homework

December 13, 2018

1 case0222 - Cholesterol In Urban And Rural Guatemalans

https://www.rdocumentation.org/packages/Sleuth2/versions/2.0-4/topics/ex0222

1.0.1 Popis dat

Dataset pochází ze studie provedené na guatemalských indiánech. Míra cholesterolu byla zmena celkem 94 jedincm a byl zaznamenán jejich pvod. Bylo nameno 49 pozorování na venkov a 45 ve mst.

1.0.2 Formát

Dataframe obsahuje 94 pozorování na následujících 2 promnných:

- **Cholesterol** - Mnství cholesterolu v krvi lovka (v mg/l).
- **Group** - Promnná obsahující hodnoty "Rural" a "Urban" oznaující, jestli je subjekt z venkova, nebo z msta.

1.0.3 Zdroj

Ramsey, F.L. and Schafer, D.W. (2002). The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed), Duxbury.

```
In [2]: library(Sleuth2)
        str(ex0222)
```

```
'data.frame':        94 obs. of  2 variables:
 $ Cholesterol: num  133 134 155 170 175 179 181 184 188 189 ...
 $ Group      : Factor w/ 2 levels "Rural","Urban": 2 2 2 2 2 2 2 2 2 2 ...
```

- [x] (1b) Natte datový soubor a rozdlte sledovanou promnnou na písluné dv pozorované skupiny. Data strun popite. Pro kadu skupinu zvlá odhadnte stední hodnotu, rozptyl a medián písluného rozdlení.

```
In [3]: rural <- subset(ex0222, Group=="Rural", Cholesterol, drop=TRUE)
        urban <- subset(ex0222, Group=="Urban", Cholesterol, drop=TRUE)
```

Subjekty z venkova:

```
In [4]: cat("Rural area indians:\n")
        cat("EX =", mean(rural), "\n")
        cat("varX =", var(rural), "\n")
        cat("median =", median(rural))
```

```
Rural area indians:
EX = 157
varX = 1008.458
median = 152
```

Subjekty z msta:

```
In [6]: cat("Urban area indians:\n")
        cat("EX =", mean(urban), "\n")
        cat("varX =", var(urban), "\n")
        cat("median =", median(urban))
```

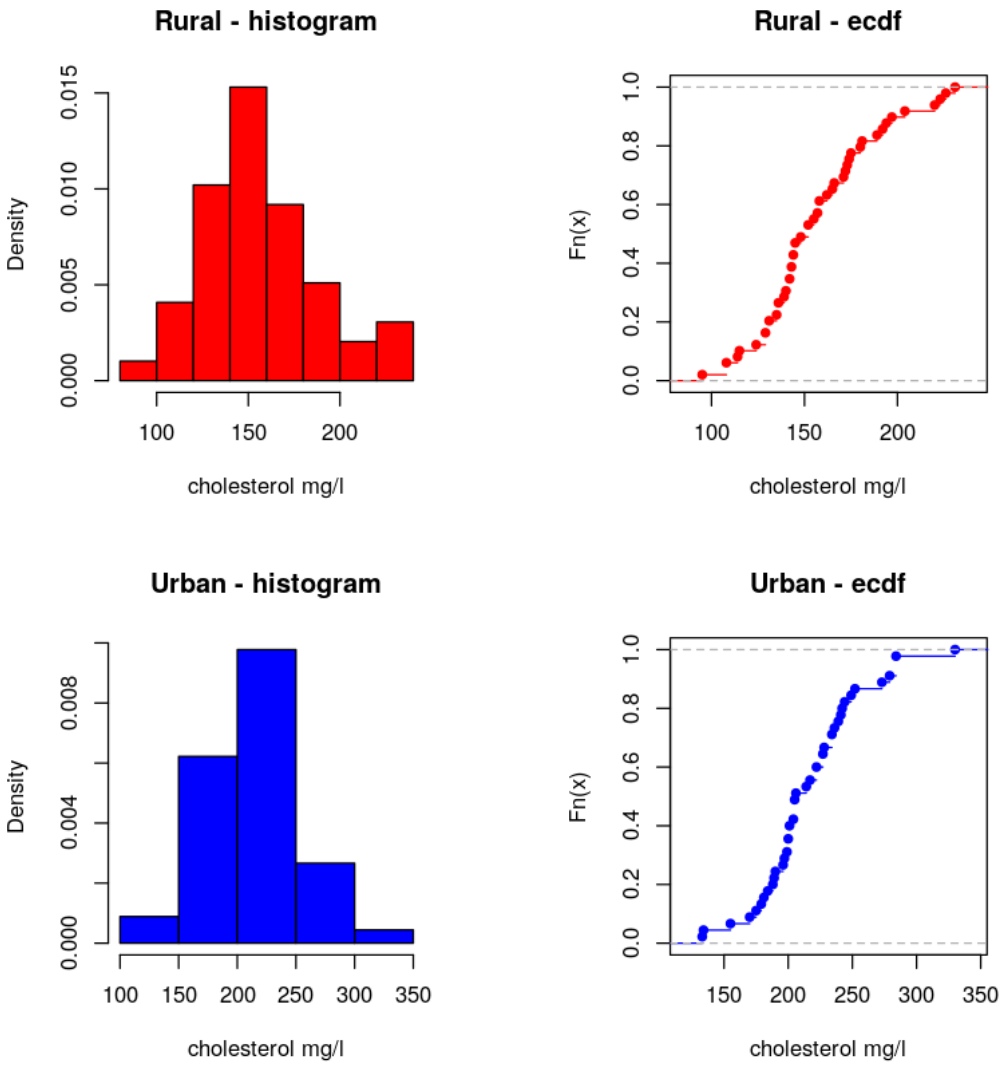
```
Urban area indians:
EX = 216.8667
varX = 1593.618
median = 206
```

- [x] (1b) Pro kadou skupinu zvlá odhadnte hustotu a distribuní funkci pomocí histogramu a empirické distribuní funkce.

```
In [18]: par(mfrow = c(2, 2),  pty="s")

         # rural
         hist(rural, col="red", main="Rural - histogram", probability=T, xlab="cholesterol mg/l
         plot.ecdf(rural, col="red", main="Rural - ecdf", xlab="cholesterol mg/l")

         # urban
         hist(urban, col="blue", main = "Urban - histogram", probability=T, xlab="cholesterol m
         plot.ecdf(urban, col="blue", main="Urban - ecdf", xlab="cholesterol mg/l")
```



- [ ] (3b) Pro kadou skupinu zvlá najdte nejblií rozdlení: Odhadnte parametry normálního, exponenciálního a rovnomrného rozdlení. Zaneste písluné hustoty s odhadnutými parametry do graf histogramu. Diskutujte, které z rozdlení odpovídá pozorovaným datm nejlépe.

Odhady rozdlení Pro provedení odhadu jsou vyuity funkce mean() a sd() zabudované do standardní knihovny jazyka R. Odhad je proveden shodn i pro mnoinu urban.

Normální rozdlení

```
EX = mean(rural)
s = sd(rural)
```

Exponenciální rozdlení

```
lambda = 1/mean(rural)
```

Uniformní rozdlení

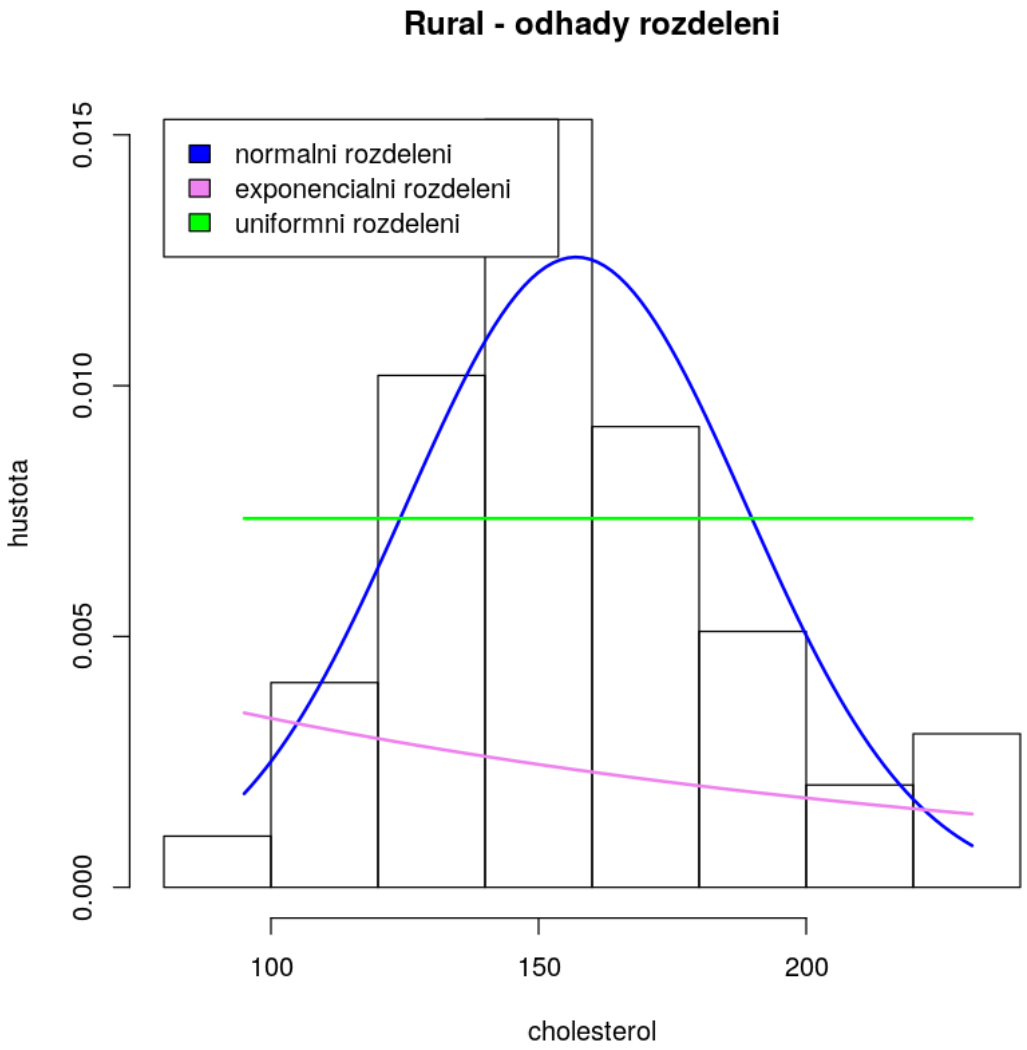
```
a = min(rural)
b = max(rural)
```

```
In [16]: x <- seq(min(rural), max(rural), length=100)

         # hodnoty pro jednotlivá rozloení
         y_norm <- dnorm(x, mean=mean(rural), sd=sd(rural))
         y_exp <- dexp(x, 1/mean(rural))
         y_unif <- dunif(x, min=min(rural), max=max(rural))

         hist(rural, probability=T, main="Rural - odhady rozdeleni", xlab="cholesterol", ylab="
         lines(x, y_norm, col="blue", lwd=2)
         lines(x, y_exp, col="violet", lwd=2)
         lines(x, y_unif, col="green", lwd=2)

         legend("topleft", inset=0.037, fill=c("blue","violet","green"),
                legend=c("normalni rozdeleni", "exponencialni rozdeleni", "uniformni rozdeleni"
```

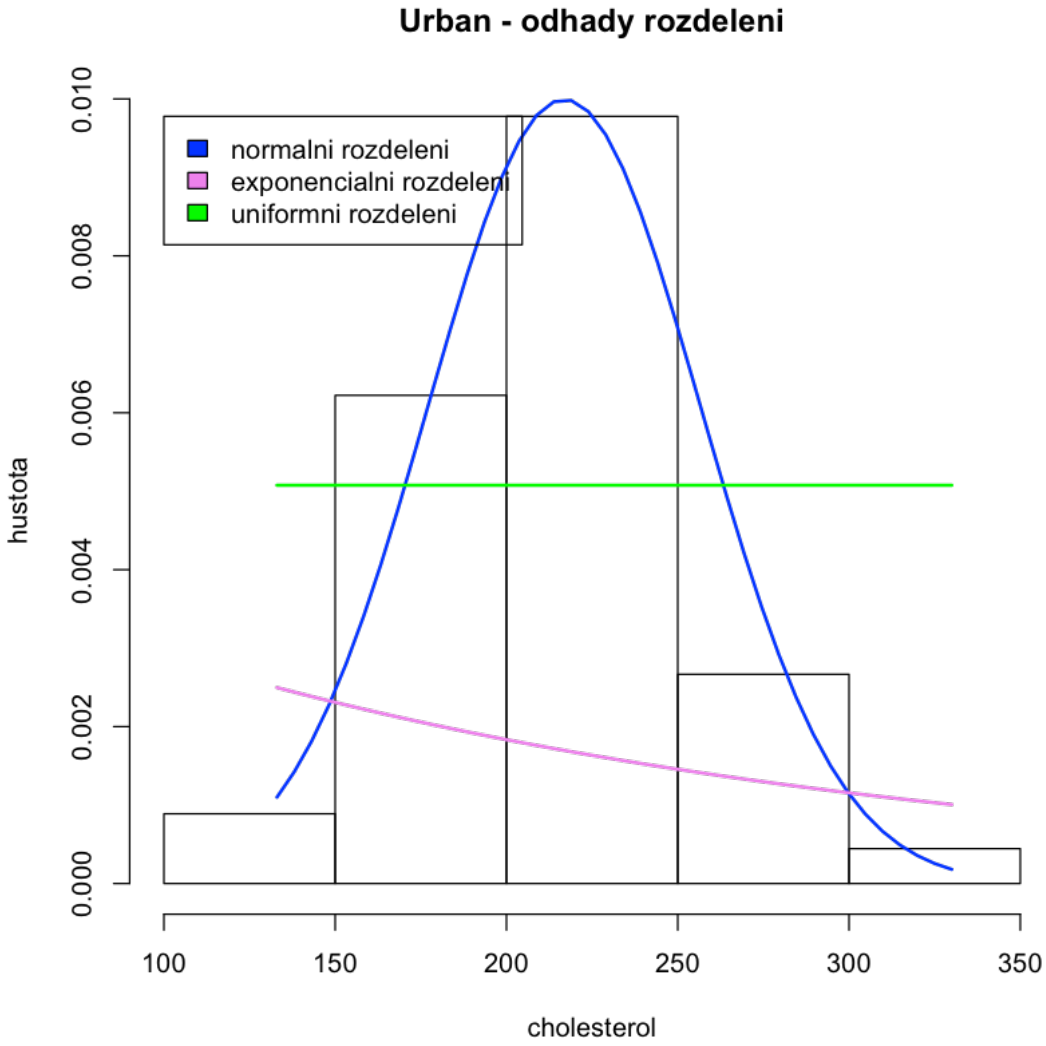



```
In [64]: x <- seq(min(urban), max(urban), length=40)

         y_norm <- dnorm(x, mean=mean(urban), sd=sd(urban))
         y_exp <- dexp(x, 1/mean(urban))
         y_unif <- dunif(x, min=min(urban), max=max(urban))

         hist(urban, probability=T, main="Urban - odhady rozdeleni", xlab="cholesterol", ylab="
         lines(x, y_norm, col="blue", lwd=2)
         lines(x, y_exp, col="violet", lwd=2)
         lines(x, y_unif, col="green", lwd=2)

         legend("topleft", inset=0.037, fill=c("blue","violet","green"),
                legend=c("normalni rozdeleni", "exponencialni rozdeleni", "uniformni rozdeleni"
```



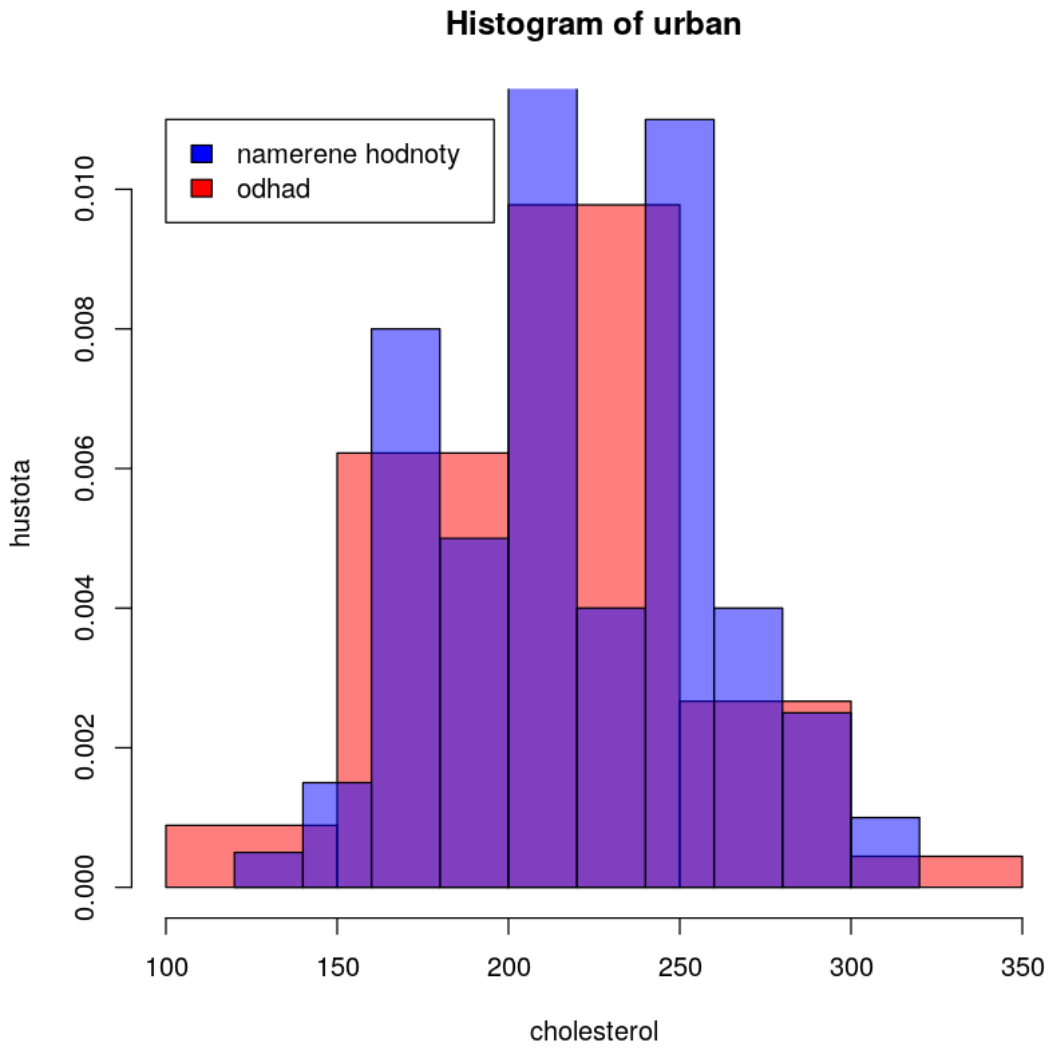
- [ ] (1b) Pro kadou skupinu zvlá vygenerujte náhodný výbr o 100 hodnotách z rozdlení, které jste zvolili jako nejblií, s parametry odhadnutými v pedchozím bod. Porovnejte histogram simulovaných hodnot s pozorovanými daty.

Na ná dataset se nejvíc hodí normální rozdlení. Parametry jsme odhadli pomocí knihovních funkcí mean() a sd(). Samotný náhodný výbr jsme vygenerovali následujícím píkazem:

```
rnorm(100, mean=mean(urban), sd=sd(urban))
```

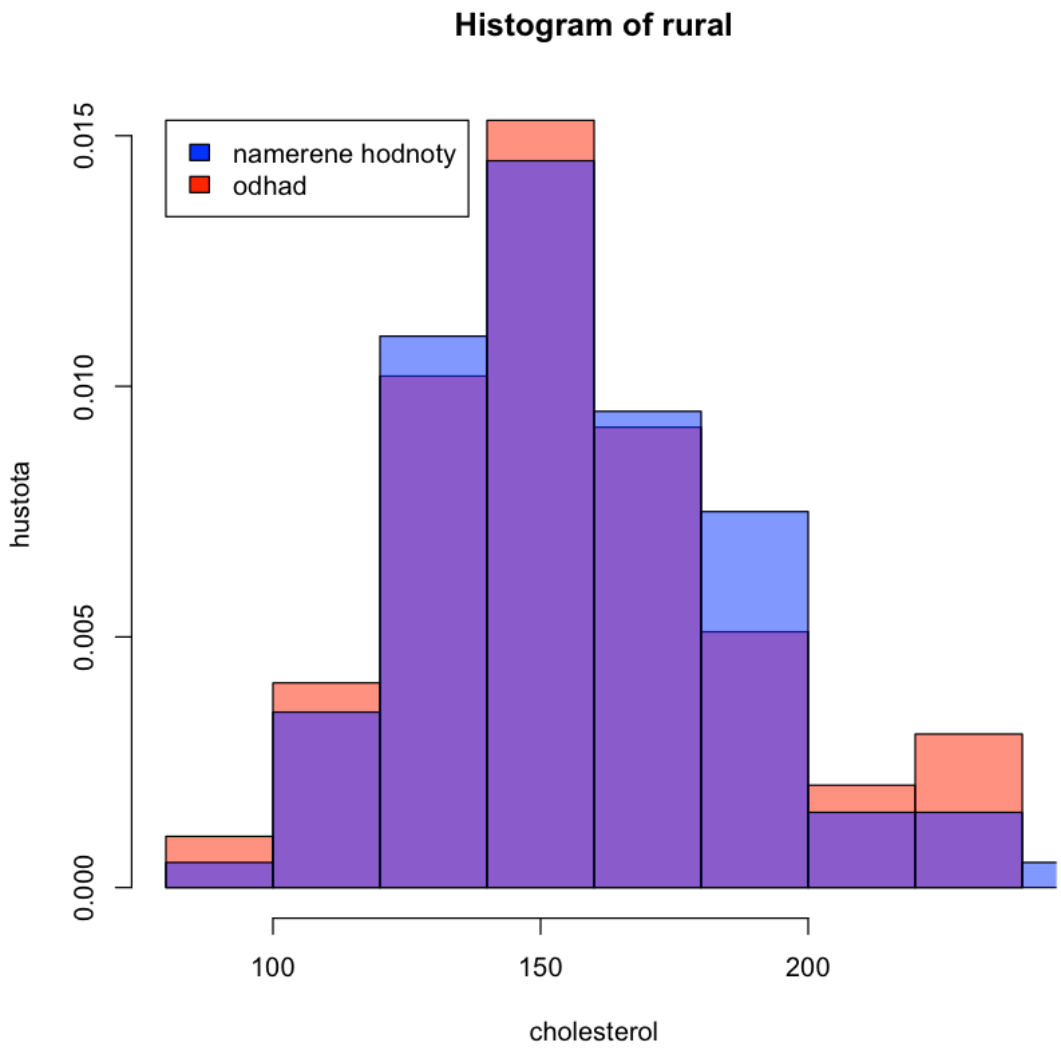
```
In [19]: y <- rnorm(100, mean=mean(urban), sd=sd(urban))

         hist(urban, probability=T, col=rgb(1, 0, 0, 0.5), ylim=c(0, 0.011), xlab="cholesterol"
         hist(y, probability=T, col=rgb(0, 0, 1, 0.5), add=T)
         legend("topleft", inset=0.037, fill=c("blue","red"),
                legend=c("namerene hodnoty", "odhad"))
```




```
In [66]: y <- rnorm(100, mean=mean(rural), sd=sd(rural))

         hist(rural, probability=T, col=rgb(1, 0, 0, 0.5), xlab="cholesterol", ylab="hustota")
         hist(y, probability=T, col=rgb(0, 0, 1, 0.5), add=T)
         legend("topleft", inset=0.037, fill=c("blue","red"),
                legend=c("namerene hodnoty", "odhad"))
```



- [ ] (1b) Pro kadou skupinu zvlá spoítejte oboustranný 95% konfidenní interval pro stední hodnotu.

```
In [6]: length(rural)
```

49

- [ ] (1b) Pro kadou skupinu zvlá otestujte na hladin významnosti 5% hypotézu, zda je stední hodnota rovná hodnot K (parametr úlohy), proti oboustranné alternativ. Mete pouít bu výsledek z pedelého bodu, nebo výstup z písluné vestavné funkce vaeho softwaru.

```
In [144]: # TODO
```

- [ ] (2b) Na hladin spolehlivosti 5% otestujte, jestli mají pozorované skupiny stejnou stední hodnotu. Typ testu a alternativy stanovte tak, aby vae volba nejlépe korespondovala s povahou zkoumaného problému.

```
In [145]: # TODO
```