

MA 0008 – teorie psti

cvičení 02: popisná statistika

Příklady viz nová učebnice pro SŠ:

Robová, Hála, Calda: Komplexní čísla, kombinatorika, pravděpodobnost, statistika

Část STATISTIKA: str. 148-194, neučte se pojem výběrového rozptylu a výběrové směrodatné odchylky na str. 180-182

Jazyk R je pouze pro zájemce, všechno lze počítat i s kalkulačkou!!

Příklad A, str. 150:

- Jsou zadány četnosti jednotlivých typů SŠ, odkud jsou studenti
 - a) sestavte histogram četností z těchto dat
 - b) spočtete relativní četnosti a zobrazte je v kruhovém diagramu

Řešení v R:

- `barplot(c(48,20,160,92))` # nakreslí obdélníčky dané výšky
- `pie(c(48,20,160,92))` # nakreslí koláčový graf
- `relc<- (1/320)* c(48,20,160,92)` # relativní četnosti
- `pie(relc)` # nakreslí koláč četností relativních

Příklad B, str. 152:

- Jsou zadány velikosti prodaných obleků během jednoho týdne v dané prodejně ...
 - a) sestavte histogram četností a polygon četností z těchto dat,
 - b) sestavte tabulku relativních četností, kumulativních absolutních četností, kumulativních relativních četností pro tato data
 - c) určete modus a medián, průměr, rozptyl a směrodatnou odchylku velikostí obleku
 - d) Určete variační rozpětí a mezikvartilové rozpětí velikosti obleků

v R:

- `obleky<- c(39,41,40,42,41,40,42,42,40,43,42,41,43,39,42,41,42,39,41,37,43,41,38,43,42,41,40,41,38,40,40,39,41,40,42,40,41,42,40,43,38,39,41,41,42,45)`
- `hist(obleky,col=6:7,breaks=36.5:45.5) # histogram, strida barvy 6-7`
`# a stredy obdelniciku umisti do celočíselných hodnot, hranice jsou posunuty`
- `table(obleky) # spocte cetnosti`
- `x<- c(37,38,39,40,41,42,43,45) # opiseme hodnoty znaku do vektoru x`
- `y<- c(1,3,5,9,12,10,5,1) # opiseme cetnosti do y`
- `plot(x,y,pch=16) # nakresli body v modu 16 = vyplnene kolecko`
- `lines(x,y) # spoji nakreslene body ... najedeme na file – lze uložit obrazek v jpg, pdf`

v R, pokračování příkladu:

- `relx <- (1/length(obleky))*y # spocte rel cetnosti`
- `kumy <- y # do promenne kumy si pripravime vektor cetnosti,`
- `for (i in 2:length(kumy)) kumy[i] <- kumy[i]+kumy[i-1] # kum cetnosti jsou hotovy!!!!`
- `relkumy <- (1/length(obleky))*kumy # rel kum cetnosti`

c) Modus = 41 = median ... vidíme z tabulky četností

- `mean(obleky) # vypocte prumer`
- `rozptyl <- function (x) ((length(x)-1)/length(x))*var(x) # definuje funkci rozptylu`
- `rozptyl(obleky) # vypocte rozptyl merenych hodnot 2.534972`
- `sqrt(rozptyl(obleky)) # vypocte smerodatnou odchylku mereni 1.592159`

v R, dokončení příkladu:

d) Určete variační rozpětí a mezikvartilové rozpětí velikosti obleků:

- `max(obleky)-min(obleky)` # variační rozpětí
- `quantile(obleky, c(0.25,0.75))` # najde dolní a horní kvartil
- # odečtením obou hodnot máme mezikvartilové rozpětí

Příklad o 75 učitelích z Hindlse (str.23):

- Jsou zadány počty let praxe jednotlivých 75 učitelů...
 - a) sestavte intervalové rozdělení četností pro tato data,
 - b) vypočtete vážený průměr, vážený rozptyl a směrodatnou odchylku jen zhruba pomocí těchto četností.

Příklad o 75 učitelích z Hindlse (str.23):

Zadání tabulky dat:

- `mojedata<- data.frame(trida=numeric(0),praxe=numeric(0))`
- `mojedata<-edit(mojedata)`
- `# a) nadefinujeme sloupce „platová třída“ a „délka praxe“`
- `# b) edit(moje data) vyvolá tabulku, do které data napíšeme`
- `> attach(mojedata) # tento příkaz aktivizuje práci s tabulkou`

Příklad o 75 učitelích z Hindlse (str.23):

- `table(praxe)` # rozdělení četnosti je nedostatečné, protože ve většine skupin je málo měření ... musíme některé četnosti sloučit
- `hist(praxe)` # program si sám sloučí četnosti do interval delky 5 jednotek
- `hist(praxe, col=6:7, breaks= c(0,10,20,30,40,50))`
sloučí četnosti do intervalu delky 10

Abychom získali i četnosti číselně, musíme „nasekat“ hodnoty do intervalů:

- `meze<- c(0,10,20,30,40,50)`
- `intervaly<- cut(praxe, meze)`
- `table(intervaly)` # získáme četnosti (21,29,15,8,2)

A zbývá vypočítat průměr, rozptyl a odchylku:

Přesně, pokud máme k dispozici měření ve vektoru „praxe“:

- `mean(praxe) # = 16.68`
- `rozptyl(praxe) # = 104.7243`
- `sqrt(rozptyl(praxe)) # = 10.23349`

○ **# pokud bychom měli k dispozici jen tabulku četností, lze tyto parametry odhadnout pomocí vzorců pro četnosti, jako x_j vezmeme středy intervalů četností:**

$$\bar{x} \doteq \frac{1}{75} \cdot (21 \cdot 5 + 29 \cdot 15 + 15 \cdot 25 + 8 \cdot 35 + 2 \cdot 45) = 17.13333333$$

$$s^2 \doteq \frac{1}{75} \cdot (21 \cdot 5^2 + 29 \cdot 15^2 + 15 \cdot 25^2 + 8 \cdot 35^2 + 2 \cdot 45^2) - 17.13333333^2 = 110.1156$$

$$s \doteq \sqrt{110.1156} = 10.4936$$

Příklad D, str. 159: Domácí úkol – nastudujte (řešení v R viz následující tři slajdy)

- Jsou zadány kupní ceny bytů ve velkých městech v roce 2007 ...
 - a) provedte pro ně intervalové rozdělení četností
 - b) sestavte tabulku relativních četností, kumulativních absolutních četností, kumulativních relativních četností pro tato data

Příklad D v jazyce R:

Do vektoru „byty“ si zadáme naměřené ceny:

➤ `byty ← c(45061, 29031, 25436, 25078, 24567, 22768, 22425, 22215, 22088, 21794, 21456, 20894, 20319, 20162, 19221, 18200, 17332, 17327, 17217, 16369, 16343, 14897, 14546, 14316, 13829, 12975, 12736)`

○ **počet intervalů se doporučuje určit jako $1 + 3.3 * \log_{10} n$:**

➤ `1 + 3.3 * log10(27)` # to je zhruba 5.7, čili zaokr. 6 intervalů

Šířku intervalů tedy určíme jako $(\max - \min) / 6$:

➤ `(max(byty) - min(byty)) / 6` # spočte se 5387.5, zaokrouhline na 5400

Příklad D v jazyce R, druhá část:

Určíme meze s krokem 5400, které pokrývají všechna měření:

➤ `bmeze <- c(12700, 18100, 23500, 28900, 34300, 39700, 45100)`

nasekáme hodnoty do daných intervalů pomocí funkce `cut`:

➤ `binterval <- cut(byty, bmeze)`

➤ `table(binterval)` # získali jsme četnosti (11,11,3,1,0,1)

➤ `cetnost <- c(11,11,3,1,0,1)`

Relativní a kumulativní četnosti budou teď už malina

Příklad D v jazyce R, třetí část:

Relativní četnosti v jazyce R:

➤ `rcetnost <- (1/length(byty))* cetnost`

kumulativní četnosti:

➤ `kcetnost <- cetnost # jen priprava vektoru na kum cetnosti`

➤ `for (i in 2:length(kcetnost)) kcetnost[i]<-kcetnost[i]+kcetnost[i-1]`

➤ `rkcetnost <- (1/length(byty))*kcetnost`

`kcetnost ... vector kum cetnosti,`

`rkcetnost ... vector rel kum cetnosti`

**Příklady H a I (í), str. 170: Domácí úkol – nastudujte
(řešení v R máte na násl dvou slajdech)**

- Jsou zadány míry inflace v deseti následných letech
 - a) jaká je průměrná jednoroční míra inflace?
 - b) o kolik procent se zvýšila inflace celkově za 10 let?

V jazyku R, ad příklad H:

```
➤ inflace<- c(1.021, 1.039, 1.047, 1.018, 1.001, 1.028, 1.019, 1.025,  
1.028,1.063)
```

musíme inflaci přeložit do přesahů sta procent, ale v řeči desetinného čísla blízkého jedné

```
geometric(inflace) # vypocte geom prumer
```

(pokud jste ulozili workspace ze cviceni 1, R si nadefinovanou funkci geometric „pamatuje“ při svém dalším spuštění)

```
geometric(inflace)^(10) # procentualni narust za 10 let
```

Jazyk R, Ad příklad 1 (i), str. 170:

Z hodnot čistého zisku

➤ `zisk <- c(1,2.5,4.4,9.2,18.0)`

vypočteme hodnoty růstů řetěžitým dělením

```
rust <- c(1,1,1,1) # jen si předdefinujeme čtyřprvkový vektor
```

```
for (i in 1:4) rust[i] <- zisk[i+1]/zisk[i] # spočte podily zisku
```

```
geometric(rust) # vypočte se zhruba 2.06 ... každý rok se zisk  
firmy více než zdvojnásobí (poroste o 106 %)
```

Příklad L, str. 173:

- Pro daná data určete všechny kvartily a 85-procentní kvantil a 13-procentní kvantil

V jazyku R:

Vytvoříme soubor všech měření (zadáme 70 hodnot):

- `prikladL<- c(1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 6,6,6,6,6,6,6,7,7,7,7,7,7,7,7,7,7,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,9,9,9,9,9,9,9,10,10,10,10,10,10,10,10,10,11)`
- `quantile(prikladL,c(0.13,0.25,0.75,0.85))`
najde dane kvantily

Hindls, str. 44-45 (nebo uč. Pro SŠ, př. J-str.184):

- Podle variačního koeficientu porovnejte denní produkci ve dvou firmách: ve které firmě je denní produkce rovnoměrnější (= vykazuje menší výkyvy)?

výpočet: menší výkyvy bude vykazovat soubor s menším variačním koeficientem

- `xx<- c(1,2,2,3,2,4,2,1,2,4)`
- `yy<- c(6,6,5,8,9,4,4,6,5,7)`
- `sqrt(rozptyl(xx))/mean(xx) # variacni koef souboru xx`
- `sqrt(rozptyl(yy))/mean(yy) # variacni koef souboru yy`

Význam: variační koef udává v jedné hodnotě míru rozptylu měření ve srovnání s průměrem měření