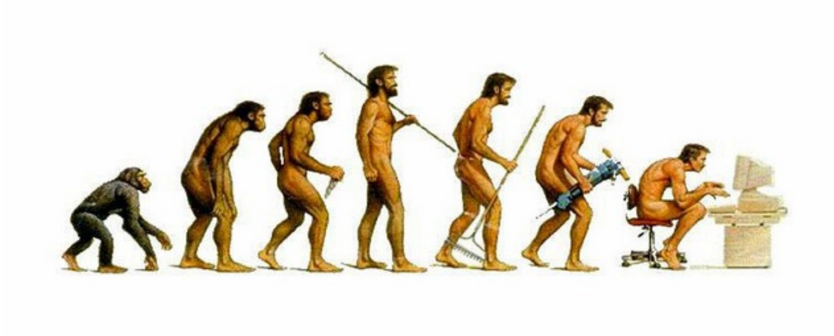
Masarykova univerzita v Brně
Přírodovědecká fakulta

SBÍRKA ŘEŠENÝCH PŘÍKLADŮ K PŘEDMĚTU APLIKOVANÁ STATISTIKA I

(Verze s výsledky)



Brno, 2015

1 - Základní práce se softwarem R - Příkazy

```
#promenna, vektor, matice
a<-3
a<-c(1,2,3)
vec<-c(1.1,5.3,6.4)
(A<-matrix(c(1,2,3,4,5,6),2,3,byrow=T))
(B<-matrix(c(1,2,3,4,5,6),ncol=3,nrow=2,byrow=T))

# zakladni operace
3+2-6*9/(8+9-5)
a<-15
b<-5
(a+b)/b

# scitani vektoru a matic
x<-c(1,2,3)
y<-c(3,2,1)
x+y

z<-c(0,1,2,3)
x+y+z

B<-matrix(c(1,1,1,1,1,1),2,3)
A-B

#dimenze vektoru a matice
length(a)
dim(A)

# Operace s promennymi
#mocnina
3^2
a<-4
(a2<-a^2)
x
(x2<-x^2)
A
(A2<-A^2)

# odmocnina
sqrt(9)
sqrt(a2)
sqrt(x2)
sqrt(A2)

# min a max
min(a)
max(a)
x
min(x)
max(x)
A
min(A)
max(A)

# absolutni hodnota
```

```
(C<-(-1)*A)
abs(C)
(y<-c(-1,0,2,-5))
abs(y)

# log/exp
log(3) # ln()
log(3,10)
log(9,3)

exp(3)
exp(log(3))

# sum
sum(x)
sum(A)

#zaokrouhlovani
(odmocnina<-sqrt(2))
round(odmocnina, digits=3)
round(odmocnina,digits=2)
ceiling(odmocnina)
floor(odmocnina)
signif(odmocnina, digits=6)
signif(odmocnina, digits=3)

#vytvareni posloupnosti
#CTRL+L
#Clear workspace
(x<-1:10)
(y<-50:55)

(pst1<-seq(from=0,to=1,length=1000))
(pst2<-seq(from=0,to=1,by=0.1))

vaha<-c(58.7, 61.6, 57.8, 59.5, 59.9, 53.9,63.6, 71.0, 66.1, 69.8)
(divky<-rep(1,6))
(chlapci<-rep(2,4))
(pohlavi<-c(divky,chlapci))

#rbind/cbind
vaha
pohlavi
(hmotnost<-matrix(c(vaha,pohlavi),nrow=10))
(hmotnost.c<-cbind(vaha,pohlavi))
(hmotnost.r<-rbind(vaha,pohlavi))

#podmnoziny
hmotnost
hmotnost[ ,1]
hmotnost[ ,2]
hmotnost[6, ]
hmotnost[1, ]
hmotnost[8,1]

vyska<-c(133,132,145,129)
vyska[4]
```

```
apply(hmotnost,1,sum)
apply(hmotnost,2,sum)

#porovnavani < > == <= >=
teploty<-c
   (10,9,9,8,8,9,11,12,13,14,16,18,18,19,18,16,15,14,14,13,13,14,14,14)
hodiny<-1:24
mean(teploty)
teploty==13.0
(1*(teploty==13))
1*(teploty<13)
1*(teploty<=13)
1*(teploty>13)
sum(1*(teploty==13))
sum(1*(teploty>13))
sum(1*(teploty<13))

#grafy
plot(hodiny,teploty,main='Teplota␣12.9.14',xlab='hodina',ylab='teplota',
      cex=1.2,pch=19,col='orchid4',lwd=2,bg='orchid4',type='l',xlim=c(0,25),
      ylim<-c(7,20))
legend(18,10,legend='teplota␣-␣12.9',fil='orchid4')

plot(hodiny,teploty,main='Teplota␣12.9.14',xlab='hodina',ylab='teplota',
      cex=1.2,pch=19,col='orchid1',lwd=2,type='l',xlim=c(0,25),
      ylim<-c(7,20))
points(hodiny,teploty,cex=1.2,pch=19,col='orchid4')
legend(18,10,legend='teplota␣-␣12.9',fil='orchid4')

pdf('pocasi.pdf')
plot(hodiny,teploty,main='Teplota␣12.9.14',xlab='hodina',ylab='teplota',
      cex=1.2,pch=19,col='dodgerblue',lwd=2,bg='red',type='n',xlim=c(0,25),
      ylim<-c(7,20))
lines(hodiny,teploty,lwd=2,col='orchid1')
points(hodiny,teploty,cex=1.2,pch=19,col='orchid4')
legend(18,10,legend='teplota␣-␣12.9',fil='orchid4')
dev.off()

# prace s datovym souborem
getwd()
setwd('C:/Users/Veronika/Documents')
dir()
setwd('C:/Users/Veronika/Documents/Data_cviceni_txt')
getwd()
dir()

data<-read.delim('znamky.txt',sep='',dec='.')
data
head(data)
dim(data)
(matematika<-data$math)
(english<-data$english)
(pohlavi<-data$sex)
data[data$sex==0,]
```

2 - Bodové a intervalové rozložení četností

Příklad č.1: Načtěte datový soubor znamky.txt.

```
#Zobrazeni prvnich sesti radku:
  math english sex
1     2       2   0
2     1       3   1
3     4       3   1
4     1       1   0
5     1       2   1
6     4       4   1
```

1. Vytvořte variační řadu (tabulku rozložení četností)

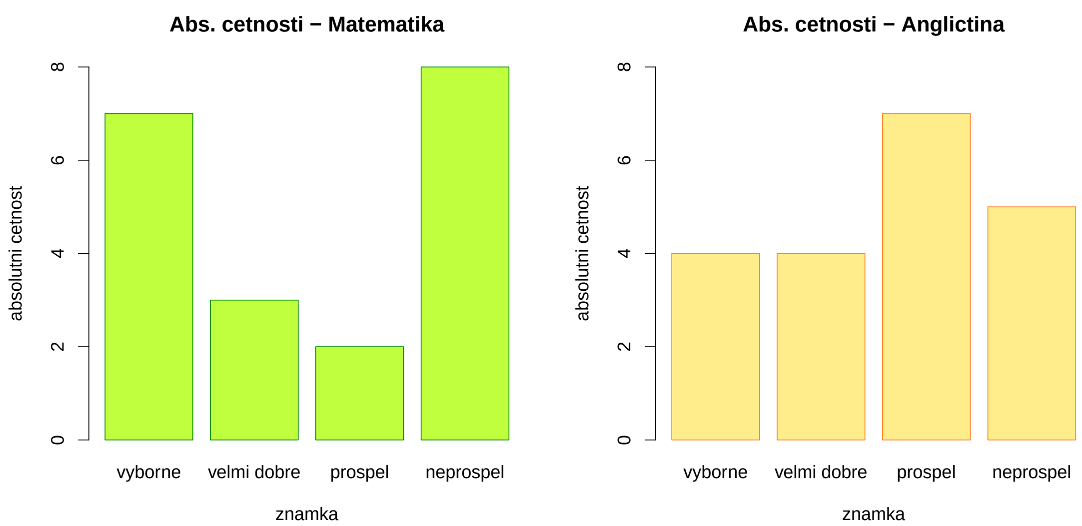
   (a) známek z matematiky (znak X);

```
             nj   pj Nj   Fj
Vyborne       7 0.35  7 0.35
Velmi_dobre   3 0.15 10 0.50
Prospel       2 0.10 12 0.60
Neprospel     8 0.40 20 1.00
```

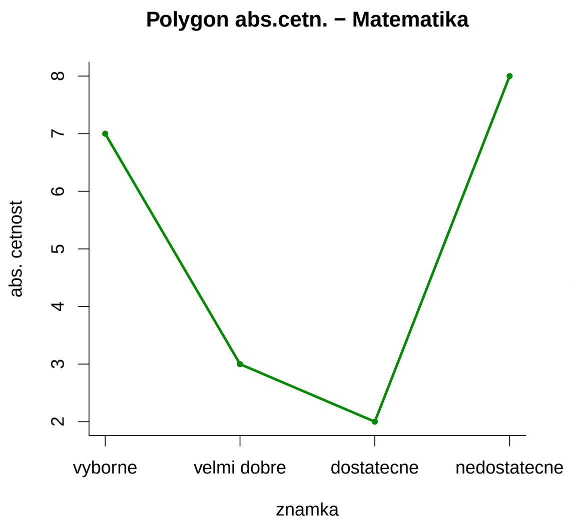
   (b) známek z angličtiny (znak Y).

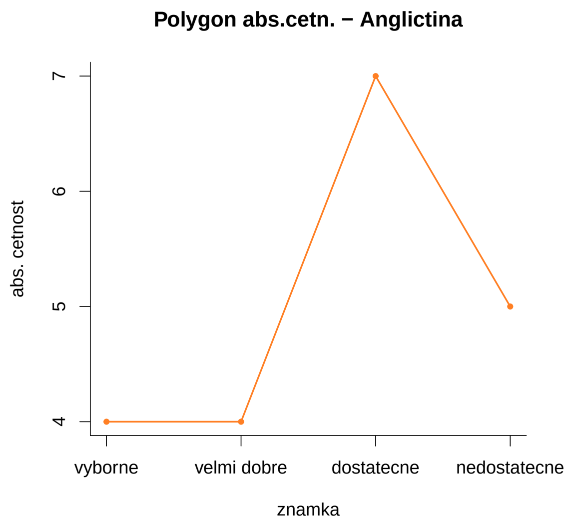
```
             nj   pj Nj   Fj
Vyborne       4 0.20  4 0.20
Velmi_dobre   4 0.20  8 0.40
Prospel       7 0.35 15 0.75
Neprospel     5 0.25 20 1.00
```

2. Vytvořte sloupkový diagram absolutních četností znaků X a Y.



3. Vytvořte polygon absolutních četností znaků X a Y.





4. Vytvořte variační řady (tabulky rozložení četností) známek z matematiky a angličtiny

   (a) pouze pro ženy;

```
#Variacni rada znamek z matematiky - zeny
            nj  pj Nj  Fj
Vyborne      5 0.5  5 0.5
Velmi_dobre  2 0.2  7 0.7
Prospel      1 0.1  8 0.8
Neprospel    2 0.2 10 1.0

#Variacni rada znamek z anglictiny - zeny
            nj  pj Nj  Fj
Vyborne      4 0.4  4 0.4
Velmi_dobre  2 0.2  6 0.6
Prospel      1 0.1  7 0.7
Neprospel    3 0.3 10 1.0
```

   (b) pouze pro muže.

```
#Variacni rada znamek z matematiky - muzi
            nj  pj Nj  Fj
Vyborne      2 0.2  2 0.2
Velmi_dobre  1 0.1  3 0.3
Prospel      1 0.1  4 0.4
Neprospel    6 0.6 10 1.0

#Variacni rada znamek z anglictiny - muzi
            nj  pj Nj  Fj
Vyborne      0 0.0  0 0.0
Velmi_dobre  2 0.2  2 0.2
Prospel      6 0.6  8 0.8
Neprospel    2 0.2 10 1.0
```

5. Vytvořte kontingenční tabulku simultánních absolutních četností znaků X a Y.

	E_Vyborne	E_Velmi.dobre	E_Prospel	E_Neprospel	E_Celkem
M_Vyborne	4	1	2	0	7
M_Velmi_dobre	0	2	1	0	3
M_Prospel	0	0	1	1	2
M_Neprospel	0	1	3	4	8
M_celkem	4	4	7	5	20

6. Vytvořte kontingenční tabulku

   (a) sloupcově podmíněných relativních četností znaků X a Y;

	E_Vyborne	E_Velmi.dobre	E_Prospel	E_Neprospel
M_Vyborne	1	0.25	0.29	0.0
M_Velmi_dobre	0	0.50	0.14	0.0
M_Prospel	0	0.00	0.14	0.2
M_Neprospel	0	0.25	0.43	0.8
Celkem	1	1.00	1.00	1.0

   (b) řádkově podmíněných relativních četností znaků X a Y.

	E_Vyborne	E_Velmi.dobre	E_Prospel	E_Neprospel	Celkem
M_Vyborne	0.57	0.14	0.29	0.0	1
M_Velmi_dobre	0.00	0.67	0.33	0.0	1
M_Prospel	0.00	0.00	0.50	0.5	1
M_Neprospel	0.00	0.12	0.38	0.5	1

   (c) Kolik procent studentů, kteří prospěli z angličtiny, neudělalo zkoušku z matematiky? (43%).
   Jaký je podíl studentů, kteří neudělali zkoušku z angličtiny a neprospěli ani z matematiky? (0.8). Kolik je to studentů? $(0.8 * 5 = 4)$
   Kolik procent studentů, kteří prospěli z matematiky, neudělalo zkoušku z angličtiny? 50%.
   Jaký je podíl studentů, kteří neudělali zkoušku z matematiky a neprospěli ani z angličtiny? (0.5) Kolik je to studentů? $(0.5 * 8 = 4)$

Příklad č.2: Načtěte soubor ocel.txt.

```
#prvnich sest pozorovani ze souboru ocel.txt
  mez_platicity mez_pevnosti
1           154          178
2           133          164
3            58           75
4           145          161
5            94          107
6           113          141
```

1. Podle Sturgersova pravidla najděte optimální počet třídicích intervalů pro znaky *plasticita* a *pevnost* a vhodně stanovte meze třídicích intervalů pro každý znak.

   ```
   #pocet tridicich intervalu
   [1] 7
   # rozsah platicity
   [1]  33 160
   # rozsah pevnosti
   [1]  52 189
   ```

   Dolní mez prvního třídicího intervalu pro platicitu zvolíme rovnu 30, horní mez posledního intervalu pro plasticitu zvolíme 170. Rozpětí mezi hodnotami 30 a 170 je 140. Po vydělení 7 dostaneme, že šíře jednoho intervalu bude rovná 20. Získáme tedy intervaly: $(30; 50\rangle$, $(50; 70\rangle$, $(70; 90\rangle$, $(90; 110\rangle$, $(110; 130\rangle$, $(130; 150\rangle$, $(150; 170\rangle$.

   Poznámka: Pro úplnost bychom měli ještě stanovit krajní intervaly $(-\infty; 30\rangle$ a $(170; \infty)$. Tyto intervaly ale neobsahují žádné pozorování.

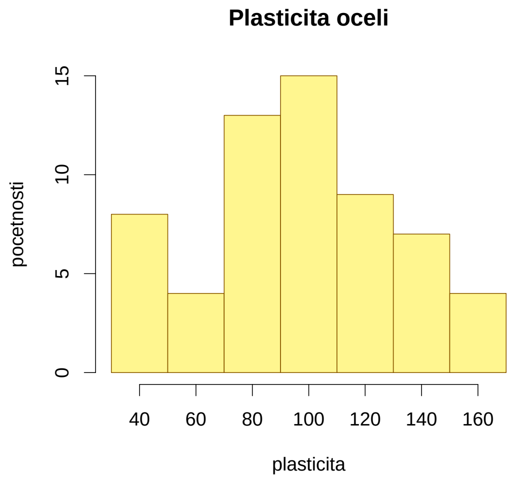
   Dolní mez prvního třídicího intervalu pro pevnost zvolíme rovnu 50, horní mez posledního intervalu pro plasticitu zvolíme 190. Rozpětí mezi hodnotami 50 a 190 je 140. Po vydělení 7 dostaneme, že šíře jednoho intervalu bude rovná 20. Získáme tedy intervaly: $(50; 70\rangle$, $(70; 90\rangle$, $(90; 110\rangle$, $(110; 130\rangle$, $(130; 150\rangle$, $(150; 170\rangle$, $(170; 190\rangle$.

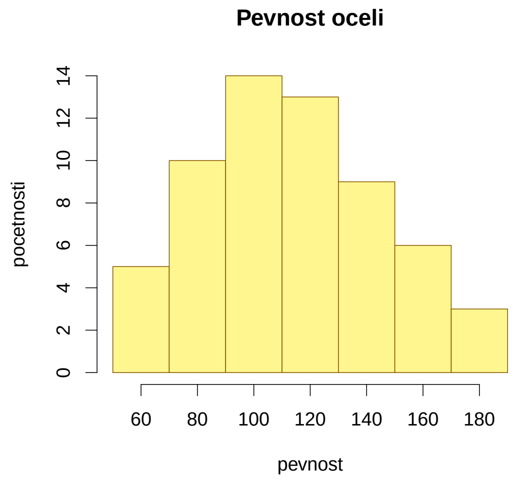
   Poznámka: Pro úplnost bychom měli ještě stanovit krajní intervaly $(-\infty; 50\rangle$ a $(190; \infty)$. Tyto intervaly ale neobsahují žádné pozorování.

   Dále určete středy těchto intervalů a příslušné variační řady.

   ```
   #Plasticita
      dh  hh stred nj   pj Nj   Fj
   1  30  50    40  8 0.13  8 0.13
   2  50  70    60  4 0.07 12 0.20
   3  70  90    80 13 0.22 25 0.42
   4  90 110   100 15 0.25 40 0.67
   5 110 130   120  9 0.15 49 0.82
   6 130 150   140  7 0.12 56 0.93
   7 150 170   160  4 0.07 60 1.00

   #Pevnost
      dh  hh stred nj   pj Nj   Fj
   1  50  70    60  5 0.08  5 0.08
   2  70  90    80 10 0.17 15 0.25
   3  90 110   100 14 0.23 29 0.48
   4 110 130   120 13 0.22 42 0.70
   5 130 150   140  9 0.15 51 0.85
   6 150 170   160  6 0.10 57 0.95
   7 170 190   180  3 0.05 60 1.00
   ```

2. Vytvořte histogram pro *plasticitu* a pro *pevnost*.





3. Sestavte kontingenční tabulky absolutních četností a relativních četností dvourozměrných třídicích intervalů pro dvojici znaků (*plasticita*, *pevnost*).

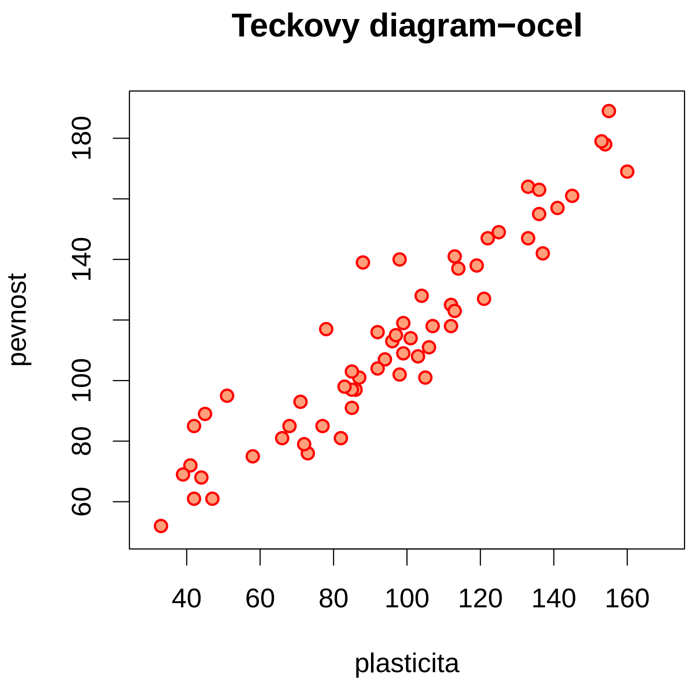
#Kontingencni tabulka absolutnich cetnosti

	pev.I	pev.II	pev.III	pev.IV	pev.V	pev.VI	pev.VII	Celkem
pl.I	5	3	0	0	0	0	0	8
pl.II	0	3	1	0	0	0	0	4
pl.III	0	4	7	1	1	0	0	13
pl.IV	0	0	6	8	1	0	0	15
pl.V	0	0	0	4	5	0	0	9
pl.VI	0	0	0	0	2	5	0	7
pl.VII	0	0	0	0	0	1	3	4
Celkem	5	10	14	13	9	6	3	60

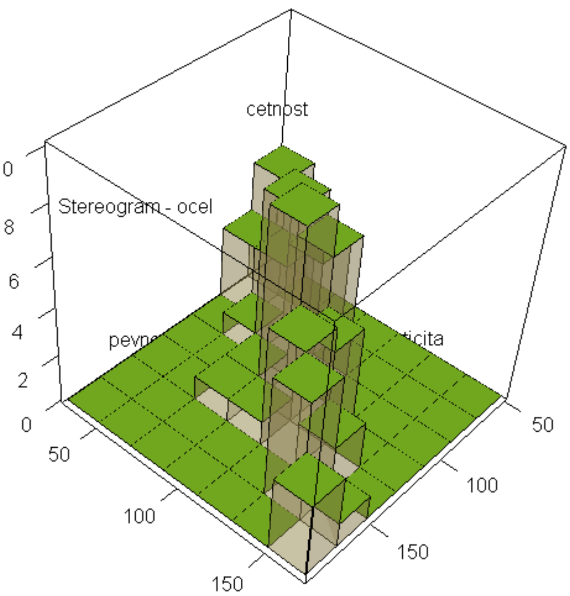
#Kontingencni tabulka relativnich cetnosti

	pev.I	pev.II	pev.III	pev.IV	pev.V	pev.VI	pev.VII	Celkem
pl.I	0.08	0.05	0.00	0.00	0.00	0.00	0.00	0.13
pl.II	0.00	0.05	0.02	0.00	0.00	0.00	0.00	0.07
pl.III	0.00	0.07	0.12	0.02	0.02	0.00	0.00	0.22
pl.IV	0.00	0.00	0.10	0.13	0.02	0.00	0.00	0.25
pl.V	0.00	0.00	0.00	0.07	0.08	0.00	0.00	0.15
pl.VI	0.00	0.00	0.00	0.00	0.03	0.08	0.00	0.12
pl.VII	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.07
Celkem	0.08	0.17	0.23	0.22	0.15	0.10	0.05	1.00

4. Nakreslete dvourozměrný tečkový diagram pro (*plasticita*, *pevnost*).



5. Dobrovolný úkol: Vytvořte stereogram pro (*plasticita*, *pevnost*).



3 - Výpočet číselných charakteristik jednorozměrného a dvourozměrného datového souboru

Příklad č.1 U 100 náhodně vybraných osob jsme zjišťovali barvu jejich vlasů (znak X, varianty 1=blond, 2=černé, 3=hnědá) a barvu jejich očí. (znak Y, varianty 1 = hnědá, 2 = zelená, 3 = modrá).

	hnědá	zelená	modrá
blond	13	15	14
černá	11	7	2
hnědá	19	9	10

(a) Pro oba znaky určete modus.

```
              hneda zelena  modra celkem
blond            13     15     14     42
cerna            11      7      2     20
hneda            19      9     10     38
celkem           43     31     26    100

#Modus pro barvu oci:
 43
#Nejcastejsi variantou barvy oci je hneda (celkem 43 pripadu).

#Modus pro barvu vlasu:
 42
#Nejcastejsi variantou barvy vlasu je blond (celkem 42 pripadu).
```

(b) Určete, zda mezi znaky *vlasy* a *oci* existuje nějaká závislost (Pokud ano, jaká?). (Nápověda: Protože oba znaky jsou nominálního typu, použijeme na zhodnocení závislosti **Cramérův koeficient**.)

Pro připomenutí zde uvádíme tabulku stupňů lineární závislosti pro Cramérův koeficient:

Cramérův koeficient	interpretace
0 – 0,1	zanedbatelná závislost
0.1 – 0.3	slabá závislost
0.3 – 0.7	střední závislost
0.7 – 1	silná závislost

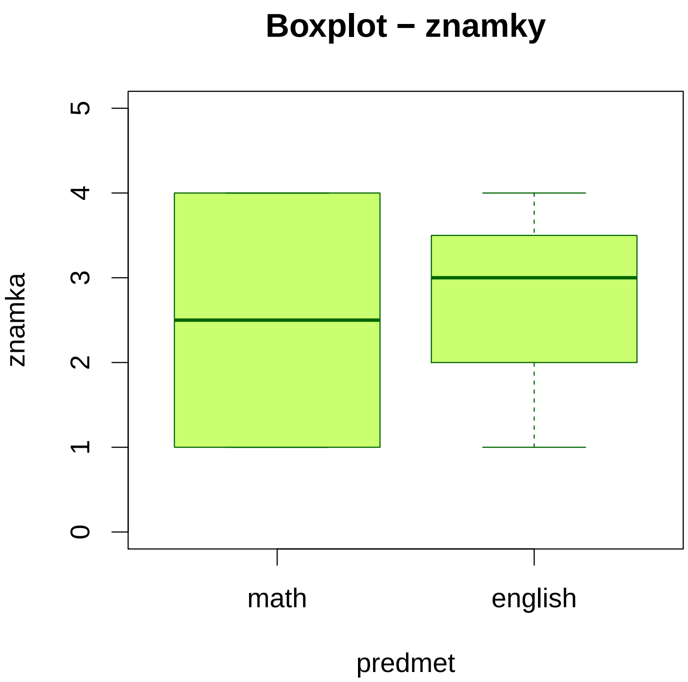
```
#Crameruv koeficient
0.1791687
#Interpretace: Mezi znaky barva oci a barva vlasu existuje pouze slaba
   linearni zavislost.
```

Příklad č.2 Otevřete datový soubor znamky.txt.

```
# Prvnich sest radku datoveho souboru
  math english sex
1    2       2   0
2    1       3   1
3    4       3   1
4    1       1   0
5    1       2   1
6    4       4   1
```

(a) Pro známky z matematiky a angličtiny vypočtěte medián, dolní a horní kvartil, kvartilovou odchylku a vytvořte krabicový diagram.

```
#Charakteristiky pro znamky z matematiky:
[1] "median=2.5"
[1] "kv1=1"
[1] "kv3=4"
[1] "rozpeti=3"

#Charakteristiky pro znamky z anglictiny:
[1] "median=3"
[1] "kv1=2"
[1] "kv3=3.25"
[1] "rozpeti=1.25"
```



(b) Určete vzájemnou závislot *známek z matematiky* a *známek z angličtiny* pro všechny studenty, pak zvlášť pro muže a zvlášť pro ženy. Získané výsledky interpretujte. (Nápověda: Protože oba znaky jsou ordinálního charakteru, použijeme na zhodnocení závislosti **Spearmanův korelační koeficient**.)

Pro připomenutí zde uvádíme tabulku stupňů pořadové závislosti pro Spearmanův korelační koeficient:

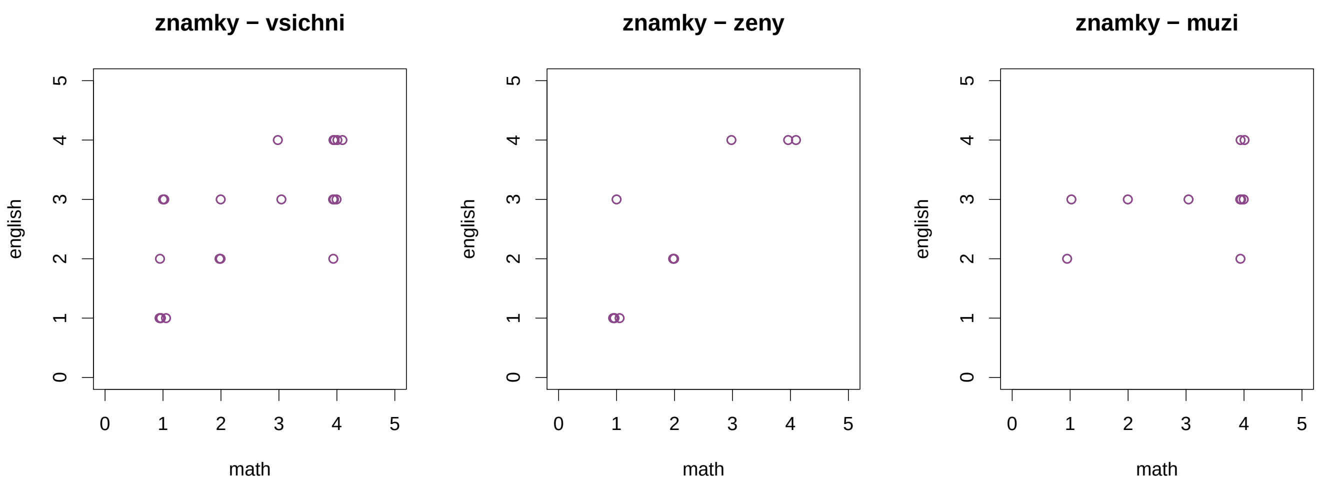
Abs.hod. korel.koef.	Interpretace hodnoty
0	pořadová nezávislost
$(0; 0.1)$	velmi nízký stupeň závislosti
$[0.1; 0.3)$	nízký stupeň závislosti
$[0.30; 0.50)$	mírný stupeň závislosti
$[0.50; 0.70)$	význačný stupeň závislosti
$[0.70; 0.90)$	vysoký stupeň závislosti
$[0.90; 1)$	velmi vysoký stupeň závislosti
1	úplná pořadová závislost

(c) Svůj závěr o (ne)závislost znaků *známka z matematiky* a *známka z angličtiny* doložte tečkovými diagramy.

```
#Korelacni koeficient - vsichni studenti
0.6884422
#Interpretace: Mezi hodnotou znamek z matematiky a anglictiny existuje
   vyznacny stupen poradove zavislosti

#Korelacni koeficient - zeny
0.8603138
#Interpretace: Mezi hodnotou znamek z matematiky a anglictiny pro zeny
   existuje vysoky stupen poradove zavislosti

#Korelacni koeficient - muzi
0.3735437
#Interpretace: Mezi hodnotou znamek z matematiky a anglictiny pro muze
   existuje mirny stupen poradove zavislosti
```

Vidíme, že nejsilnější přímá pořadová závislost mezi známkami z matematiky a angličtiny je u žen, $korel.koef = 0.86$. U mužů je tato závislost mnohem slabší, $korel.koef = 0.37$. U žen tedy dochází k tomu, že se sdružují podobné známky z obou předmětů, zatímco u mužů se projevuje spíše tendence k různým známkám.

(d) Svůj závěr o (ne)závislost znaků *známka z matematiky* a *známka z angličtiny* doložte tečkovými diagramy.



Příklad č.3 Otevřete datový soubor ocel.txt.

```
  mez_plasticity mez_pevnosti
1            154          178
2            133          164
3             58           75
4            145          161
5             94          107
6            113          141
```

(a) Pro *mez plasticity* a *mez pevnosti* vypočtěte aritmetický průměr, směrodatnou odchylku, rozptyl, koeficient variace, šikmost a špičatost.

```
#Tabulka zakladnich charakteristik pro mez plasticity a mez pevnosti
             prumer     sd  rozptyl koef.var sikmost spicatost
plasticita   95.883 32.715 1070.240    0.341  -0.044    -0.732
pevnost     114.400 32.789 1075.125    0.287   0.283    -0.721
```

(b) Vypočtěte Pearsonův koeficient korelace *meze plasticity* a *meze pevnosti*. Dále vypočtěte také kovarianci a kovarianční matici.

Pro připomenutí zde uvádíme tabulku stupňů lineární závislosti pro Pearsonův korelační koeficient:

Abs.hod. korel.koef.	Interpretace hodnoty
0	lineární nezávislost
$(0; 0.1)$	velmi nízký stupeň závislosti
$[0.1; 0.3)$	nízký stupeň závislosti
$[0.30; 0.50)$	mírný stupeň závislosti
$[0.50; 0.70)$	význačný stupeň závislosti
$[0.70; 0.90)$	vysoký stupeň závislosti
$[0.90; 1)$	velmi vysoký stupeň závislosti
1	úplná lineární závislost

```
#Pearsonuv koeficient korelace pro mez plasticity a mez pevnosti
0.9345481

#Kovariance meze plasticity a meze pevnosti
1002.471

#Kovariancni matice meze plasticity a meze pevnosti
           plasticita  pevnost
plasticita   1070.240 1002.471
pevnost      1002.471 1075.125
```

Vysvětlení kovarianční matice: Na hlavní diagonále jsou rozptyly proměnných X, Y, mimo hlavní diagonálu je kovariance.

Příklad č.4 Je třeba si uvědomit, že průměr a rozptyl nepopisují rozložení četností jednoznačně. Existují datové soubory, které mají shodný průměr i rozptyl, ale přesto se jejich rozložení četností velmi liší. Tuto skutečnost dobře ilustruje následující příklad: Tři skupiny studentů o počtech 149, 69 a 11 odpovídaly při testu na 10 otázek. Znak X je počet správně zodpovězených otázek. Známe absolutní četnosti znaku X ve všech třech skupinách. *Poznámka:*

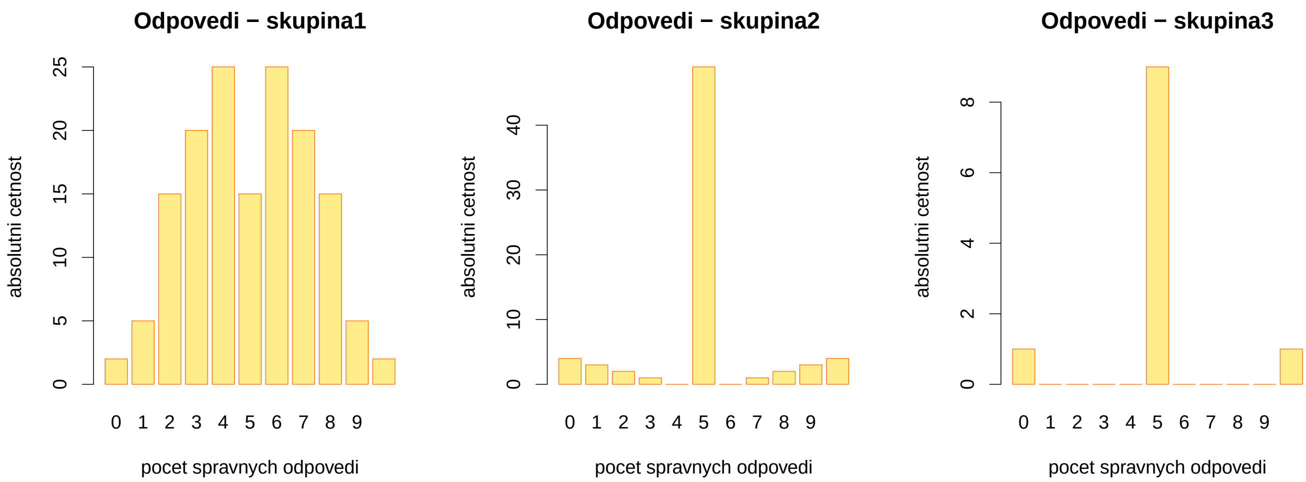
c.sk / X	0	1	2	3	4	5	6	7	8	9	10
1	2	5	15	20	25	15	25	20	15	5	2
2	4	3	2	1	0	49	0	1	2	3	4
3	1	0	0	0	0	9	0	0	0	0	1

Data k tomuto příkladu lze nalézt v souboru odpovedi.txt.

```
  X0 X1 X2 X3 X4 X5 X6 X7 X8 X9 X10
1  2  5 15 20 25 15 25 20 15  5   2
2  4  3  2  1  0 49  0  1  2  3   4
3  1  0  0  0  0  9  0  0  0  0   1
```

Vypočtěte průměr, rozptyl, šikmost a špičatost počtu správně zodpovězených otázek ve všech třech skupinách. Nakreslete sloupkové diagramy absolutních četností.

```
#Charakteristiky pro skupinu c.1
  prumer rozptyl sikmost spicatost
1      5       5       0    -0.804

#Charakteristiky pro skupinu c.2
  prumer rozptyl sikmost spicatost
1      5       5       0    0.9954

#Charakteristiky pro skupinu c.3
  prumer rozptyl sikmost spicatost
1      5       5       0    1.5455
```



Všechny tři skupiny mají týž průměr, rozptyl a šikmost, liší se pouze ve špičatosti. Sloupkové diagramy počtu správně zodpovězených otázek v každé ze tří uvažovaných skupin mají naprosto odlišný vzhled.

4 - Využití systému R při řešení příkladů na opakované pokusy

Vyřešte následující příklady. Ke každému příkladu zobrazte tvar příslušné distribuční funkce a hustoty.

Binomické rozložení pravděpodobností:

Příklad č.1: Pojišťovna zjistila, že 12 % pojistných událostí je způsobeno vloupáním. Jaká je pravděpodobnost, že mezi 30 náhodně vybranými pojistnými událostmi bude způsobeno vloupáním

(a) nejvýše 6;

(b) alespoň 6;

(c) právě 6;

(d) od dvou do pěti?

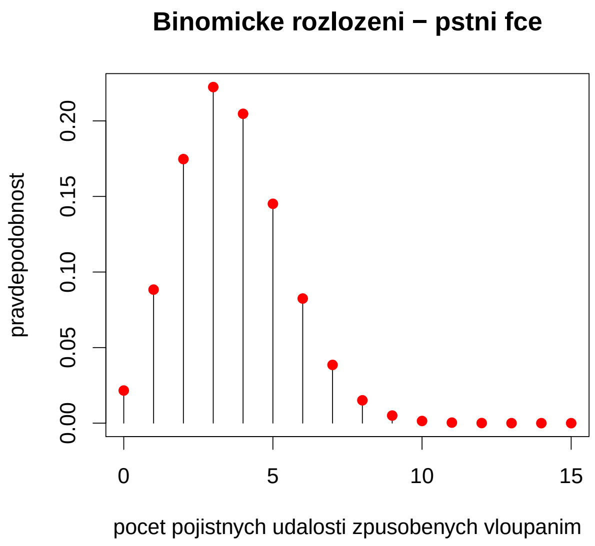
```
# a)
[1] 0.9393926
# b)
[1] 0.1430769
# c)
[1] 0.08246953
# d)
[1] 0.7469528
```

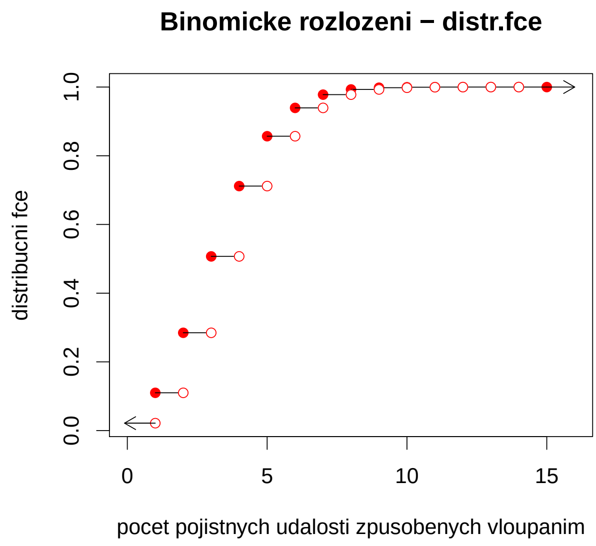
(a) Pravděpodobnost, že mezi 30 náhodně vybranými pojistnými událostmi bude nejvýše 6 událostí způsobeno vloupáním, je 0.939.

(b) Pravděpodobnost, že mezi 30 náhodně vybranými pojistnými událostmi bude alespoň 6 událostí způsobeno vloupáním, je 0.143.

(c) Pravděpodobnost, že mezi 30 náhodně vybranými pojistnými událostmi bude právě 6 událostí způsobeno vloupáním, je 0.082.

(d) Pravděpodobnost, že mezi 30 náhodně vybranými pojistnými událostmi bude 2–5 událostí způsobeno vloupáním, je 0.747.





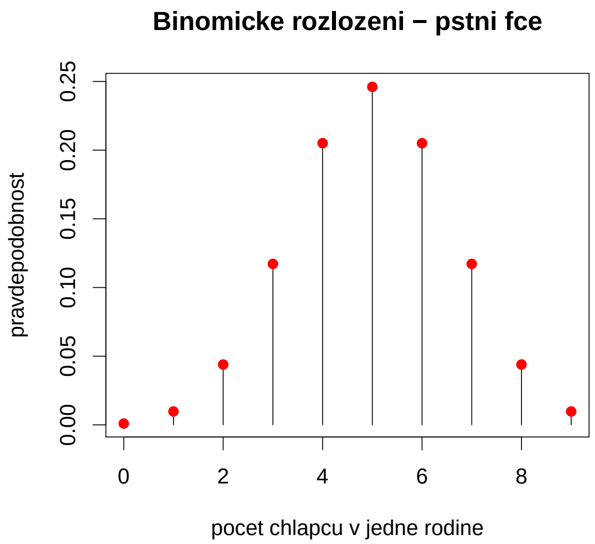
Příklad č.2: V rodině je 10 dětí. Za předpokladu, že chlapci i dívky se rodí s pravděpodobností 0.5 a pohlaví se formuje nezávisle na sobě, určete pravděpodobnost, že v této rodině je

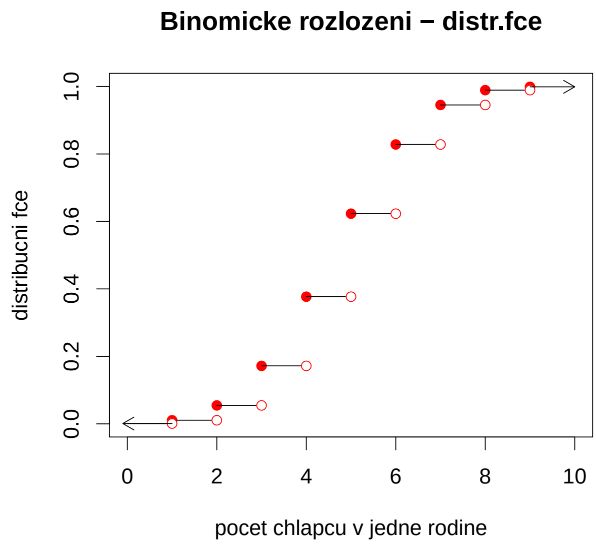
(a) právě 5 chlapců;

(b) nejméně 3 a nejvýše 8 chlapců.

```
# a)
[1] 0.2460938
# b)
[1] 0.9345703
```

Pravděpodobnost, že v rodině bude právě 5 chlapců, je 0.246. Pravděpodobnost, že v rodině bude 3–8 chlapců, je 0.935.





Příklad č.3: Na dvoukolejném železničním mostě se potkají během 24 hodin nejvýše dva vlaky, a to s pravděpodobností 0.2. Za předpokladu, že denní provozy jsou nezávislé, určete pravděpodobnost, že během týdne se dva vlaky na mostě potkají

(a) právě třikrát;

(b) nejvýše třikrát;

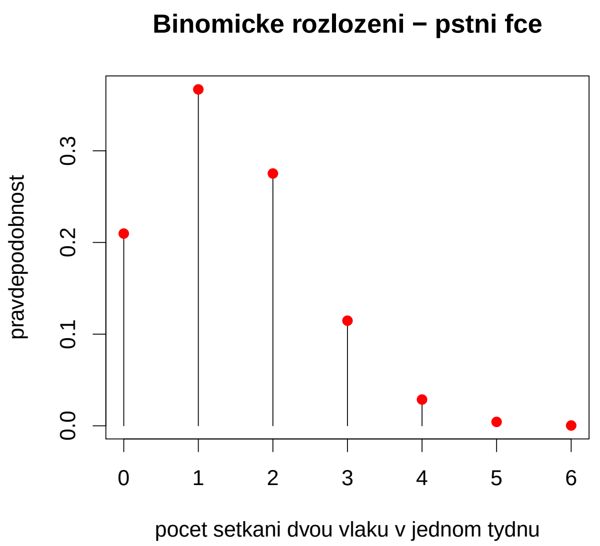
(c) alespoň třikrát.

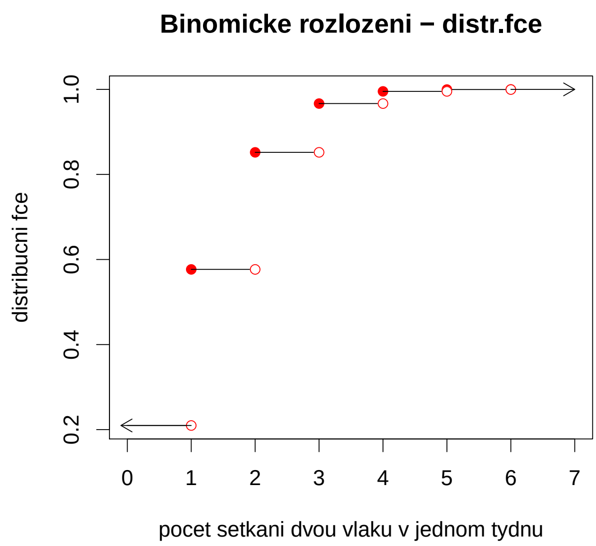
```
# a)
[1] 0.114688
# b)
[1] 0.966656
# c)
[1] 0.148032
```

(a) Pravděpodobnost, že se dva vlaky potkají na mostě právě třikrát za týden, je 0.115.

(b) Pravděpodobnost, že se dva vlaky potkají na mostě nejvýše třikrát za týden, je 0.967.

(c) Pravděpodobnost, že se dva vlaky potkají na mostě alespoň třikrát za týden, je 0.148.

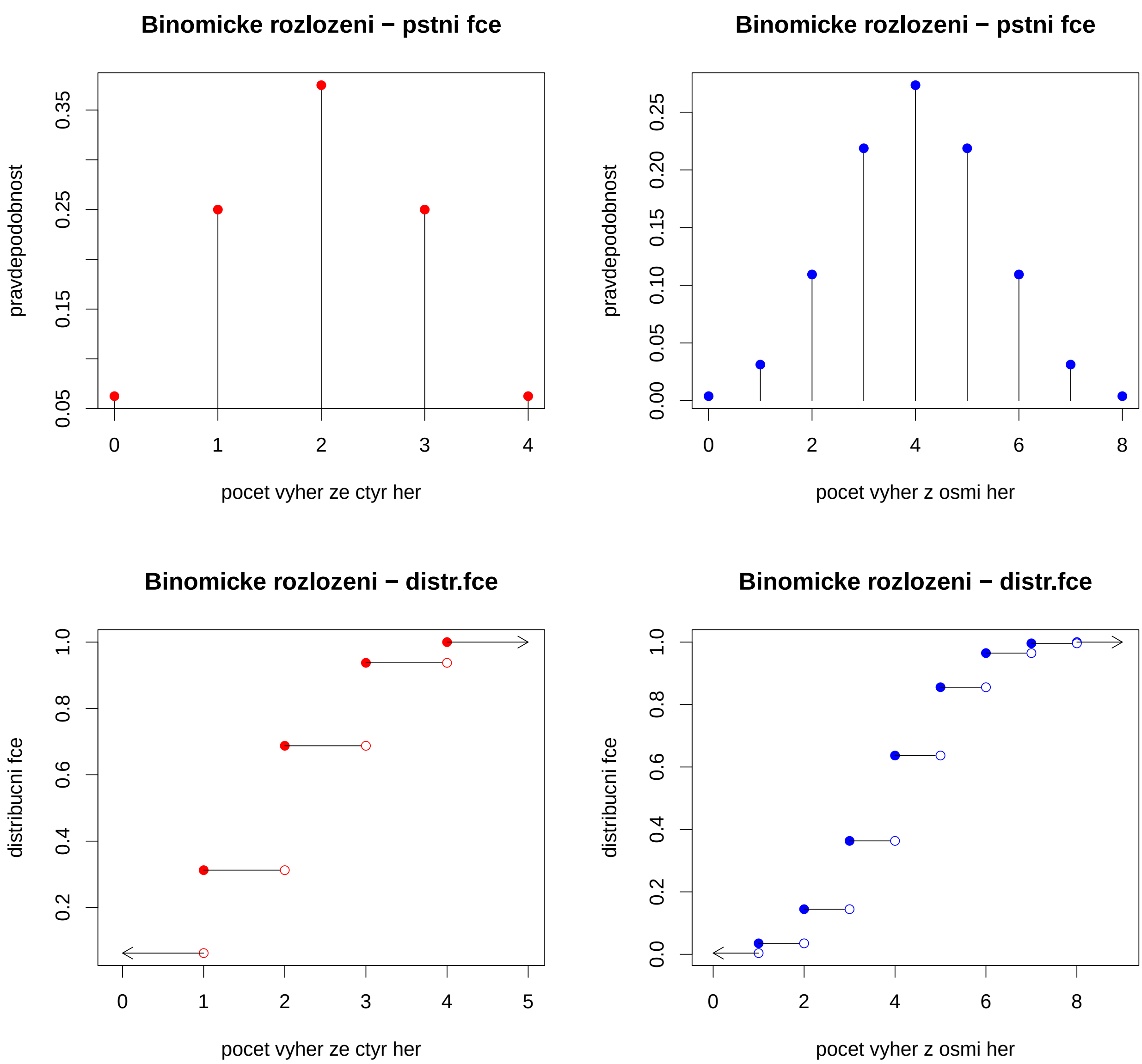




Příklad č.4: Je pravděpodobnější vyhrát se stejně silným soupeřem tři partie ze čtyř nebo pět partií z osmi, když nerozhodný výsledek je vyloučen a výsledky jsou nezávislé? Úspěch je výhra partie se stejně silným soupeřem, když remíza je vyloučena, pravděpodobnost úspěchu $\vartheta = 0.5$.

```
# pst, ze vyhraji 3 partie ze 4
[1] 0.9375
# pst, ze vyhraji 5 partii z 8
[1] 0.8554688
```

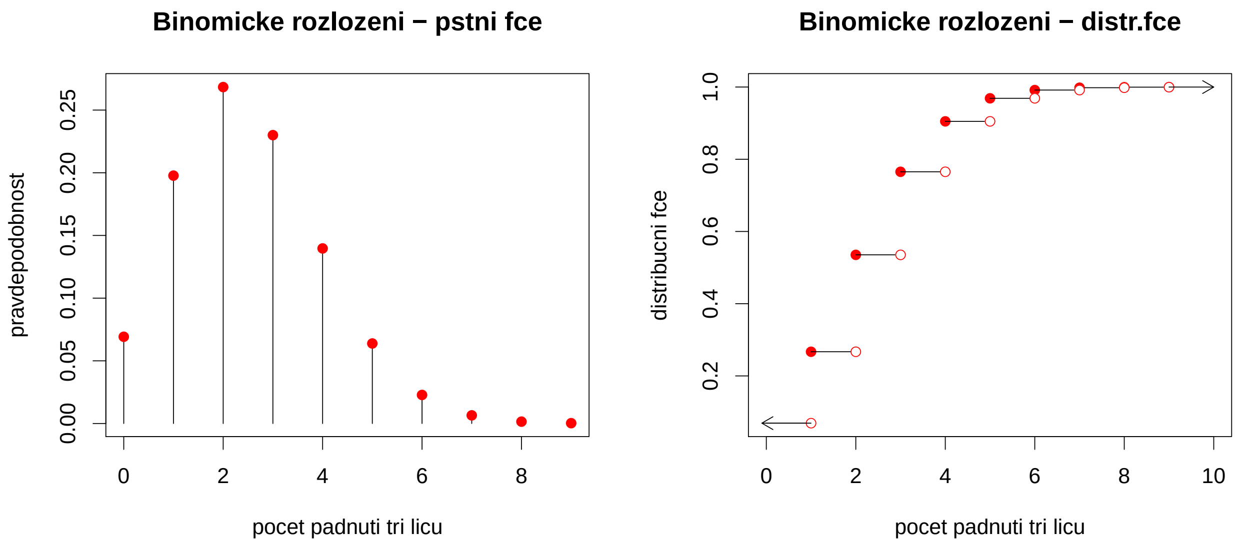
Pravděpodobnější je, že vyhraji tři partie ze čtyř, než pět partií z osmi.



Příklad č.5: Dvacetkrát nezávisle na sobě házíme třemi mincemi. Jaká je pravděpodobnost, že alespoň v jednom hodě padnou tři líce?

```
[1] 0.9307912
```

Pravděpodobnost, že v alespoň jednom hodu padnou tři líce, je 0.931.

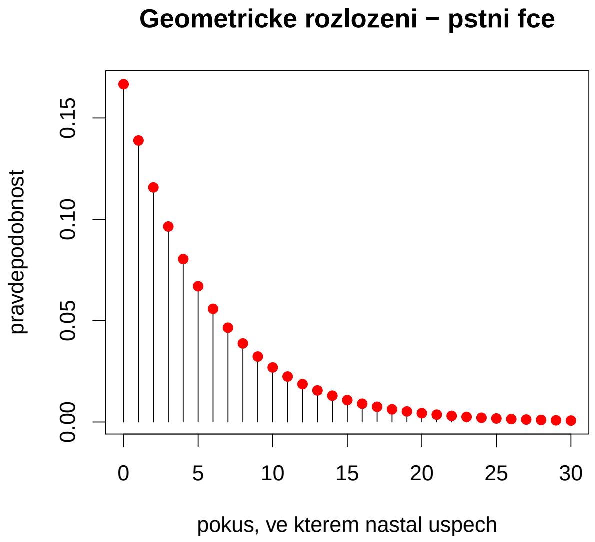

Binomicke rozlozeni − pstni fce
pravdepodobnost
0.00
0.05
0.10
0.15
0.20
0.25
0
2
4
6
8
pocet padnuti tri licu
Binomicke rozlozeni − distr.fce
distribucni fce
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
pocet padnuti tri licu

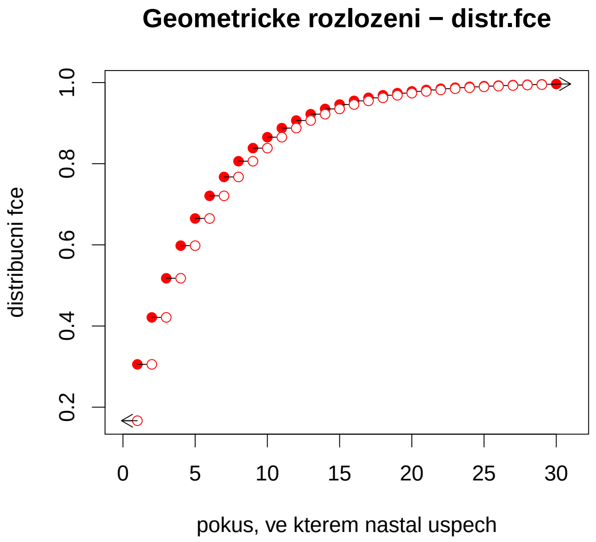
Geometrické rozložení pravděpodobností:

Příklad č.6: Jaká je pravděpodobnost, že při hře „Člověče, nezlob se!“ nasadíme figurku nejpozději při třetím hodu?

```
[1] 0.4212963
```

Pravděpodobnost, že nasadím figurku nejpozději při třetím hodu, je 0.421.

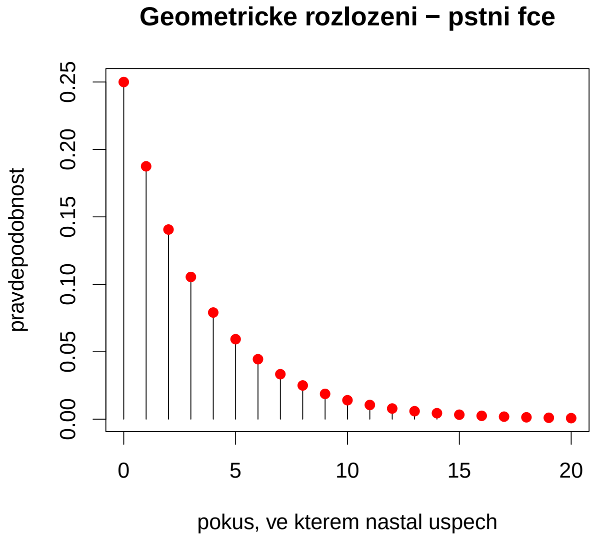


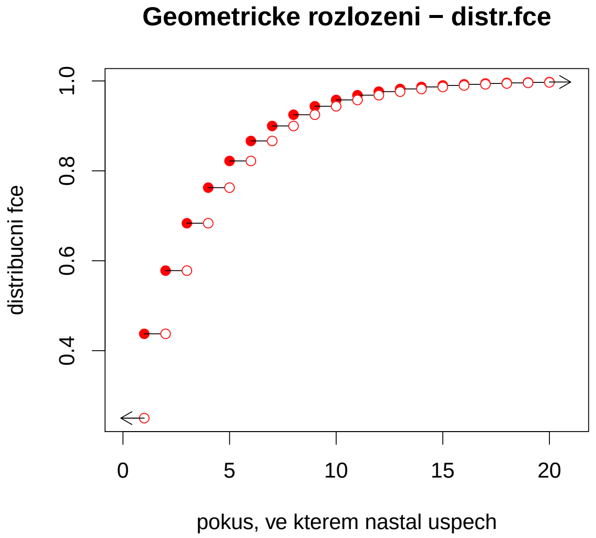


Příklad č.7: Studenti biologie zkoumají barvu očí octomilek. Pravděpodobnost, že octomilka má bílou barvu očí, je 0.25, červenou 0.75. Jaká je pravděpodobnost, že až čtvrtá zkoumaná octomilka má bílou barvu očí?

```
[1] 0.1054688
```

Pravděpodobnost, že až čtvrtá zkoumaná octomilka má bílou barvu očí, je 0.105.





Hypergeometrické rozložení pravděpodobností:

Příklad č.8: Koupili jsme 10 cibulek červených tulipánů a 5 cibulek žlutých tulipánů. Zasadili jsme 8 náhodně vybraných cibulek.

(a) Jaká je pravděpodobnost, že žádná cibulka nebude cibulka žlutých tulipánů?

(b) Jaká je pravděpodobnost, že jsme zasadili všech 5 cibulek žlutých tulipánů?

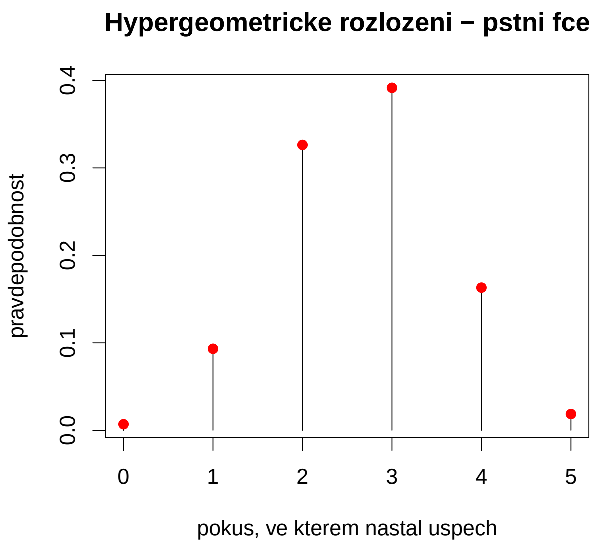
(c) Jaká je pravděpodobnost, že aspoň dvě cibulky budou cibulky žlutých tulipánů?

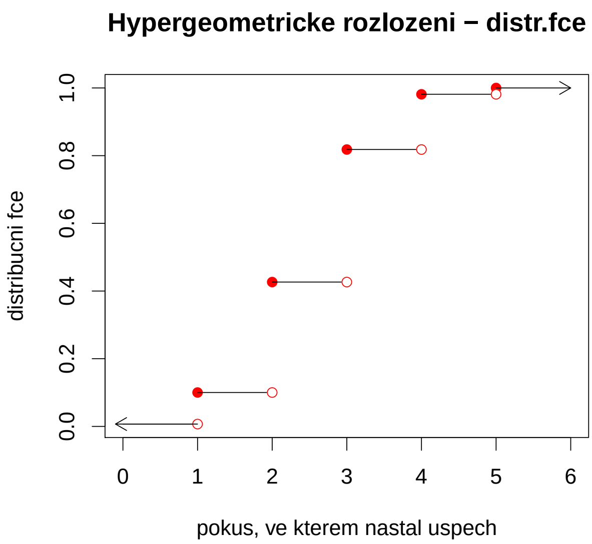
```
# a)
[1] 0.006993007
# b)
[1] 0.01864802
# c)
[1] 0.8997669
```

(a) Pravděpodobnost, že mezi vybranými cibulkami nebude žádná cibulka žlutých tulipánů, je 0.007.

(b) Pravděpodobnost, že mezi vybranými cibulkami bude všech 5 cibulek žlutých tulipánů, je 0.019.

(c) Pravděpodobnost, že mezi vybranými cibulkami budou alespoň dvě cibulky žlutých tulipánů, je 0.900.



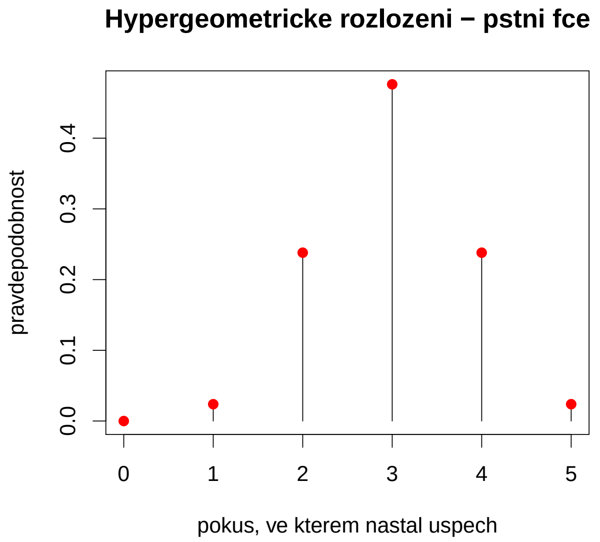


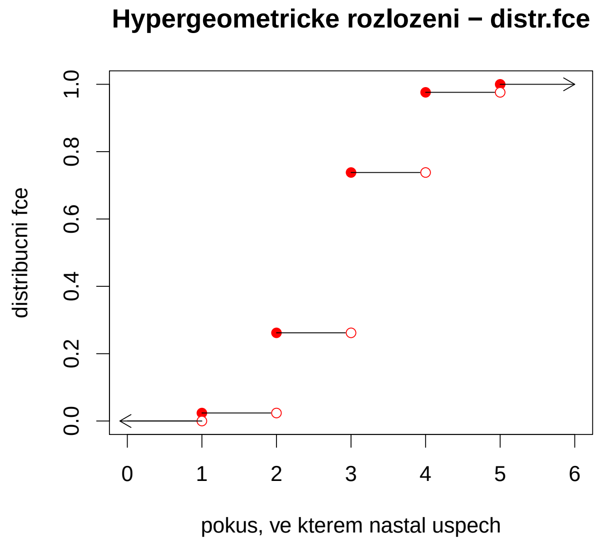
Příklad č.9: Dítě dostalo sáček, v němž bylo 5 červených a 5 žlutých bonbónů. Dítě náhodně vybralo ze sáčku 6 bonbónů. Jaká je pravděpodobnost, že mezi vybranými bonbóny budou právě 2 červené?

```
[1] 0.2380952
```

Pravděpodobnost, že mezi vybranými bonbóny budou právě 2 červené, je 0.238.

Výsledek testu	skutečnost		Celkem
	H (pozitivní)	H (negativní)	
A (pozitivní)	a=50	b=300	350
A (negativní)	c=25	d=870	895
celkem	75	1170	1245





Diagnostické testy - Nepovinné

Příklad č.10: Provádělo se ověřování kvality nového testu pro diagnostikování jisté poruchy slu-chu, která se vyskytuje u 12 % osob v populaci. Test byl ověřován u 1245 osob, u nichž byl stav sluchu vyšetřen již dříve podrobnými klinickými postupy. Výsledky máme v tabulce:

Vypočtěte prediktivní validitu pozitivního i negativního testu.

5 - Pravděpodobnostní funkce, hustoty a distribuční funkce v systému R, výpočet pravděpodobností pomocí distribučních funkcí

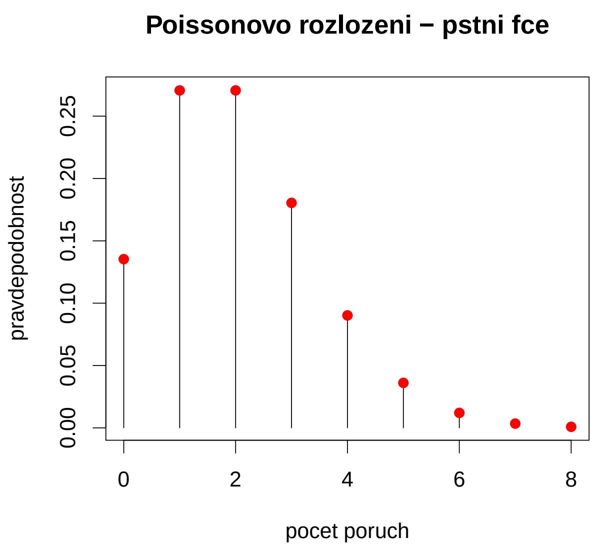
Vyřešte následující příklady. Ke každému příkladu zobrazte tvar příslušné distribuční funkce a hustoty.

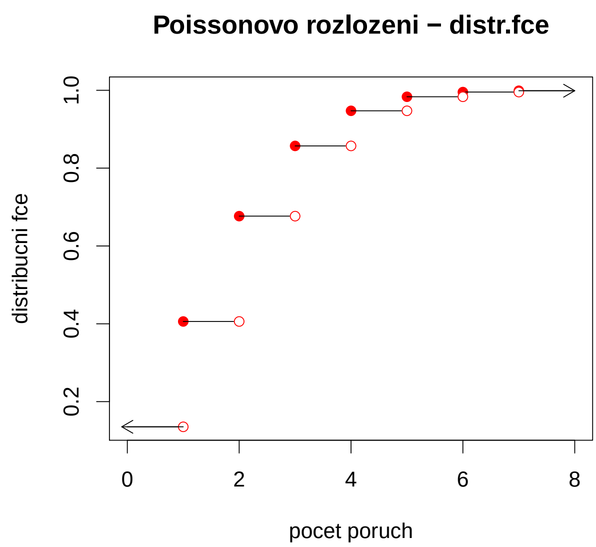
Poissonovo rozložení

Příklad č.1: Při provozu balicího automatu vznikají během směny náhodné poruchy, které se řídí rozložením Po(2). Jaká je pravděpodobnost, že během směny dojde k aspoň jedné poruše?

```
[1] 0.8646647
```

Pravděpodobnost, že během směny dojde k alespoň jedné poruše je 0.86.



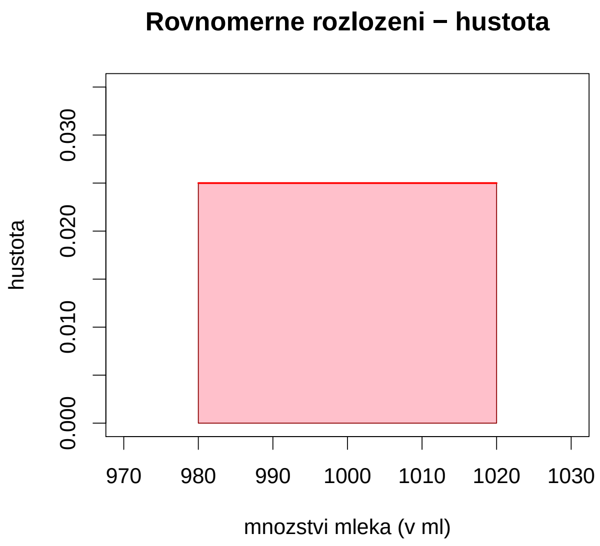


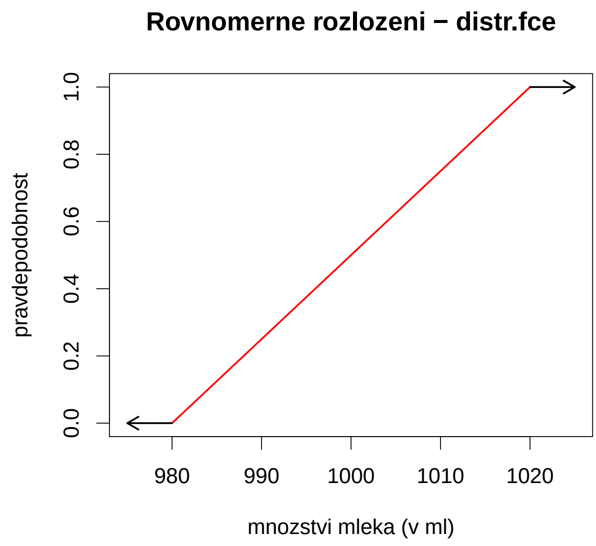
Rovnoměrné rozložení

Příklad č.2: Na automatické lince se plní lahve mlékem. Působením náhodných vlivů množství mléka kolísá v intervalu (980 ml, 1020 ml). Každé množství mléka v tomto intervalu považujeme za stejně možné. Jaká je pravděpodobnost, že v náhodně vybrané lahvi bude aspoň 1010 ml mléka?

```
[1] 0.25
```

Pravděpodobnost, že v náhodné lahvi bude alespoň 1010 ml mléka je 0.25.



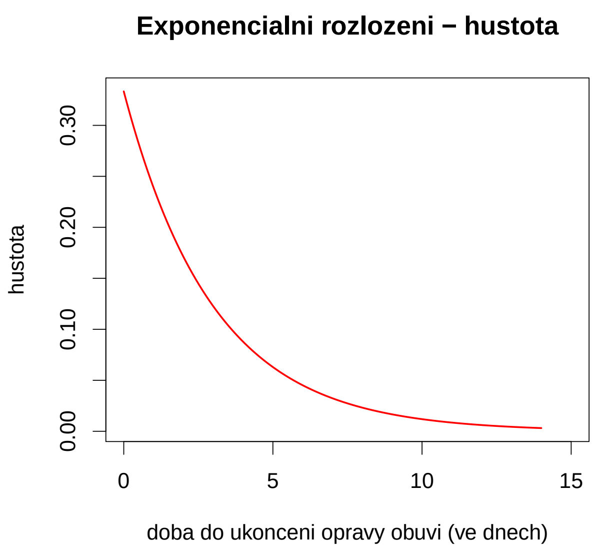


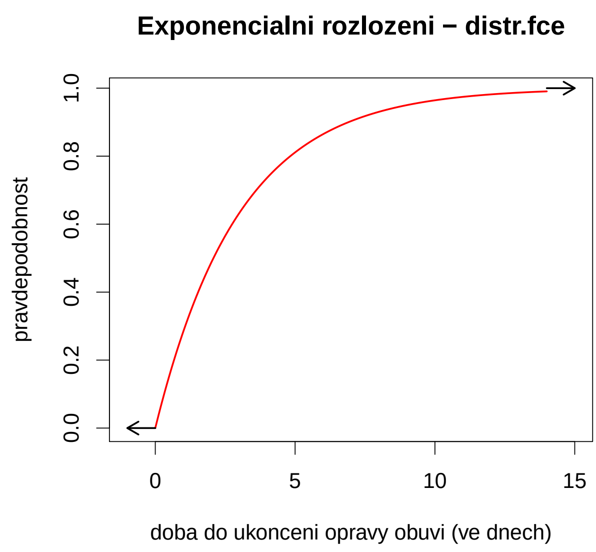
Exponenciální rozložení

Příklad č.3: Doba do ukončení opravy v opravně obuvi je náhodná veličina, která se řídí exponenciálním rozložením se střední dobou opravy 3 dny. Jaká je pravděpodobnost, že oprava bude ukončena do dvou dnů?

```
[1] 0.4865829
```

Pravděodobnost, že oprava obuvi bude dokončena do dvou dnů je 0.49.

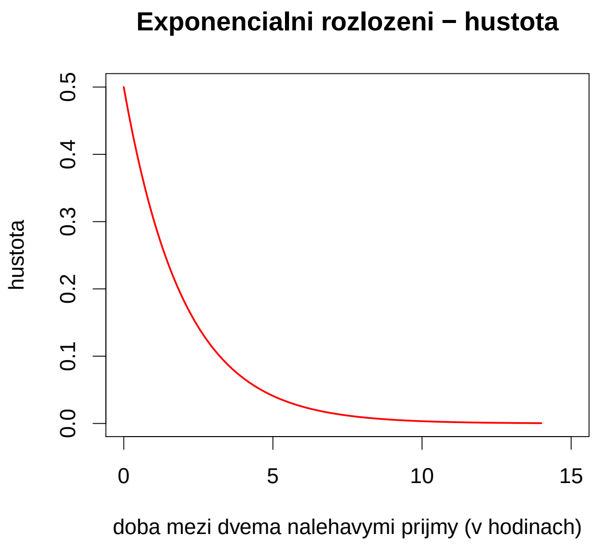


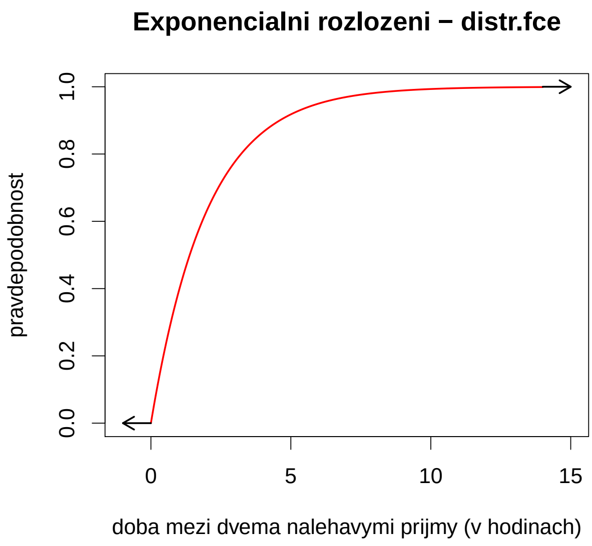


Příklad č.4: Doba (v hodinách), která uplyne mezi dvěma naléhavými příjmy v jisté nemocnici, se řídí exponenciálním rozložením se střední dobou čekání 2 h. Jaká je pravděpodobnost, že uplyne více než 5 h bez naléhavého příjmu?

```
[1] 0.082085
```

Pravděpodobnost, že uplyně více než 5 h bez naléhavého příjmu je 0.08.



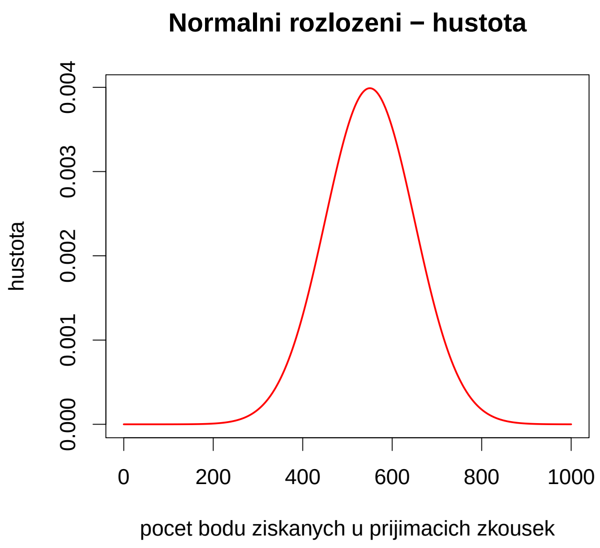


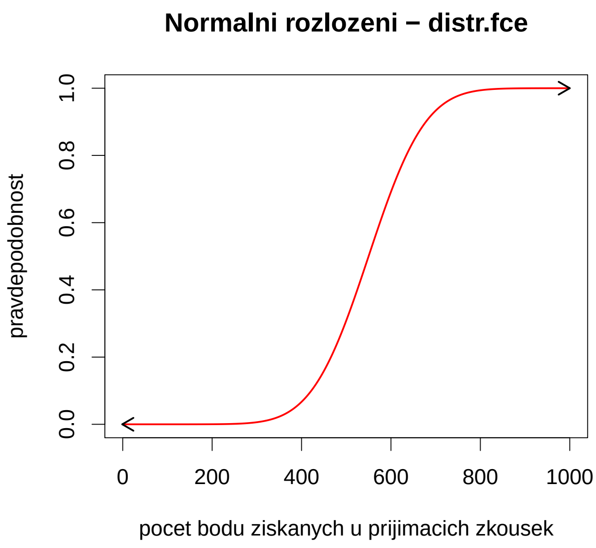
Normální rozložení

Příklad č.5: Výsledky u přijímacích zkoušek na jistou VŠ jsou normálně rozloženy s parametry $\mu = 550$ bodů, $\sigma = 100$ bodů. S jakou pravděpodobností bude mít náhodně vybraný uchazeč aspoň 600 bodů?

```
[1] 0.3085375
```

Pravěpodobnost, že vybraný uchazeč bude mít alespoň 600 bodů je 0.31.





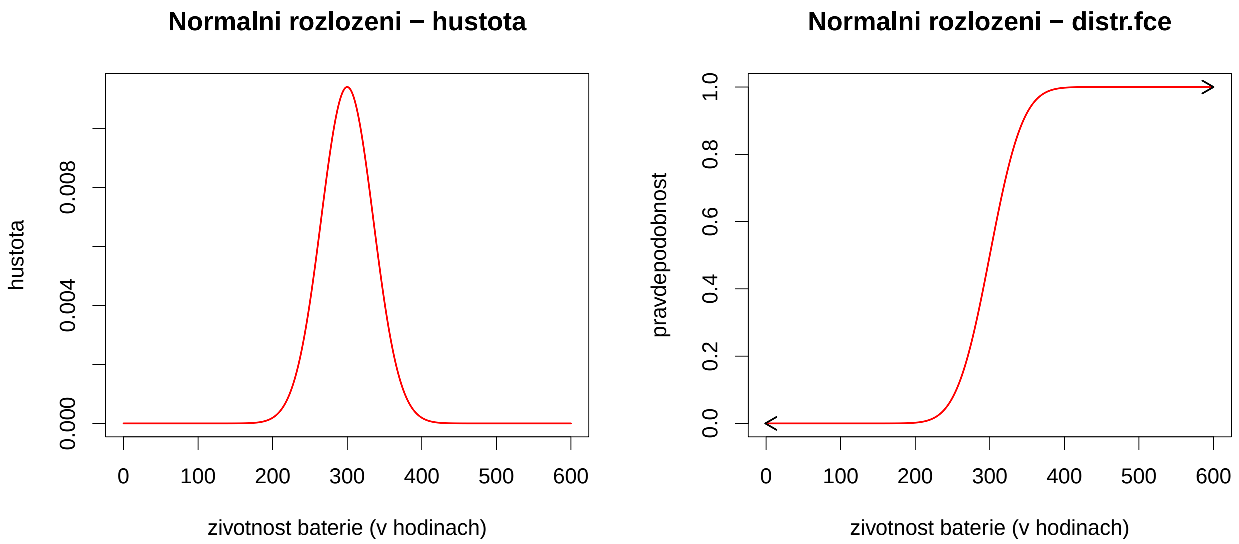
Příklad č.6: : Životnost baterie v hodinách je náhodná veličina, která má normální rozložení se střední hodnotou 300 hodin a směrodatnou odchylkou 35 hodin. Jaká je pravděpodobnost, že náhodně vybraná baterie bude mít životnost

(a) aspoň 320 hodin?

(b) nejvýše 310 hodin?

```
# a)
[1] 0.2838546
# b)
[1] 0.6124515
```

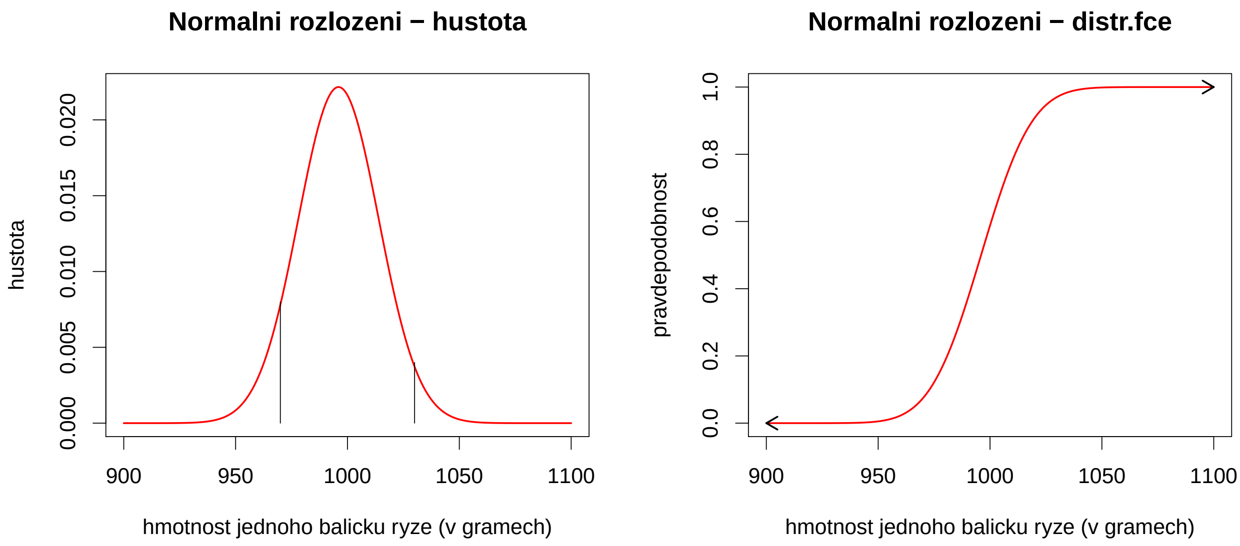
Pravděpodobnost, že náhodná baterie vydrží alespoň 320 hodin je 0.28, pravděpodobnost, že náhodná baterie vydrží nejvýše 310 hodin je 0.61.



Příklad č.7: Na výrobní lince jsou automaticky baleny balíčky rýže o deklarované hmotnosti 1000 g. Působením náhodných vlivů hmotnost balíčků kolísá. Lze ji považovat za náhodnou veličinu, která se řídí normálním rozložením se střední hodnotou 996 g a směrodatnou odchylkou 18 g. Jaká je pravděpodobnost, že náhodně vybraný balíček rýže neprojde výstupní kontrolou, jestliže je povolená tolerance ± 30 g od deklarované hmotnosti 1000 g?

```
[1] 0.1037604
```

Pravděpodobnost, že náhodný balíček rýže neprojde vstupní kontrolou je 0.1.


Normalni rozlozeni − hustota
hustota
0.000
0.005
0.010
0.015
0.020
900
950
1000
1050
1100
hmotnost jednoho balicku ryze (v gramech)
Normalni rozlozeni − distr.fce
pravdepodobnost
0.0
0.2
0.4
0.6
0.8
1.0
900
950
1000
1050
1100
hmotnost jednoho balicku ryze (v gramech)

6 - Výpočet číselných charakteristik náhodných veličin pomocí software R

Příklad č.1:

(a) Nechť $U \sim N(0,1)$. Najděte medián a horní a dolní kvartil.

(b) Nechť $X \sim N(3,5)$. Najděte dolní kvartil.

(c) Určete $\chi^2_{0.025}(25)$.

(d) Určete $t_{0.99}(30)$ a $t_{0.05}(14)$.

(e) Určete $F_{0.975}(5,20)$ a $F_{0.05}(2,10)$.

```
#a)
# median
[1] 0
# dolni kvartil
[1] -0.6744898
# horni kvartil
[1] 0.6744898

#b)
[1] 1.491795

#c)
[1] 13.11972

#d)
[1] 2.457262
[1] -1.76131

#e)
[1] 3.289056
[1] 0.0515573
```

Příklad č.2: Postupně se zkouší spolehlivost čtyř přístrojů. Další přístroj se zkouší jen tehdy, když předchozí je spolehlivý. Každý z přístrojů vydrží zkoušku s pravděpodobností 0.8. Náhodná veličina X udává počet zkoušených přístrojů. Vypočtěte střední hodnotu a rozptyl náhodné veličiny X.

```
#pstni funkce
(p1 <- 0.2)
[1] 0.2
(p2 <- 0.8*0.2)
[1] 0.16
(p3 <- 0.8*0.8*0.2)
[1] 0.128
(p4 <- 0.8*0.8*0.8*0.2+0.8^4)
[1] 0.512

# stredni hodnota
[1] EX = 2.952

# rozptyl
[1] DX = 1.47
```

Příklad č.3: Náhodná veličina X udává počet ok při hodu kostkou. Vypočtěte střední hodnotu a rozptyl náhodné veličiny X.

```
# pstni fce
p1 = p2 = p3 = p4 = p5 = p6 = 1/6

# stredni hodnota
[1] EX = 3.5

# rozptyl
[1] DX = 2.917
```

Příklad č.4: Náhodná veličina X udává příjem manžela (v tisících dolarů) a náhodná veličina Y příjem manželky (v tisících dolarů). Je známa simultánní pravděpodobnostní funkce $\pi(x, y)$ diskrétního náhodného vektoru (X, Y): Vypočtěte koeficient korelace příjmů manžela a manželky.

Tabulka simultánní pstní fce $\pi(X, Y)$				
X - příjem manžela	Y - příjem manželky			
	10	20	30	40
10	0.2	0.04	0.01	0
20	0.1	0.36	0.09	0
30	0	0.05	0.1	0
40	0	0	0	0.05

Vytvořte, funkci **corel.koef**, jejímž vstupem bude matice simultánních pstních fcí A, vektor $x = (10, 20, 30, 40)$ a vektor $y = (10, 20, 30, 40)$ a výstupem bude hledaný koeficient korelace.

```
# Pomocne vysledky mezivypoctu:
# marginalni pstni fce:
#pX
[1] 0.25 0.55 0.15 0.05
#pY
[1] 0.30 0.45 0.20 0.05

#EX
[1] 20
#EY
[1] 20
#DX
[1] 60
#DY
[1] 70
#C(X,Y)
[1] 49

# Korelacni koeficient
> cor.koef(A,x,y)
[1] 0.756
```

Přílad č.5: Diskrétní náhodný vektor (X_1, X_2) má simultánní pravděpodobnostní funkci s hodnotami $\pi(0,-1) = c$, $\pi(0,0) = \pi(0,1) = \pi(1,-1) = \pi(2,-1) = 0$, $\pi(1,0) = \pi(1,1) = \pi(2,1) = 2c$, $\pi(2,0) = 3c$, $\pi(x,y) = 0$ jinak. Určete konstantu c a vypočtěte $R(X_1, X_2)$.

```
# c = 0.1
# Matice simultannich psti B:
[,1] [,2] [,3]
[1,]  0.1  0.0  0.0
[2,]  0.0  0.2  0.2
[3,]  0.0  0.3  0.2

# x:
[1] 0 1 2
# y:
[1] -1  0  1

# Korelacni koeficient:
cor.koef(B,x,y)
[1] 0.424
```

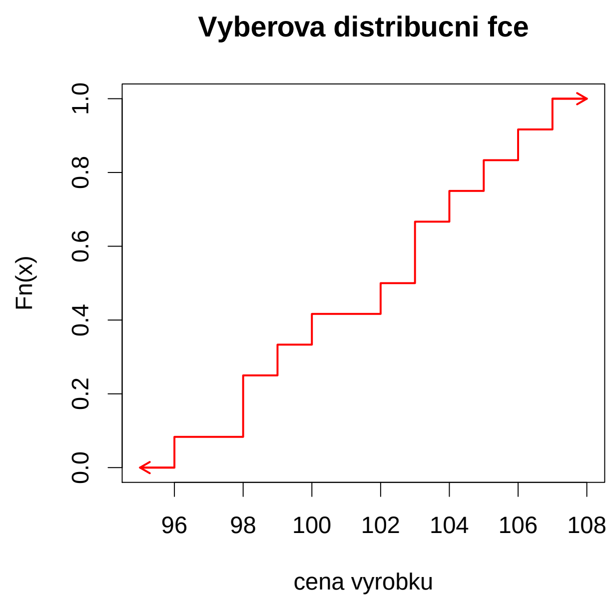
7 - Základní pojmy matematické statistiky

Příklad č.1: Ve 12-ti náhodně vybraných prodejnách ve městě byly zjištěny následující ceny určitého výrobku (v Kč): $102, 99, 106, 103, 96, 98, 100, 105, 103, 98, 104, 107$. Těchto 12 hodnot považujeme za realizace náhodného výběru $X_1, \ldots, X_{12}$ z rozložení, které má střední hodnotu μ a rozptyl σ^2.

(a) Určete nestranné bodové odhady neznámé střední hodnoty μ a neznámého rozptylu σ^2 a směrodatné odchylky σ.

(b) Najděte výběrovou distribuční funkci $F_{12}(x)$ a nakreslete její graf.

```
# a)
# odhad stredni hodnoty
[1] 101.75
# odhad rozptylu
[1] 12.38636
# odhad smerodatne odchylky
[1] 3.519427

# b)Vyberova distribucni fce:
   cena distr.fce
1    95      0.00
2    96      0.08
3    97      0.08
4    98      0.25
5    99      0.33
6   100      0.42
7   101      0.42
8   102      0.50
9   103      0.67
10  104      0.75
11  105      0.83
12  106      0.92
13  107      1.00
14  108      1.00
```



Příklad č.2: Přírůstky cen akcií v % na burze v New Yorku u 10-ti náhodně vybraných společností dosáhly těchto hodnot: $10, 16, 5, 10, 12, 8, 4, 6, 5, 4$.

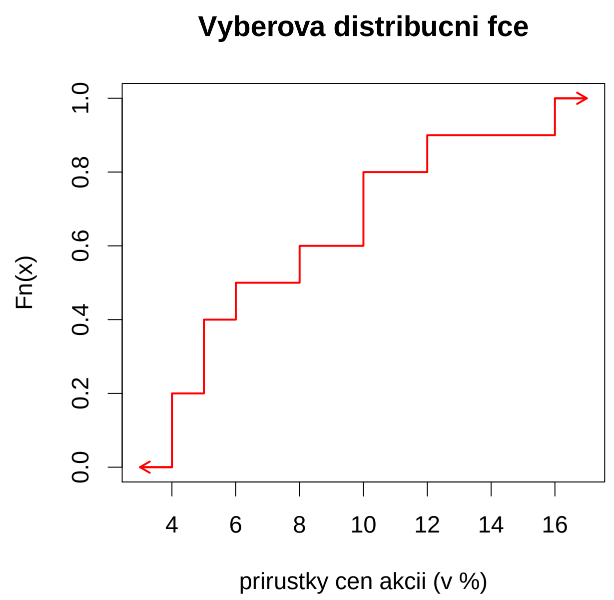
(a) Odhadněte střední hodnotu μ a směrodatnou odchylku σ růstu cen akcií.

(b) Odhadněte pravděpodobnost růstu cen akcií aspoň o 8.5 %.

(c) Nakreslete distrubuční fci.

```
# a)
# odhad stredni hodnoty
[1] 8
# odhad smerodatne odchylky
[1] 3.972125

# b) pravdepodobnost rustu cen akcii aspon o 8.5%
[1] 0.4
```

Pravděpodobnost, že akcie na burze porostou aspoň o 8.5 % je 0.4.



Příklad č.3: Bylo zkoumáno 9 vzorků půdy s různým obsahem fosforu (veličina X). Hodnoty veličiny Y označují obsah fosforu v obilných klíčcích (po 38 dnech), jež vyrostly na těchto vzorcích půdy.

číslo vzorku	1	2	3	4	5	6	7	8	9
X	1	4	5	9	11	13	23	23	28
Y	64	71	54	81	76	93	77	95	109

Těchto 9 dvojic hodnot považujeme za realizace náhodného výběru $(X_1, Y_1), \ldots, (X_9, Y_9)$ z dvourozměrného rozložení s kovariancí σ_{12} a koeficientem korelace ρ. Najděte bodové odhady kovariance σ_{12} a koeficientu korelace ρ. Výslednou hodnotu koeficientu korelace interpretujte.

```
# odhad kovariance
[1] 130
# odhad koeficientu korelace
[1] 0.8049892
```

Mezi obsahem fosforu v půdě a obsahem fosforu v obilných klíčcích existuje silný stupeň přímé lineární závislosti.

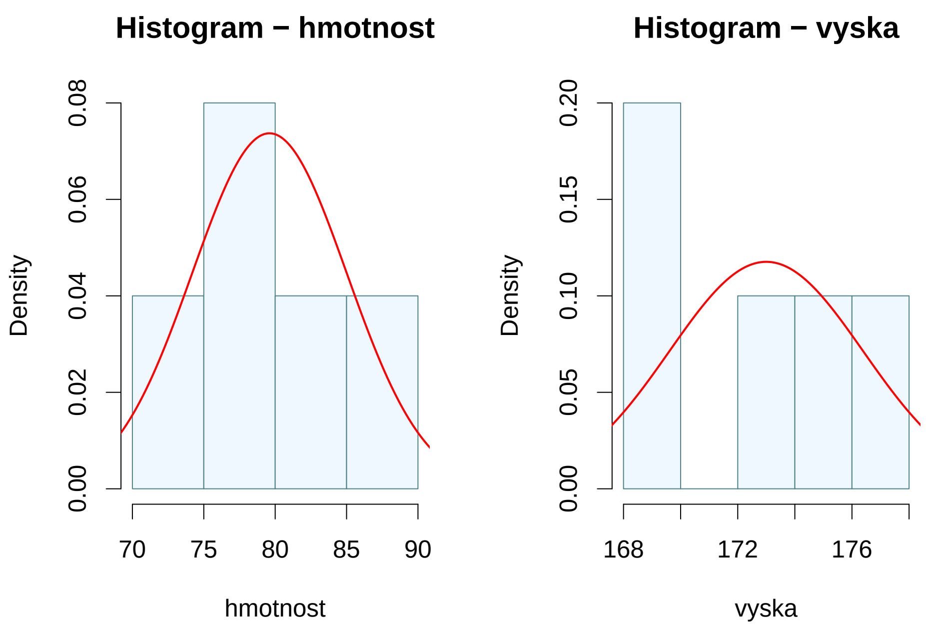
Příklad č.4: Pět mužů zjistilo a zapsalo svou hmotnost (v kg) a výšku (v cm):

Číslo muže	1	2	3	4	5
Hmotnost	76	86	73	84	79
Výška	170	177	169	174	175

Najděte nestranný bodový odhad rozptylu hmotnosti, rozptylu výšky a kovariance hmotnosti a výšky. Vypočtěte rovněž realizaci výběrového koeficientu korelace hmotnosti a výšky. Výslednou hodnotu koeficientu korelace interpretujte. Dále vytvořte histogramy pro hmotnost a výšku.

```
# odhad rozptylu hmotnosti
[1] 29.3
# odhad rozptylu vysky
[1] 11.5
# kovariance hmotnosti a vysky
[1] 16.5
# realizace vyberoveho koeficientu korelace
[1] 0.898879
```

Mezi hmotností a výškou muže existuje silný stupeň přímé lineární závislosti.



Příklad č.5: Při kontrolních zkouškách životnosti 16-ti žárovek byl stanoven odhad $m = 3000\,h$ střední hodnoty jejich životnosti. Z dřívějších zkoušek je známo, že životnost žárovky se řídí normálním rozložením se směrodatnou odchylkou $\sigma = 20\,h$. Vypočtěte

(a) 99 % empirický interval spolehlivosti pro střední hodnotu životnosti;

(b) 90 % levostranný empirický interval spolehlivosti pro střední hodnotu životnosti;

(c) 95 % pravostranný empirický interval spolehlivosti pro střední hodnotu životnosti.

Poznámka: Výsledek zaokrouhlete na jedno desetinné místo a vyjádřete v hodinách a minutách.

```
# a) oboustranny empiricky IS
# dolni hranice
[1] 2987.1
# horni hranice
[1] 3012.9
```

99 % empirický interval spolehlivosti pro střední hodnotu μ životnosti žárovky je

$$(2987.1\,\mathrm{h}\,;\,3012.9\,\mathrm{h})$$

což je v přepočtu na hodiny a minuty:

$$(2987\,\mathrm{h}\ \ 6\,\mathrm{min}\,;\,3012\,\mathrm{h}\ \ 54\,\mathrm{min}).$$

```
# b) levostranny empiricky IS - dolni hranice
[1] 2993.6
```

90 % levostranný empirický interval spolehlivosti pro střední hodnotu μ životnosti žárovky je

$$(2993.6\,\mathrm{h}\,;\,\infty)$$

což je v přepočtu na hodiny a minuty:

$$(2993\,\mathrm{h}\ \ 36\,\mathrm{min}\,;\,\infty).$$

```
# c) pravostranny empiricky IS - horni hranice
[1] 3008.2
```

95 % pravostranný empirický interval spolehlivosti pro střední hodnotu μ životnosti žárovky je

$$(-\infty\,;\,3008.2\,\mathrm{h})$$

což je v přepočtu na hodiny a minuty:

$$(-\infty\,;\,3008\,\mathrm{h}\ \ 12\,\mathrm{min}).$$

Příklad č.6: Víme, že výška hochů ve věku 9.5 let až 10 let má normální rozložení s neznámou střední hodnotou μ a známým rozptylem $\sigma^2 = 39.112\,\mathrm{cm}^2$. Dětský lékař náhodně vybral 15 hochů uvedeného věku, změřil je a vypočítal realizaci výběrového průměru $m = 139.13\,\mathrm{cm}$. Podle jeho názoru by výška hochů v tomto věku neměla přesáhnout 142 cm s pravděpodobností 0.95. Lze tvrzení lékaře akceptovat? Ověřte všemi třemi známými způsoby.

Nulová hypotéza: $H_0 : \mu \leq 142$
Alternativní hypotéza: $H_1 : \mu > 142$.

(a) pomocí kritického oboru

```
#statistika t0:
[1] -1.777348
#dolni hranice kritickeho oboru:
[1] 1.644854
```

Hodnota testové statistiky $t_0 = -1.777$. Kritický obor má tvar $W = \langle 1.645\,;\,\infty)$. Protože $t_0 \notin W$, nulovou hypotézu H_0 nezamítáme na hladině významnosti $\alpha = 0.05$.

(b) pomocí empirického intervalu spolehlivosti
Proti pravostranné alternativě postavíme levostranný empirický interval spolehlivosti:

```
#dolni hranice levostranneho IS
[1] 136.4739
```

Levostranný empirický interval spolehlivosti má tvar $\langle 136.47\,;\,\infty)$. Protože testovaná hodnota $142 \in \langle 136.47\,;\,\infty)$, H_0 nezamítáme na hladině významnosti $\alpha = 0.05$.

(c) pomocí p-hodnoty

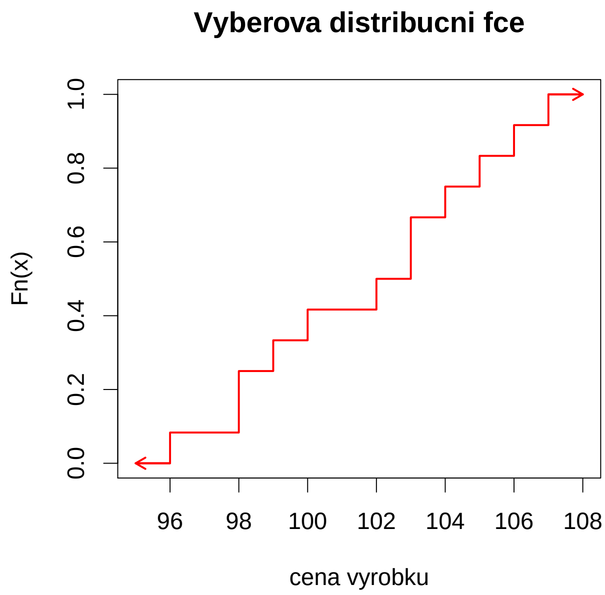
```
#t0
[1] -1.777348
#p-hodnota pro t0
[1] 0.9622445
```

Protože p-hodnota $= 0.9622 > 0.05$, H_0 nezamítáme na hladině významnosti $\alpha = 0.05$.

Závěr: Všechny tři způsoby testování vedou ke stejnému závěru: Tvrzení lékaře o tom, že výška hochů ve věku 9.5 let až 10 let by neměla přesáhnout 142 cm s 95 % pravděpodobností, lze akceptovat.

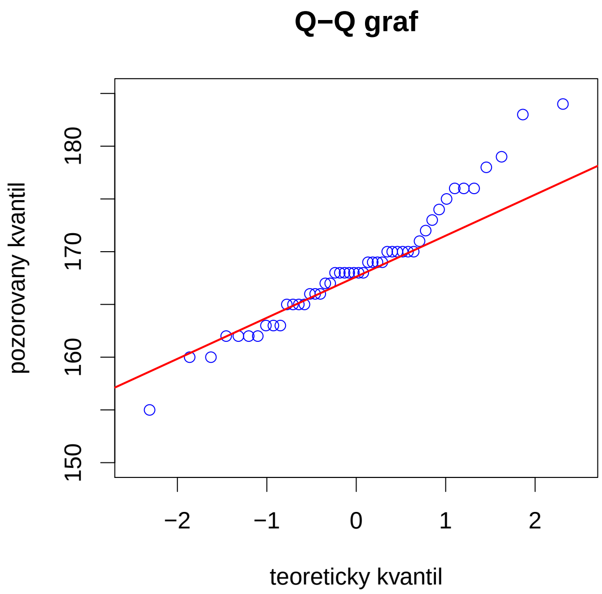
8 - Ověřování normality a parametrické úlohy o jednom náhodném výběru z normálního rozložení a dvourozměrného rozložení

Příklad č.1: Při nanášení tenkých kovových vrstev stříbra na polymerní materiál se vyžaduje, aby tloušťka vrstvy byla 0.020 μm. Pomocí atomové absorpční spektroskopie se zjistily hodnoty, jež jsou uvedeny v tabulce a uloženy v souboru vrstva_stribra.txt. Posuďte Q-Q grafem, zda se výsledky měření řídí normálním rozložením.



Příklad č.2:

1. U 48 studentek VŠE v Praze byla zjišťována výška a obor studia (1 – národní hospodářství, 2 – informatika). Hodnoty jsou uloženy v souboru vyska.txt. Pomocí Q-Q grafu posuďte vizuálně předpoklad normality. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že data pocházejí z normálního rozložení. Hypotézu otestujte pomocí



   (a) Lillieforsovy modifikace K-S testu;

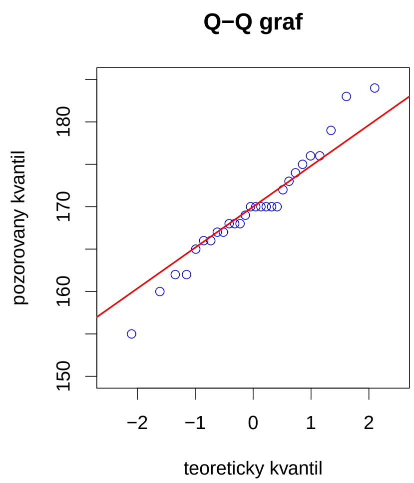
   (b) Shapirova-Wilkova testu;

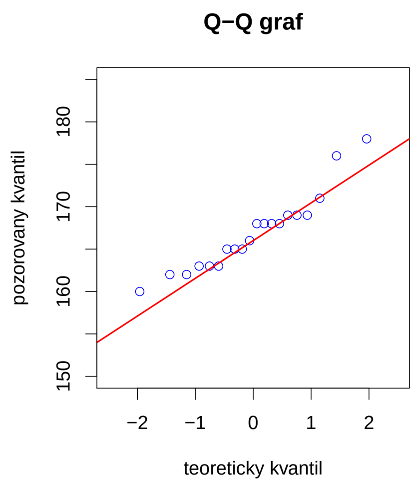
   (c) Andersonova-Darlingova testu;

   (d) Pearsonova χ^2 testu;

```
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  vyska
D = 0.1556, p-value = 0.005258

#-------------------------------------------------------
        Shapiro-Wilk normality test

data:  vyska
W = 0.966, p-value = 0.176
#-------------------------------------------------------
        Anderson-Darling normality test

data:  vyska
A = 0.661, p-value = 0.07933
#-------------------------------------------------------
        Pearson chi-square normality test

data:  vyska
P = 13.25, p-value = 0.06625
```

2. Testy normality a grafické ověření normality proveďte jak pro výšky studentek oboru národní hospodářství, tak pro výška studentek oboru informatiky.






```
#studium narodniho hospodarstvi
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  vyska.h
D = 0.1675, p-value = 0.04293
#--------------------------------------------------------
        Shapiro-Wilk normality test

data:  vyska.h
W = 0.971, p-value = 0.6068
#--------------------------------------------------------
        Anderson-Darling normality test

data:  vyska.h
A = 0.4192, p-value = 0.3053
#--------------------------------------------------------
        Pearson chi-square normality test

data:  vyska.h
P = 4, p-value = 0.5494
#--------------------------------------------------------

#studium informatiky
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  vyska.i
D = 0.1723, p-value = 0.124
#-------------------------------------------------------
        Shapiro-Wilk normality test

data:  vyska.i
W = 0.9227, p-value = 0.1119
#-------------------------------------------------------


        Anderson-Darling normality test
```

```
data:  vyska.i
A = 0.566, p-value = 0.1237
#-------------------------------------------------------
        Pearson chi-square normality test

data:  vyska.i
P = 10.8, p-value = 0.02891
```

Příklad č.3: Předpokládejme, že velký ročník na vysoké škole má výsledky ze statistiky normálně rozloženy kolem střední hodnoty 72 bodů se směrodatnou odchylkou 9 bodů. Najděte pravděpodobnost, že průměr výsledků náhodného výběru 10-ti studentů bude větší než 80 bodů.

```
[1] 0.002470053
```

Příklad č.4: Z populace stejně starých selat téhož plemene bylo vylosováno šest selat a po dobu půl roku jim byla podávána táž výkrmná dieta. Byly zaznamenávány průměrné denní přírůstky hmotnosti v Dg. Z dřívějších pokusů je známo, že v populaci mívají takové přírůstky normální rozložení, avšak střední hodnota i rozptyl se měnívají. Přírůstky v Dg: 62, 54, 55, 60, 53, 58.

(a) Najděte 95% empirický levostranný interval spolehlivosti pro neznámou střední hodnotu μ při neznámé směrodatné odchylce σ.

(b) Najděte 95% empirický interval spolehlivosti pro směrodatnou odchylku σ.

Poznámka: Nezapomeňte před tvorbou intervalů spolehlivosti **ověřit normalitu dat**, která je nezbytným předpokladem zaručujícím spolehlivost intervalů.

Ověření normality dat:

```
        Shapiro-Wilk normality test

data:  selata
W = 0.935, p-value = 0.6195
#-----------------------------------------------
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  selata
D = 0.2119, p-value = 0.5431
```

(a) 95% empirický levostranný interval spolehlivosti pro neznámou střední hodnotu μ:

$$\langle 54.0568\,;\,\infty)$$

(b) Najděte 95% empirický interval spolehlivosti pro směrodatnou odchylku σ:

$$\langle 2.233234\,;\,8.774739)$$

Příklad č.5: Systematická chyba měřicího přístroje se eliminuje nastavením přístroje a měřením etalonu, jehož správná hodnota je $\mu = 10.00$. Nezávislými měřeními za stejných podmínek byly získány hodnoty: 10.24, 10.12, 9.91, 10.19, 9.78, 10.14, 9.86, 10.17, 10.05, které považujeme za realizace náhodného výběru rozsahu 9 z rozložení $N(\mu, \sigma^2)$. Je možné při riziku 0.05 vysvětlit odchylky od hodnoty 10.00 působením náhodných vlivů? Hypotézu otestujte pomocí

(a) kritického oboru;

(b) intervalu spolehlivosti;

(c) p-hodnoty.

Poznámka: Nezapomeňte před samotným testováním hypotéz **ověřit normalitu dat**, která je nezbytným předpokladem zaručujícím spolehlivost testů.

Ověření normality dat:

```
        Shapiro-Wilk normality test

data:  hodnoty
W = 0.9058, p-value = 0.2873
#----------------------------------------------------
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  hodnoty
D = 0.2196, p-value = 0.2404

# a) Testovani pomoci kritickeho oboru:
# statistika t0
[1] 0.9426111
#kriticky obor:
W = (-inf ; -2.306004> a <2.306004; inf)

# b) Testovani pomoci IS:
# dolni hranice IS
[1] 9.926073
# horni hranice IS
[1] 10.17615

# c) Testovani pomoci p-hodnoty:
#p-hodnota
[1] 0.3734702
```

Příklad č.6: U 25-ti náhodně vybraných dvoulitrových lahví s nealkoholickým nápojem byl zjištěn přesný objem nápoje. Výběrový průměr činil $m = 1.99\,\mathrm{l}$ a výběrová směrodatná odchylka $s = 0.1\,\mathrm{l}$. Předpokládejme, že objem nápoje v láhvi je náhodná veličina s normálním rozložením. Na hladině významnosti $\alpha = 0.05$ ověřte tvrzení výrobce, že směrodatná odchylka je $0.08\,\mathrm{l}$. Tvrzení ověřte pomocí

(a) kritického oboru;

(b) intervalu spolehlivosti;

(c) p-hodnoty.

```
# a) Testovani pomoci kritickeho oboru:
# statistika t0
[1] 37.5
#kriticky obor:
W = (0 ; 12.40115> a <39.36408 ; inf)

# b) Testovani pomoci IS:
# dolni hranice IS
[1] 0.006096929
# horni hranice IS
[1] 0.01935304

# c) Testovani pomoci p-hodnoty:
#p-hodnota
[1] 0.0779636
```

Příklad č.7: Bylo vylosováno 6 vrhů selat a z nich vždy dva sourozenci. Jeden z nich vždy dostal náhodně dietu č.1 a druhý dietu č.2. Přírůstky v Dg jsou následující: (62;52), (54;56), (55;49), (60;50), (53;51), (58;50). Za předpokladu, že uvedené dvojice tvoří náhodný výběr z dvourozměrného rozložení s vektorem středních hodnot (μ_1, μ_2) a jejich rozdíly se řídí normálním rozložením, sestrojte 95% interval spolehlivosti pro rozdíl středních hodnot. Pomocí tohoto intervalu otestujte hypotézu, že výkrmná dieta nemá vliv na hmotnostní přírůstky selat.

```
# Interval spolehlivosti:
# dolni hranice
[1] 0.6264613
# horni hranice
[1] 10.70687
```

Příklad č.8: Bylo vybráno šest nových vozů téže značky a po určité době bylo zjištěno, o kolik mm se sjely jejich levé a pravé přední pneumatiky. Výsledky: (1.8; 1.5), (1.0; 1.1), (2.2; 2.0), (0.9; 1.1), (1.5; 1.4), (1.6; 1.4). Za předpokladu, že uvedené dvojice tvoří náhodný výběr z dvourozměrného rozložení s vektorem středních hodnot (μ_1, μ_2) a jejich rozdíly se řídí normálním rozložením, testujte na hladině významnosti $\alpha = 0.05$ hypotézu, že obě pneumatiky se sjíždějí stejně rychle.

```
# a) Testovani pomoci kritickeho oboru:
# statistika t0
[1] 1.051758
#kriticky obor:
W = (-inf ; -2.57058> a <2.57058 ; inf)

# b) Testovani pomoci IS:
# dolni hranice IS
[1] -0.1203401
# horni hranice IS
[1] 0.2870068

# c) Testovani pomoci p-hodnoty:
#p-hodnota
[1] 0.341062
```

9 - Parametrické úlohy o dvou nezávislých náhodných výběrech z normálních rozložení a jednom náhodném výběru z alternativního rozložení

Příklad č.1: Intervaly spolehlivosti pro parametrické funkce $\mu_1 - \mu_2$, σ_1^2/σ_2^2
Bylo vylosováno 11 stejně starých selat téhož plemene. Šesti z nich byla předepsána výkrmná dieta č.1 a zbylým pěti výkrmná dieta č.2. Průměrné denní přírůstky v Dg za dobu půl roku jsou následující:

dieta č.1:	62	54	55	60	53	58
dieta č.2:	52	56	49	50	51	

Zjištěné hodnoty považujeme za realizace dvou nezávislých náhodných výběrů pocházejících z rozložení $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$. Ověřte předpoklad normality.

```
# overeni normality pro dietu c.1:
Shapiro-Wilk normality test

data:  d1
W = 0.935, p-value = 0.6195
#------------------------------------------------------
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  d1
D = 0.2119, p-value = 0.5431


# Overeni normality pro dietu c.2
        Shapiro-Wilk normality test

data:  d2
W = 0.9031, p-value = 0.4272
#------------------------------------------------------
        Lilliefors (Kolmogorov-Smirnov) normality test

data:  d2
D = 0.2412, p-value = 0.4412
```

(a) Sestrojte 95 % empirický interval spolehlivosti pro podíl rozptylů. Pomocí tohoto intervalu otestujte hypotézu, že rozptyly σ_1^2 a σ_2^2 jsou shodné.

Empirický interval spolehlivosti má tvar:

$$\langle 0.1872423\,;\, 12.9541\rangle.$$

(b) Za předpokladu, že data pocházejí z rozložení $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$, sestrojte 95 % empirický interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$.

Poznámka - Mezivýpočet: Vážený průměr výběrových rozptylů má tvar: $S_*^2 = 10.35556$.

Empirický interval spolehlivosti má tvar:

$$\langle 0.9919634\,;\, 9.808037\rangle$$

Příklad č.2: Jsou dány dva nezávislé náhodné výběry o rozsazích $n_1 = 25$, $n_2 = 10$, první pochází z rozložení $N(\mu_1, \sigma_1^2)$, druhý z rozložení $N(\mu_2, \sigma_2^2)$, kde parametry μ_1, μ_2, σ_1^2, σ_2^2 neznáme. Byly vypočteny realizace výběrových rozptylů: $\sigma_1^2 = 1.7482, \sigma_2^2 = 1.7121$. Sestrojte 95 % empirický interval spolehlivosti pro podíl rozptylů.

Empirický interval spolehlivosti pro podíl rozptylů σ_1^2/σ_2^2 má tvar:

$$\langle 0.2825207\,;\, 2.759698)$$

Příklad č.3: Testování hypotéz o parametrických funkcích $\mu_1 - \mu_2$, σ_1^2/σ_2^2
Pro datový soubor z příkladu č.1 testujte na hladině významnosti $\alpha = 0.05$ hypotézu, že

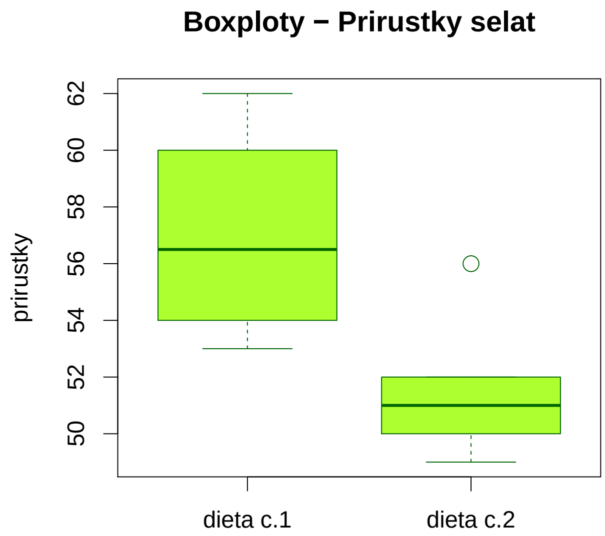
(a) rozptyly hmotnostních přírůstků selat při obou výkrmných dietách jsou shodné;

```
# Testovani pomoci kritickeho oboru:
# statistika t0
[1] 1.753425
#kriticky obor
W = (0 ; 0.1353567> a <9.364471 ; inf)

# Testovani pomoci IS:
# dolni hranice
[1] 0.1872423
# horni hranice
[1] 12.9541

# Testovani pomoci p-hodnoty
# p-hodnota
[1] 0.6063451
```

(b) obě výkrmné diety mají stejný vliv na hmotnostní přírůstky selat.

```
# Testovani pomoci kritickeho oboru:
# statistika t0
[1] 2.771222
# kriticky obor
W = (-inf ;  -2.262157> a <2.262157 ; inf)

# Testovani pomoci IS:
# dolni hranice
[1] 0.9919634
# horni hranice
[1] 9.808037

# Testovani pomoci p-hodnoty
# p-hodnota
[1] 0.02171008
```

Dále sestrojte krabicové grafy pro hmotnostní přírůstky selat obou výkrmných diet.



Příklad č.4: Načtěte datový soubor vyska.txt, který obsahuje údaje o výšce 48 studentek VŠE v Praze (proměnná vyska) a obor jejich studia (1 – národní hospodářství, 2 – informatika).

(a) Pomocí S-W testu ověřte na hladině významnosti $\alpha = 0.1$ předpoklad o normalitě výšek v obou skupinách studentek.

```
# Testovani normality dat pro studentky z oboru narodniho hospodarstvi:
        Shapiro-Wilk normality test

data:  vyska.h
W = 0.971, p-value = 0.6068

# Testovani normality dat pro studentky z oboru informatiky:
        Shapiro-Wilk normality test

data:  vyska.i
W = 0.9227, p-value = 0.1119
```

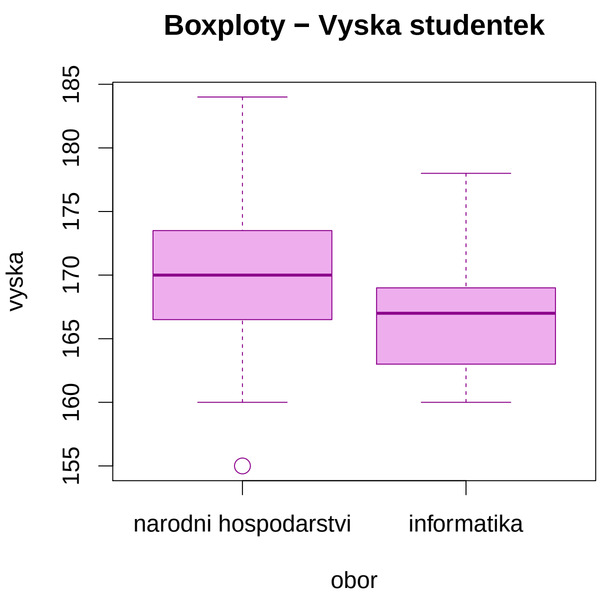
(b) Na hladině významnosti $\alpha = 0.1$ testujte hypotézu o shodě rozptylů výšek studentek v daných dvou oborech studia.

```
# Testovani pomoci kritickeho oboru:
# statistika t0
[1] 1.987288
# kriticky obor
W = (0 ; 0.503273> a <2.090489 ; inf)

# Testovani pomoci IS:
# dolni hranice
[1] 0.9506332
# horni hranice
[1] 3.948727

# Testovani pomoci p-hodnoty
# p-hodnota
[1] 0.1249251
```

(c) Na hladině významnosti $\alpha = 0.1$ testujte hypotézu o shodě středních hodnot výšek studentek v daných dvou oborech studia.

```
# Testovani pomoci kritickeho oboru:
# statistika t0
[1] 1.744008
# kriticky obor
W = (-inf ; -1.67866> a <1.67866 ; inf)

# Testovani pomoci IS:
# dolni hranice
[1] 0.1094654
# horni hranice
[1] 5.733392

# Testovani pomoci p-hodnoty
# p-hodnota
[1] 0.08783749
```

(d) Výpočet doplňte krabicovými diagramy.



Příklad č.5: Asymptotický interval spolehlivosti pro parametr θ alternativního rozložení Může politická strana, pro niž se v předvolebním průzkumu vyslovilo 60 z 1000 dotázaných osob, očekávat se spolehlivostí 0.95, že by v této době ve volbách překročila 5 % hranici pro vstup do parlamentu? Pro stanovení závěru využijte interval spolehlivosti.
Poznámka: Nezapomeňte před samotným výpočtem **ověřit tzv. podmínku dobré aproximace** (Haldovu podmínku), jejíž splnění je nezbytné pro relevantnost závěru.

Ověření Haldovy podmínky:

```
# Haldova podminka
[1] 56.4 > 9
```

95 % levostranný empirický IS má tvar

$$\langle 0.04765\,;\,\infty).$$

V intervalu jsou zahrnuty i hodnoty menší než 0.05 (tedy 5 %). Může tedy nastat situace, že politická strana 5 % hlasů pro vstup do parlamentu nezíská.

Příklad č.6: Přírůstky cen akcií na burze (v %) u 10-ti náhodně vybraných společností dosáhly těchto hodnot: $10, 16, 5, 10, 12, 8, 4, 6, 5, 4$. Sestrojte 95 % asymptotický empirický interval spolehlivosti pro pravděpodobnost, že přírůstek ceny akcie překročí 8.5 %.

Ověření Haldovy podmínky:

```
# Haldova podminka
[1] 2.4 < 9
```

Haldova podmínka není splněna, i přesto si interval cvičně vypočítáme. V praxi by nám ale nesplnění podmínky mělo být varováním, že výsledný interval není zcela spolehlivý.

95 % asymptotický empirický IS pro pravděpodobnost ϑ má tvar

$$(0.096\,;\,0.704).$$

To znamená, že pravděpodobnost, že přírůstek ceny akcie překročí 8.5 %, je alespoň 9.6 % a nanejvýš 70.4 % (při spolehlivosti 95 %).

Příklad č.7: Určitá cestovní kancelář organizuje zahraniční zájezdy podle individuálních přání zákazníků. Z několika minulých let ví, že 30 % všech takto organizovaných zájezdů má za cíl zemi X. Po zhoršení politických podmínek v této zemi se cestovní kancelář obává, že se zájem o tuto zemi mezi zákazníky sníží. Ze 150-ti náhodně vybraných zákazníků v tomto roce má 38 za cíl právě zemi X. Potvrzují nejnovější data pokles zájmu o tuto zemi? Volte hladinu významnosti $\alpha = 0.05$.

Ověření Haldovy podmínky:

```
# Haldova podminka
[1] 31.5 > 9
```

Testování hypotézy:

```
# Testovani pomoci kritickeho oboru:
# statistika t0
[1] -1.247219
# kriticky obor
W = (-inf ; -1.644854>

# Testovani pomoci IS:
# dolni hranice
[1] 0
#horni hranice
[1] 0.3117439

# Testovani pomoci p-hodnoty
# p-hodnota
[1] 0.1061586
```

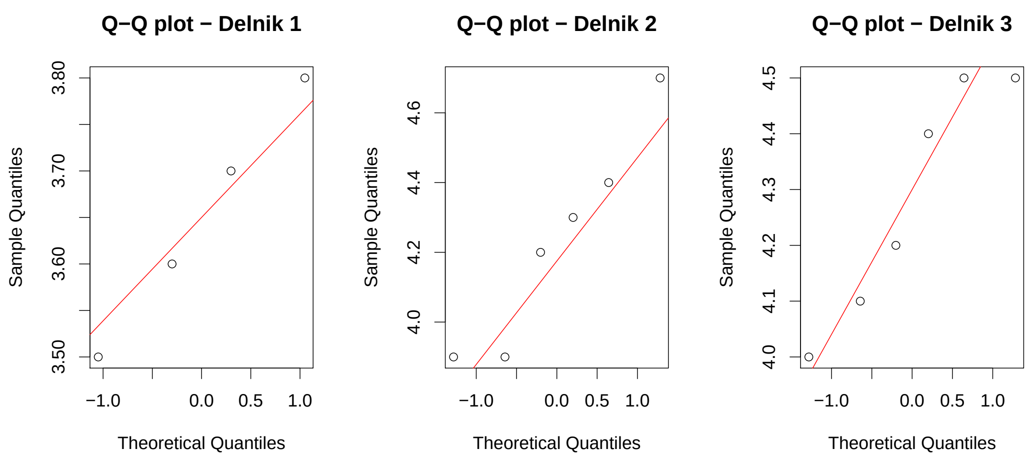
10 - Analýza rozptylu jednoduchého třídění

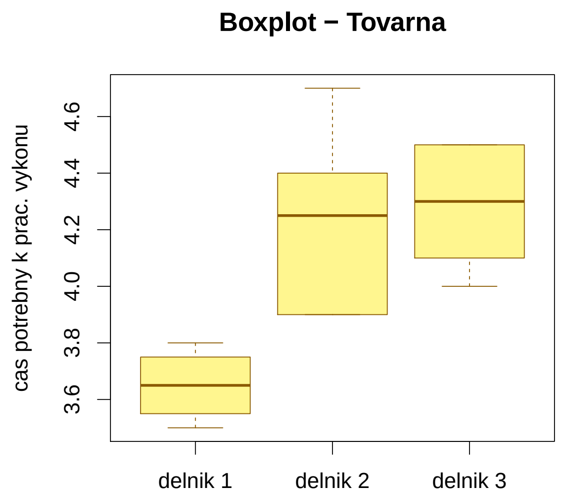
Příklad č.1: Ústav antropologie vypsal konkurz na přijetí nového antropologa do svých řad. Ředitel ústavu se rozhodl, že nedá na hezký obličejík a naučené fráze a vezme někoho, kdo je ve svém oboru zručný. Každý uchazeč měl za úkol provést v rámci pohovoru několik měření a byl mu stopován čas potřebný k měření. Konkurzu se zúčastnili tři kandidáti. Časy jejich měření v minutách jsou zaznamenány v tabulce:

1.antropolog:	3.6	3.8	3.7	3.5		
2.antropolog:	4.3	3.9	4.2	3.9	4.4	4.7
3.antropolog:	4.2	4.5	4.0	4.1	4.5	4.4

Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že rychlost měření těchto tří antropologů jsou stejné. Zamítnete-li nulovou hypotézu, určete, výkony kterých antropologů se liší na dané hladině významnosti $\alpha = 0.05$ a stanovte závěr, který by ředitele ústavu mohl zajímat.

Poznámka: Před samotným testováním **nezapomeňte ověřit, že všechny tři výběry pochází z normálních rozložení a že rozptyly těchto výběrů jsou shodné**. Jsou to důležité předpoklady, které musí být splněny, abychom mohli analýzu rozptylu použít. Normalitu otestujte pomocí S-W testu a graficky pomocí Q-Q grafu, shodu rozptylů potom ověřte pomocí Levenova testu a graficky pomocí krabicových diagramů. Proč nemůžeme k otestování shody rozptylů použít Bartlettův test?





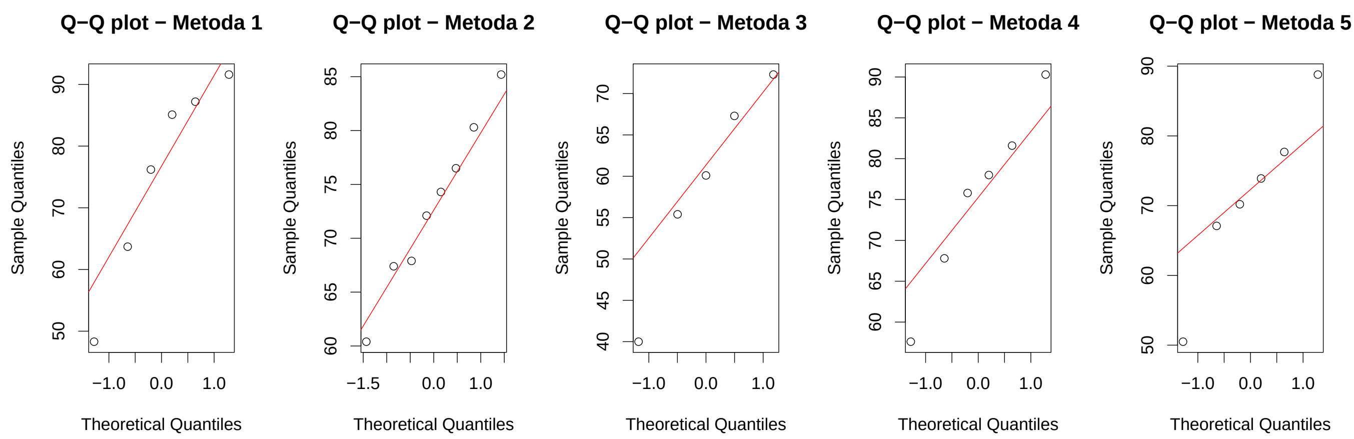

```
#[1] "Shapiro test - promenna 1 = 0.9719"
#[1] "Shapiro test - promenna 2 = 0.5819"
#[1] "Shapiro test - promenna 3 = 0.3313"

#[1] "Levene test = 0.2621"

    SA fA     SE fE     ST fT    Fa
 1.118  2 0.752 13 1.869 16 9.665

#[1] "p-value = 0.00268"

#[1] "Scheffeho motoda:"
     [,1] [,2] [,3]
[1,]    0    1    1
[2,]    1    0    0
[3,]    1    0    0
```

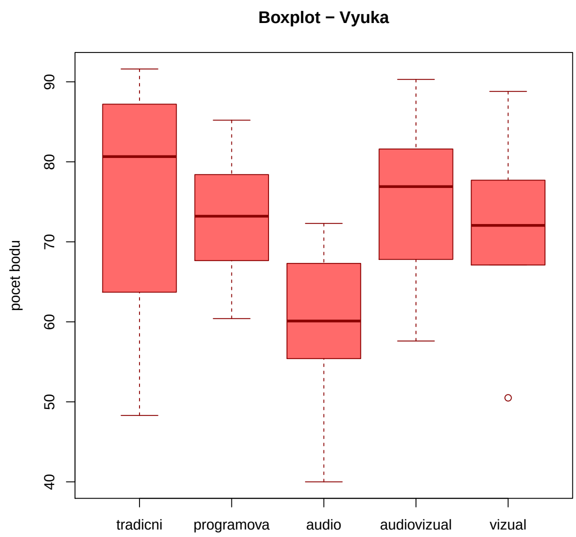
Příklad č.2: Na střední škole byl uskutečněn experminet zjišťující efektivitu jednotlivých pedagogických metod. Studenti byli rozděleni do pěti supin a každá skupina byla vyučována pomocí jedné z pedagogických metod: tradiční způsob, programová výuka, audiotechnika, audiovizuální technika a vizuální technika. Z každé skupiny byl potom vybrán náhodný vzorek studentů a všichni byli podrobeni témuž písemnému testu. Výsledky testu jsou uvedeny v následující tabulce a v souboru vyukove metody.txt:

metoda	počet bodů							
tradicni	76.2	48.3	85.1	63.7	91.6	87.2		
programova	85.2	74.3	76.5	80.3	67.4	67.9	72.1	60.4
audio	67.3	60.1	55.4	72.3	40.0			
audiovizualni	75.8	81.6	90.3	78.0	67.8	57.6		
vizualni	50.5	70.2	88.8	67.1	77.7	73.9		

Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že znalosti všech studentů jsou stejné a nezávisí na použité pedagogické metodě. V případě zamítnutí hypotézy zjistěte, které výběry se liší na hladině významnosti 0.05.

Poznámka: Před samotným testováním **nezapomeňte ověřit, že všechny tři výběry pochází z normálních rozložení a že rozptyly těchto výběrů jsou shodné**. Jsou to důležité předpoklady, které musí být splněny, abychom mohli analýzu rozptylu použít. Normalitu otestujte pomocí S-W testu a graficky pomocí Q-Q grafu, shodu rozptylů potom ověřte pomocí Levenova testu a Bartlettova testu a graficky pomocí krabicových diagramů.






```
#[1] "Shapiro test - promenna 1 = 0.4177"
#[1] "Shapiro test - promenna 2 = 0.9966"
#[1] "Shapiro test - promenna 3 = 0.7663"
#[1] "Shapiro test - promenna 4 = 0.9577"
#[1] "Shapiro test - promenna 5 = 0.8814"

#[1] "Bartlett test = 0.5524"

#[1] "Levene test = 0.6513"

       SA fA       SE fE       ST fT    Fa
```

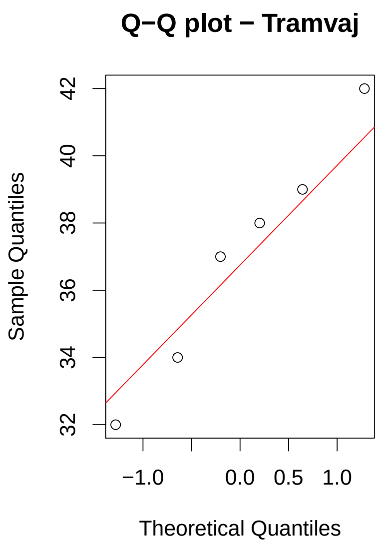
```
  966.374  4 3868.773 26 4835.147 31 1.624

#[1] "p-value = 0.19825"
```

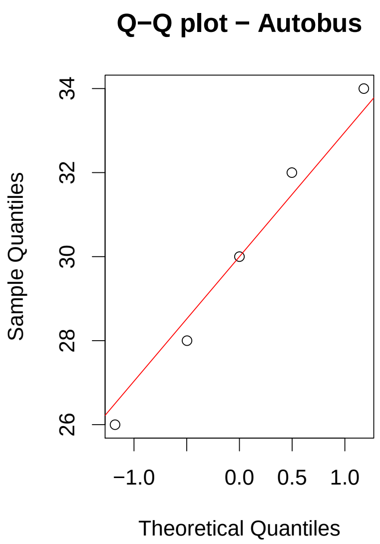
Příklad č.3: Pan Novák může cestovat z místa bydliště do místa pracoviště třemi různými způsoby: tramvají (způsob A), autobusem (způsob B) a metrem s následným přestupem na tramvaj (způsob C). Máme k dispozici jeho naměřené časy cestování do práce v době ranní špičky (včetně čekání na příslušný spoj) v minutách:

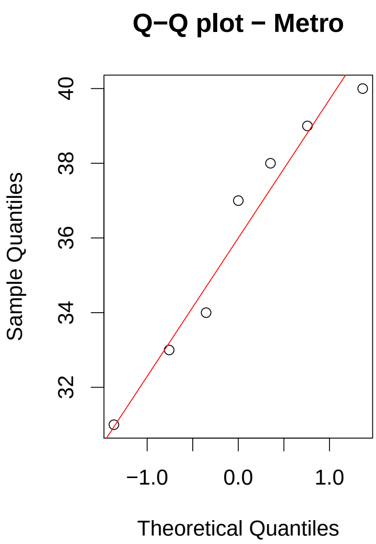
způsob A:	32	39	42	37	34	38	
způsob B:	30	34	28	26	32		
způsob C:	40	37	31	39	38	33	34

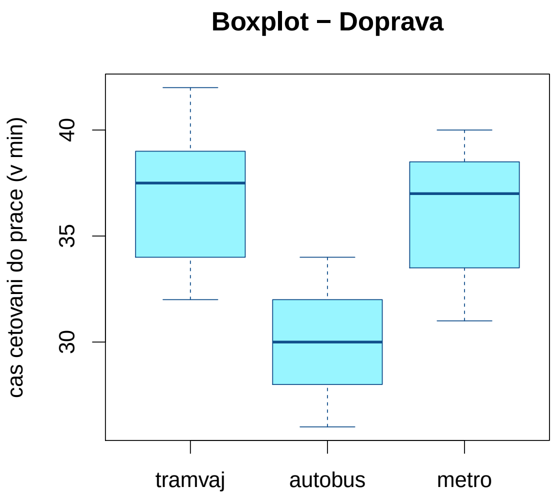
Pro všechny tři způsoby dopravy vypočtěte průměrné časy cestování. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že doba cestování do práce nezávisí na způsobu dopravy. V případě zamítnutí nulové hypotézy zjistěte, které způsoby dopravy do práce se od sebe liší na hladině významnosti $\alpha = 0.05$.

Poznámka: Před samotným testováním nezapomeňte ověřit, že všechny tři výběry pochází z normálních rozložení a že rozptyly těchto výběrů jsou shodné.









```
# Vypocet prumeru:
# tramvaj
[1] 37
# autobus
[1] 30
# metro
[1] 36
#-----------------------------------------------------------------
#[1] "Shapiro test - promenna 1 = 0.9539"
#[1] "Shapiro test - promenna 2 = 0.9672"
#[1] "Shapiro test - promenna 3 = 0.6294"

#[1] "Levene test = 0.9597"

   SA fA  SE fE  ST fT    Fa
  154  2 172 15 326 18 6.715

#[1] "p-value = 0.00827"

#[1] "Scheffeho motoda:"
     [,1] [,2] [,3]
[1,]    0    1    0
[2,]    1    0    1
[3,]    0    1    0
```

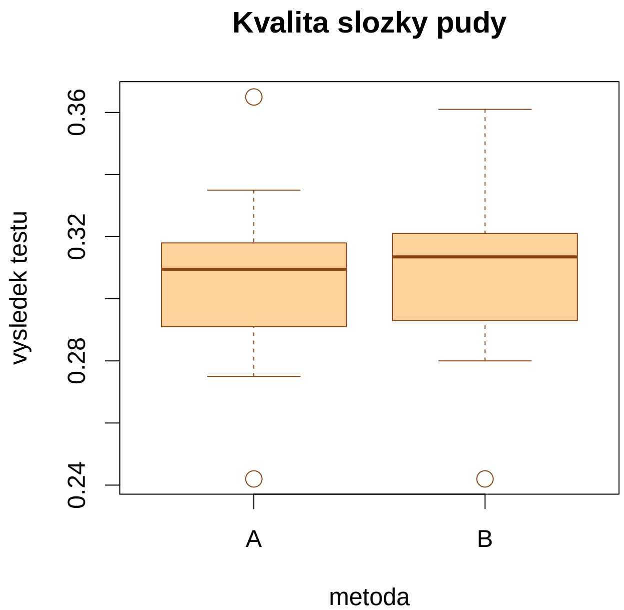
11 - Neparametrické úlohy o mediánech

Příklad č.1: Párový znaménkový test a párový Wilcoxonův test Při zjišťování kvality jedné složky půdy se používají dvě metody označené A a B. Výsledky jsou uvedeny v následující tabulce:

Vzorek	1	2	3	4	5	6	7	8	9	10	11	12
A	0.275	0.312	0.284	0.3	0.365	0.298	0.312	0.315	0.242	0.321	0.335	0.307
B	0.28	0.312	0.288	0.298	0.361	0.307	0.319	0.315	0.242	0.323	0.341	0.315

Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že metody A a B dávají stejné výsledky.
K testování použijte jak párový znaménkový test, tak párový Wilcoxonův test. Pro lepší představu sestrojte krabicové diagramy pro obě metody.

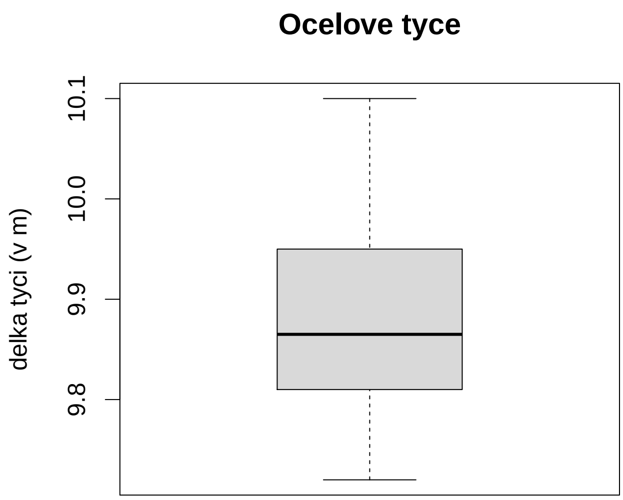
```
# Parovy znamenkovy test
Dependent-samples Sign-Test

data:  x1 and x2
S = 2, p-value = 0.1797
alternative hypothesis: true median difference is not equal to 0
#----------------------------------------------------------------------
# Parovy Wilcoxonuv test
        Wilcoxon signed rank test with continuity correction

data:  x1 and x2
V = 5, p-value = 0.04364
alternative hypothesis: true location shift is not equal to 0
```



Příklad č.2: Jednovýběrový znaménkový test a jednovýběrový Wilcoxonův test Vyráběné ocelové tyče mají kolísavou délku s předpokládanou hodnotou mediánu 10 m. Náhodný výběr 10-ti tyčí poskytl tyto výsledky: 9.83, 10.10, 9.72, 9.91, 10.04, 9.95, 9.82, 9.73, 9.81, 9.90. Na hladině významnosti 0.05 testujte hypotézu, že předpoklad o mediánu délky tyčí je oprávněný.

K testování použijte jak jednovýběrový znaménkový test, tak jednovýběrový Wilcoxonův test. Pro lepší představu sestrojte krabicový diagram.

```
# Jednovyberivy znamenkovy test
One-sample Sign-Test

data:  x
s = 2, p-value = 0.1094
alternative hypothesis: true median is not equal to 10
95 percent confidence interval:
  9.755956 10.010800
#-----------------------------------------------------------------
# Jednovyberovy Wilcoxonuv test
        Wilcoxon signed rank test with continuity correction

data:  x
V = 5.5, p-value = 0.02831
alternative hypothesis: true location is not equal to 10
```



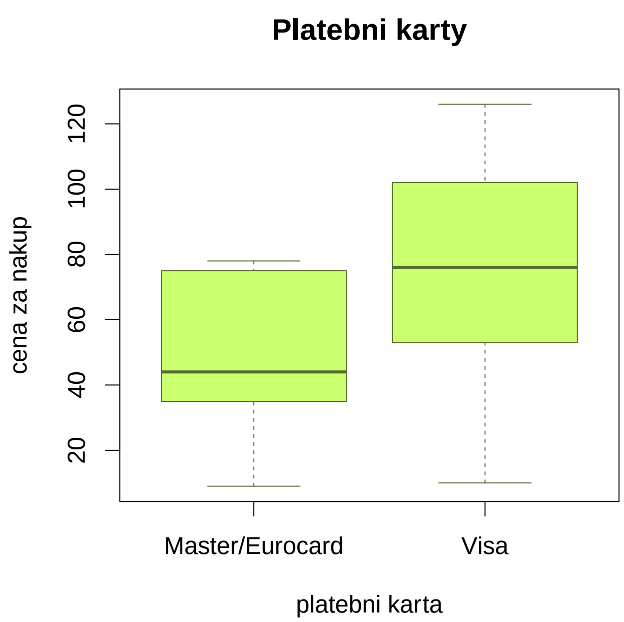
Příklad č.3: Dvouvýběrový Wilcoxonův test a dvouvýběrový K-S test Majitel obchodu chtěl zjistit, zda velikost nákupů (v dolarech) placených kreditními kartami Master/EuroCard a Visa jsou přibližně stejné. Náhodně vybral

- 7 nákupů placených Master/EuroCard: 42, 77, 46, 73, 78, 33, 37;
- 9 nákupů placených Visou: 39, 10, 119, 68, 76, 126, 53, 79, 102.

Lze na hladině významnosti $\alpha = 0.05$ tvrdit, že velikost nákupů placených těmito dvěma typy karet se shodují?

K testování použijte dvouvýběrový Wilcoxonův test a Kolmogorův-Smirnovův test. Pro lepší představu sestrojte krabicové diagramy pro oba typy platebních karet.

```
# Dvouvyberovy Wilcoxonuv test
Wilcoxon rank sum test

data:  x and y
W = 20, p-value = 0.1388
alternative hypothesis: true location shift is not equal to 0
#--------------------------------------------------------------
#Dvouvyberovy Kolmogoruv-Smirnovuv test
        Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.4444, p-value = 0.2425
alternative hypothesis: two-sided
```



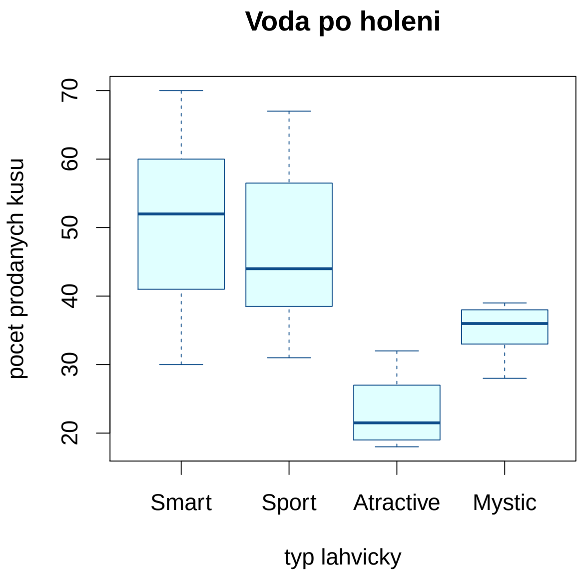
Příklad č.4: Kruskalův – Wallisův test a mediánový test Voda po holení jisté značky se prodává ve čtyřech různých lahvičkách stejného obsahu. Údaje o počtu prodaných lahviček za týden v různých obchodech jsou uvedeny v následující tabulce:

Smart:	50	35	43	30	62	52	43	57	33	70	64	58	53	65	39
Sport:	31	37	59	67	44	49	54	62	34	42	40				
Atractive:	27	19	32	20	18	23									
Mystic:	35	39	37	38	28	33									

Posuďte na 5 % hladině významnosti, zda typ lahvičky ovlivňuje úroveň prodeje. V případě zamítnutí nulové hypotézy zjistěte, prodeje kterých typů lahviček se od sebe významně liší.

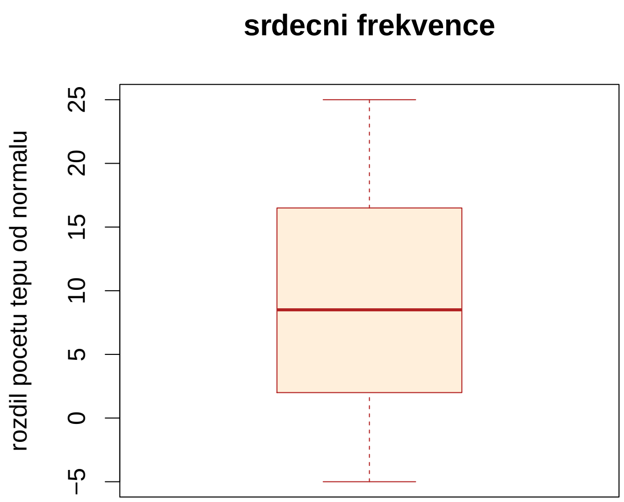
K testování použijte Kruskalův – Wallisův test i mediánový test; v případě zamítnutí nulové hypotézy použijte k zjištění významných rozdílů vhodnou metodu mnohonásobného porovnávání. Pro lepší představu sestrojte krabicové diagramy pro všechny typy lahviček.

```
# Medianovy test
[1] Q = 17.5394
# dolni hranice kritickeho oboru W
[1] dh = 7.8147
#-------------------------------------------------------------
# Kruskaluv-Wallisuv test
        Kruskal-Wallis rank sum test
data:  x and group
Kruskal-Wallis chi-squared = 18.802, df = 3, p-value = 0.0003004
#--------------------------------------------------------------
#Obecna metoda mnohonasobneho porovnavani
     [,1] [,2] [,3] [,4]
[1,]    0    0    1    0
[2,]    0    0    1    0
[3,]    1    1    0    0
[4,]    0    0    0    0
```



Příklad č.5: Ve skupině 12-ti studentů se sledovala srdeční frekvence při změně polohy z lehu do stoje. Získaly se tyto rozdíly počtu tepů srdce za 1 minutu: -2, 4, 8, 25, -5, 16, 3, 1, 12, 17, 20, 9. Za předpokladu, že tyto rozdíly mají symetrické rozložení, testujte na hladině významnosti $\alpha = 0.05$ hypotézu, že medián rozdílů obou tepových frekvencí je 15 proti oboustranné alternativě. Sestrojte krabicový diagram.

```
# Jednovyberovy Wilcoxonuv test
        Wilcoxon signed rank test

data:  x
V = 14, p-value = 0.05225
alternative hypothesis: true location is not equal to 15
#---------------------------------------------------------
# Jednovyberovy znamenkovy test
        One-sample Sign-Test

data:  x
s = 4, p-value = 0.3877
alternative hypothesis: true median is not equal to 15
95 percent confidence interval:
  1.212727 16.893636
```

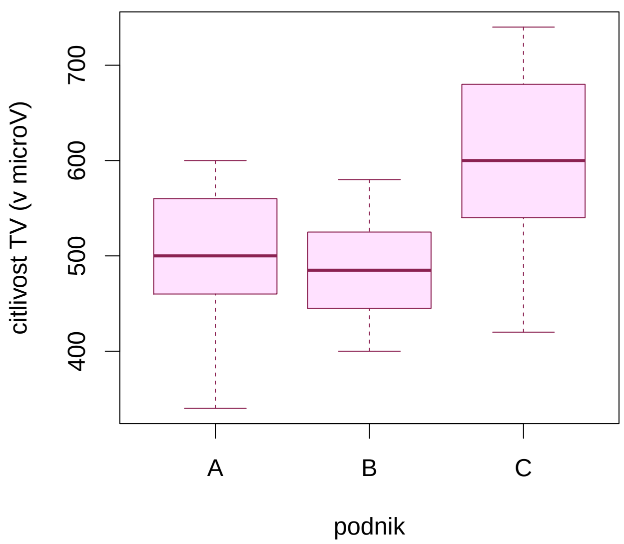


Příklad č.6: Z produkce tří podniků vyrábějících televizory bylo vylosováno 10, 8 a 12 kusů. Byly získány následující výsledky zjišťování citlivosti těchto televizorů v mikrovoltech:

podnik A:	420	560	600	490	550	570	340	480	510	460		
podnik B:	400	420	580	470	470	500	520	530				
podnik C:	450	700	630	590	420	590	610	540	740	690	540	670

Ověřte na hladině významnosti $\alpha = 0.05$ hypotézu o shodě úrovně citlivosti televizorů v jednotlivých podnicích. Sestrojte krabicové diagramy pro všechny tři podniky.

```
# Kruskal-Wallis test
Kruskal-Wallis rank sum test

data:  citlivost and podnik
Kruskal-Wallis chi-squared = 8.3047, df = 2, p-value = 0.01573

# Medianovy test
[1] "Q␣=␣10.2333"
# dolni hranice kritickeho oboru
[1] "dh␣=␣5.9915"

# Obecna metoda mnohonasobneho porovnavani
     [,1] [,2] [,3]
[1,]    0    0    0
[2,]    0    0    1
[3,]    0    1    0
```



12 - Hodnocení kontingenčních tabulek

Příklad č.1: Testování hypotézy o nezávislosti, měření síly závislosti
V roce 1950 zkoumali Yule a Kendall barvu očí a vlasů u 6800 mužů.

Barva očí	Barva vlasů			
	světlá	kaštanová	černá	rezavá
modrá	1768	807	180	47
šedá/zelená	946	1387	746	53
hnědá	115	438	288	16

Na asymptotické hladině významnosti $\alpha = 0.05$ testujte hypotézu o nezávislosti barvy očí a barvy vlasů. Vypočtěte Cramérův koeficient. *Poznámka:* Nezapomeňte před samotným testováním ověřit podmínky dobré aproximace.

```
# Podminky dobre aproximace
          [,1]     [,2]     [,3]     [,4]
[1,] 1167.2593 1085.976 500.9024 47.86217
[2,] 1304.7310 1213.875 559.8952 53.49904
[3,]  357.0097  332.149 153.2025 14.63879

# K = 1088.1485
# W = ( 12.5916 ; Inf )
# p_hodnota = 0
# Crameruv_koef = 0.283
```

Příklad č.2: Na hladině významnosti $\alpha = 0.05$ testujte hypotézu o nezávislosti pedagogické hodnosti a pohlaví. Dále vypočtěte Cramérův koeficient vyjadřující intenzitu závislosti pedagogické hodnosti na pohlaví, jsou-li k dispozici následující údaje:

Pohlaví	Pedagogická hodnost		
	odb. asistent	docent	profesor
muž	32	15	8
žena	34	8	3

Poznámka: Nezapomeňte před samotným testováním ověřit podmínky dobré aproximace.

```
# Podminky dobre aproximace
     [,1]  [,2] [,3]
[1,] 36.3 12.65 6.05
[2,] 29.7 10.35 4.95

# K = 3.4988
# W = ( 5.9915 ; Inf )
# p_hodnota = 0.1739
# Crameruv_koef = 0.1871
```

Příklad č.3: Fisherův faktoriálový test
100 náhodně vybraných mužů a žen bylo dotázáno, zda dávají přednost nealkoholickému nápoji A či B. Údaje jsou uvedeny ve čtyřpolní kontingenční tabulce.

pref. nápoj	pohlaví	
	muž	žena
A	20	30
B	30	20

Na hladině významnosti $\alpha = 0.05$ testujte pomocí Fisherova faktoriálového testu hypotézu, že preferovaný typ nápoje nezáleží na pohlaví respondenta.

```
        Fisher's␣Exact␣Test␣for␣Count␣Data

data:␣␣napoje
p-value␣=␣0.07134
alternative␣hypothesis:␣true␣odds␣ratio␣is␣not␣equal␣to␣1
```

Příklad č.4: Podíl šancí Pro údaje z příkladu č.4 vypočtěte podíl šancí a sestrojte 95 % asymptotický interval spolehlivosti pro logaritmus podílu šancí. Pomocí tohoto intervalu spolehlivosti testujte na asymptotické hladině významnosti $\alpha = 0.05$ hypotézu, že preferovaný typ nápoje nezáleží na pohlaví respondenta.

```
# podil sanci:
[1] 0.4444444
# dolni hrancie asymptotickeho IS pro logaritmus podilu sanci:
[1] -1.611082
# horni hranice asymptotickeho IS pro logaritmus podilu sanci:
[1] -0.01077827
```

Příklad č.5: 36 mužů onemocnělo určitou chorobou. Někteří z nich se léčili, jiní ne. Někteří se uzdravili, jiní zemřeli. Údaje jsou uvedeny ve čtyřpolní kontingenční tabulce.

přežití	léčení	
	ano	ne
ano	10	6
ne	12	8

Vypočtěte a interpretujte podíl šancí. Pomocí intervalu spolehlivosti pro logaritmus podílu šancí testujte na asymptotické hladině významnosti $\alpha = 0.05$ hypotézu, že přežití nezávisí na léčení proti tvrzení, že léčení zvyšuje šance na přežití.

```
# podil sanci:
[1] 1.111111
# dolni hranice levostranneho IS:
[1] -1.028277
# IS: < -1.028 ; Inf)
```