

Příkazy pro výpočty v R-ku

- `load("soubor.RData")` načtení datového souboru

Popisné statistiky číselné proměnné

Označme X číselnou proměnnou

- `library(DescTools)` aktivace knihovny popisných statistik
- `summary(X)` popisné statistiky polohy
- `quantile(X, z)` výběrový z -kvantil (z je číslo mezi 0 a 1)
- `mean(X, trim = z)` průměr s nastavením parametru pro useknutí
- `HuberM(X)` Huberův odhad
- `var(X)` rozptyl
- `sd(X)` směrodatná odchylka
- `IQR(X)` mezikvartilové rozpětí
- `MAD(X)` mediánová absolutní odchylka od mediánu
- `CoefVar(X)` koeficient variace
- `Skew(X)` šikmost
- `Kurt(X)` špičatost
- `hist(X)` histogram
- `boxplot(X)` boxplot

Popisné statistiky kategorické proměnné

Označme X kategorickou proměnnou

- `table(X)` tabulka absolutních četností
- `cumsum(Y)` kumulativní součty vektoru Y
- `prop.table(table(X))` relativní četnosti
- `barplot(table(X))` sloupcový graf
- `pie(table(X))` koláčový graf

Vztah dvou proměnných

Označme X a Y dvě proměnné (mohou být číselné i kategorické)

- `plot(Y ~ X)` graf závislosti, funguje pro všechny typy proměnných
- `lines(X, Y)` přikreslí křivku procházející body X, Y do existujícího grafu
- `cov(X, Y)` koeficient výběrové kovariance pro dvě číselné proměnné
- `cor(X, Y)` Pearsonův korelační koeficient pro dvě číselné proměnné
- `tapply(Y, X, f)` z číselné proměnné Y spočítá funkci f zvlášť pro každou skupinu z kategorické proměnné X
- `table(X, Y)` kontingenční tabulka absolutních četností pro dvě kategorické proměnné
- `addmargins(table(X, Y))` kontingenční tabulka absolutních četností s marginálními součty
- `prop.table(table(X, Y), 1)` kontingenční tabulka řádkových procent (hodnota 2 dá sloupcová procenta) pro dvě kategorické proměnné

Hledání modusu

Označme X proměnnou

- `density(X, bw = z)` jádrový odhad hustoty s nastavením parametru bandwidth na hodnotu z
- `which.max(X)` vrátí pozici maximální hodnoty
- `hist(X, plot = F)` vypíše charakteristiky histogramu
- `hist(X, plot = F)$breaks` dělicí body histogramu
- `hist(X, plot = F)$counts` počty pozorování ve sloupcích histogramu
- `hist(X, plot = F)` vypíše charakteristiky histogramu

Další příkazy

Označme X proměnnou

- `round(X, z)` zaokrouhlí hodnoty X na z desetinných míst
- `sort(X)` uspořádá hodnoty proměnné X
- `unique(X)` vrátí každou hodnotu proměnné pouze jednou
- `scale(X)` vypočte standardizované hodnoty (z-skóry)
- `PlotQQ(X)` Q-Q plot pro normální rozdělení

Vypočty pravdepodobnosti

Diskretní rozdelení

- Binomicke rozdelení: n - počet pokusu, p - pst uspechu
 - `pbinom(k, n, p)` distribucni funkce
 - `dbinom(k, n, p)` pravdepodobnostni funkce
- Hypergeometricke rozdelení: w - počet bilych kouli v osudi, b - počet cernych kouli v osudi, n - počet kouli tazeny z osudi
 - `phyper(k, w, b, n)` distribucni funkce
 - `dhyper(k, w, b, n)` pravdepodobnostni funkce
- Geometricke rozdelení: p - pravdepodobnost uspechu
 - `pgeom(k, p)` distribucni funkce
 - `dgeom(k, p)` pravdepodobnostni funkce
- Poissonovo rozdelení: λ - stredni hodnota
 - `ppois(k, lambda)` distribucni funkce
 - `dpois(k, lambda)` pravdepodobnostni funkce

Spojita rozdelení

- Normalni rozdelení: μ - stredni hodnota, σ - smerodatna odchylka
 - `pnorm(x,mu,sigma)` distribucni funkce
 - `qnorm(x,mu,sigma)` kvantilova funkce
- Lognormalni rozdelení: velicina, jejich logaritmus ma normalni rozdelení s parametry: μ , σ
 - `plnorm(x,mu,sigma)` distribucni funkce
 - `qlnorm(x,mu,sigma)` kvantilova funkce
- Exponencialni rozdelení: λ - intenzita
 - `pexp(x,int)` distribucni funkce
 - `qexp(x,int)` kvantilova funkce

Hledání optimálního rozdělení

Označme X proměnnou

- `library(fitdistrplus)` aktivace knihovny s odhady různých rozdělení
- `descdist(X, discrete=FALSE, boot=1000)` graf šikmosti a špičatosti proměnné
- `fitdist(X, "rozdeleni")` odhadne parametry zvoleného rozdělení ("norm", "logis", "lnorm", "beta", "gamma", "weibul", "exp")
- `PlotQQ(chol, qdist = function(p) qllogis(p, location = coef(fit2)[1], scale = coef(fit2)[2]))` ukázka Q-Q plotu pro logistické rozdělení s danými parametry

Intervaly spolehlivosti

Označme X číselnou proměnnou a Y kategorickou proměnnou

- `library(DescTools)` aktivace knihovny popisných statistik
- `MeanCI(X, conf.level = z)` interval spolehlivosti pro průměr se spolehlivostí z
- `MeanDiffCI(X ~ Y, conf.level = z)` interval spolehlivosti pro rozdíl průměrů se spolehlivostí z
- `BinomCI(k, n, conf.level = z, method = "wald")` interval spolehlivosti pro podíl se spolehlivostí z (k je počet úspěchů, n počet pokusů)
- `BinomDiffCI(k1, n1, k2, n2, conf.level = z)` interval spolehlivosti pro rozdíl podílů se spolehlivostí z ($k1, k2$ jsou počty úspěchů, $n1, n2$ počty pokusů)
- `BootCI(X, conf.level = z, FUN = f)` bootstrapový interval spolehlivosti pro funkci f se spolehlivostí z

Lineární regrese

Označme X a Y číselné proměnné

- `plot(Y ~ X)` bodový graf závislosti Y na X
- `abline(model1)` do existujícího grafu zakreslí přímku lineární regrese dané modelem 1
- `model1 <- lm(Y ~ X)` vytvoří model lineární regrese závislosti Y na X a uloží ho do proměnné `model1`
- `coef(model1)` koeficienty modelu
- `summary(model1)` základní shrnutí modelu
- `confint(model1)` intervaly spolehlivosti pro regresní parametry
- `plot(model1)` grafické testy předpokladů