočte harmonický průměr souboru „v“**

příklady

- a) Crawley, str. 118, slon chodí po obvodu čtverce o straně 2km, po každé straně jinou rychlostí (1 km/hod, 2 km/hod, 4 km/hod, 1km/hod) ... jaká je jeho průměrná rychlost?

- b) Hindls, str.36-37, př. 1.7 ... harm-průměr, vážený harm-průměr

- c) Hindls, str. 343-363 ... využití váženého harmonického průměru u Paascheho indexu v ekonomii!!!!

Domácí úkol:

Naučte se výpočty z této hodiny buď pomocí freewaru R, nebo se naučte tyto věci počítat na své kalkulačce

Do cvičení si noste minimálně kalkulačku

U prověrek budete moci použít laptop s jazykem R, nebo kalkulačku, ovšem telefony nebudou povoleny

Děkuji za pozornost

(matematika pro gymnázia – komb a pst)

- a) Statistické údaje o škodlivosti kouření vztahuje každý kuřák jen na ostatní kuřáky
- b) Při interpretaci statistických údajů buďte opatrní. Každý pátý člověk na světě je sice Číňan, ale mezi svými spolužáky patrně žádného nenajdete.
- c) Starší bratr snědl svůj oběd i oběd mladšího bratra. Oba snědli v průměru jeden oběd.