

BI-PST Domácí úkol

Patrik Jantošovič

Tomáš Zvara

Tomáš Janecký

2. prosince 2018

1 PARAMETRY A DATOVÝ SOUBOR

Reprezentant: Patrik Jantošovič

$K = \text{den narození} = 16$

$L = \text{počet písmen v příjmení} = 10$

$M = ((K+L)*46) \bmod 11 + 1 = 2$

Výsledkem je teda datový soubor: case0102, mzda dle pohlaví

1.1 VYTVOŘENÍ DATOVÉHO SOUBORU

Řešení úloh předpokládá úspěšnou instalaci knihovni Sleuth2 a vytvoření .csv souboru s příslušnými datami.

Postup uvedeme jednou na začátku aby jsme se neopakovali.

- » `install.packages("Sleuth2")`
 - Instalace package Sleuth2
- » `library(Sleuth2)`
 - Načítání package Sleuth2
- » `write.table(case0102, "C:/data.csv", row.names=F, sep=";", dec=",")`
 - Zápis dat do .csv souboru

2 ŘEŠENÍ ÚKOLŮ

2.1 ÚKOL ČÍSLO 1

(1b) Načtěte datový soubor a rozdělte sledovanou proměnnou na příslušné dvě pozorované skupiny. Data stručně popište. Pro každou skupinu zvlášť odhadněte střední hodnotu, rozptyl a medián příslušného rozdělení.

- » `data<-read.table("C:/data.csv",header=TRUE,sep=";")`
 - Načteme data z přiraveného souboru
- » `female<-data[1:61,]`
 - Načítání dat pro pozorovanou skupinu: Female
- » `male<-data[62:93,]`
 - Načítání dat pro pozorovanou skupinu: Male
- » `female<-female[,1]`
 - Odseknutí sloupce s pohlavím pro pozorovanou skupinu: Female
- » `male<-male[,1]`
 - Odseknutí sloupce s pohlavím pro pozorovanou skupinu: Male
- » `length(male)`
 - Velkost dat pro pozorovanou skupinu: Male
- » `length(female)`
 - Velkost dat pro pozorovanou skupinu: Female
- » `var(male)`
 - Rozptyl pro pozorovanou skupinu: Male
- » `var(female)`
 - Rozptyl pro pozorovanou skupinu: Female
- » `mean(male)`
 - Střední hodnota pro pozorovanou skupinu: Male
- » `mean(female)`
 - Střední hodnota pro pozorovanou skupinu: Female
- » `median(male)`
 - Medián pro pozorovanou skupinu: Male

- » median(female)
 - Medián pro pozorovanou skupinu: Female

Výsledky zapíšeme do následující tabulky:

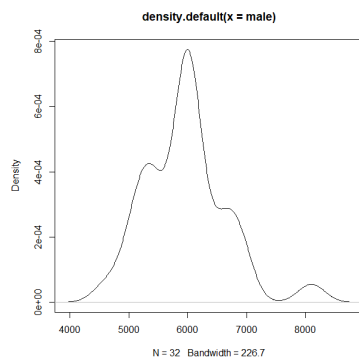
Pohlavie	Velkost dat	Střední hodnota	Rozptyl	Medián
Male	32	5956.875	477112.5	6000
Female	61	5138.852	291460.3	5220

2.2 ÚKOL ČÍSLO 2

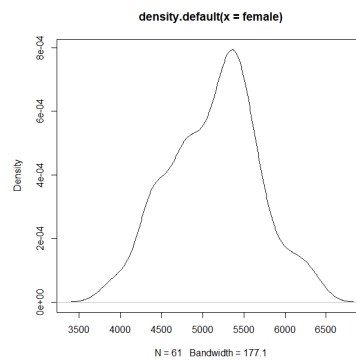
(1b) Pro každou skupinu zvlášť odhadněte hustotu a distribuční funkci pomocí histogramu a empirické distribuční funkce.

- » hist(female, freq=FALSE)
 - Vykreslení histogramu female. freq=FALSE používáme jako přepínač pro hustotu
- »hist(male, freq=FALSE)
 - Vykreslení histogramu female. freq=FALSE používáme jako přepínač pro hustotu
- »plot(density(male))
 - Vykreslení hustoty Male
- »plot(density(female))
 - Vykreslení hustoty Female
- »plot(ecdf(male))
 - Vykreslení empirický distribuční funkce pro Male
- »plot(ecdf(female))
 - Vykreslení empirický distribuční funkce pro Female

Výsledkem jsou grafy přiložené na následující stránce.

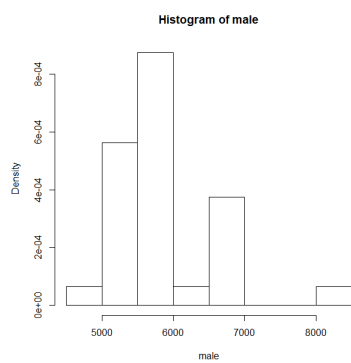


(a) lMale

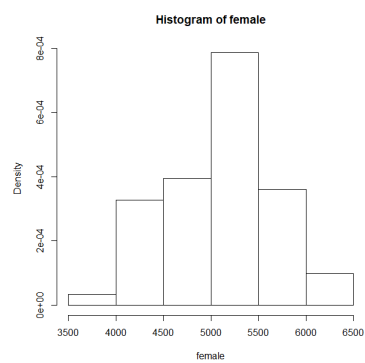


(b) Female

Obrázek 2.1: Hustota

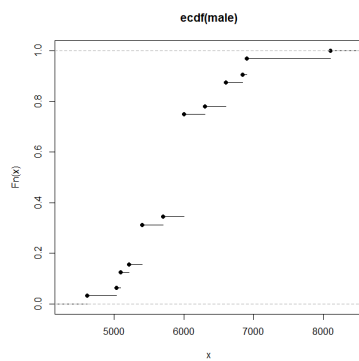


(a) Male

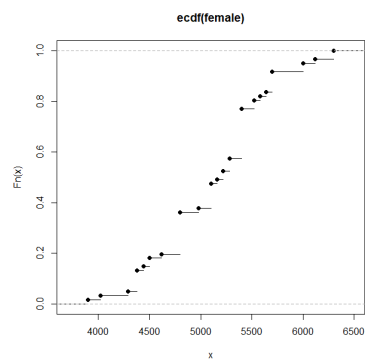


(b) Female

Obrázek 2.2: Histogram



(a) Male



(b) Female

Obrázek 2.3: Empirická distribuční funkce

2.3 ÚKOL ČÍSLO 3

(3b) Pro každou skupinu zvlášť najděte nejbližší rozdělení: Odhadněte parametry normálního, exponenciálního a rovnoměrného rozdělení. Zaneste příslušné hustoty s odhadnutými parametry do grafů histogramu. Diskutujte, které z rozdělení odpovídá pozorovaným datům nejlépe.

- Male

- » `hist(male, freq=FALSE)`

- * Porovnáváme distribuční funkce různých rozdělení na histogramu.

- Normální rozdělení

- * » `maleV<-seq(min(male),max(male),10)`

- vytvoříme si sekvenci hodnot od nejmenší po největší hodnoty

- * » `maleNorm<-dnorm(maleV, mean = mean(male), sd = sd(male))`

- využijeme funkci `dnorm` na převod pro body normálního rozdělení

- * » `lines(maleV,maleNorm, col="blue")`

- vykreslíme normální rozdělení na histogram

- Exponenciální rozdělení

- * » `lambdaMale<-1/mean(male)`

- vypočteme si parametr pro exponenciální rozdělení jako $1 / \text{střední hodnota}$

- * » `maleExp<-dexp(maleV, lambdaMale)`

- využijeme funkci `dexp` na výpočet bodu exponenciálního rozdělení

- * » `lines(maleV,maleExp, col="red")`

- vykreslíme exponenciální rozdělení na histogram

- Uniformní rozdělení

- * » `aMale<-mean(male)-sqrt(3*var(male))`

- * » `bMale<-sqrt(3*var(male))+mean(male)`

- vypočteme si parametr 'a' a 'b' pro uniformní rozdělení podle vztahu k střední hodnotě a rozptylu ze cvičení

- * » `maleUnif<-dunif(maleV, aMale,bMale)`

- využijeme funkci `dunif` na výpočet bodu uniformního rozdělení

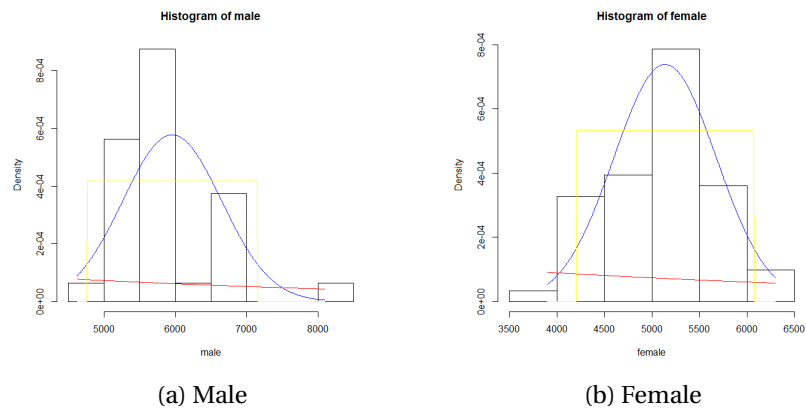
- * » `lines(maleV,maleUnif, col="yellow")`

- vykreslíme uniformní rozdělení na histogram

- Female

- » `hist(female, freq=FALSE)`
 - * Porovnáваме distribuční funkce různých rozdělení na histogramu.
- Normální rozdělení
 - * » `femaleV<-seq(min(female),max(female),10)`
 - vytvoříme si sekvenci hodnot od nejmenší po největší hodnoty
 - * » `femaleNorm<-dnorm(femaleV, mean = mean(female), sd = sd(female))`
 - využijeme funkci `dnorm` na převod pro body normálního rozdělení
 - * » `lines(femaleV,femaleNorm, col="blue")`
 - vykreslíme normální rozdělení na histogram
- Exponenciální rozdělení
 - * » `lambdaFemale<-1/mean(female)`
 - vypočteme si parametr pro exponenciální rozdělení jako $1 / \text{střední hodnota}$
 - * » `femaleExp<-dexp(femaleV, lambdaFemale)`
 - využijeme funkci `dexp` na výpočet bodu exponenciálního rozdělení
 - * » `lines(femaleV,femaleExp, col="red")`
 - vykreslíme exponenciální rozdělení na histogram
- Uniformní rozdělení
 - * » `aFemale<-mean(female)-sqrt(3*var(female))`
 - * » `bFemale<-sqrt(3*var(female))+mean(female)`
 - vypočteme si parametr 'a' a 'b' pro uniformní rozdělení podle vztahu k střední hodnotě a rozptylu ze cvičení
 - * » `femaleUnif<-dunif(femaleV, aFemale,bFemale)`
 - využijeme funkci `dunif` na výpočet bodu uniformního rozdělení
 - * » `lines(femaleV,femaleUnif, col="yellow")`
 - vykreslíme uniformní rozdělení na histogram

Výsledkem jsou následující grafy:



Obrázek 2.4: Porovnání rozdělení

Došli sme k záveru že se u obou datasetu jedná o Normální rozdělení.