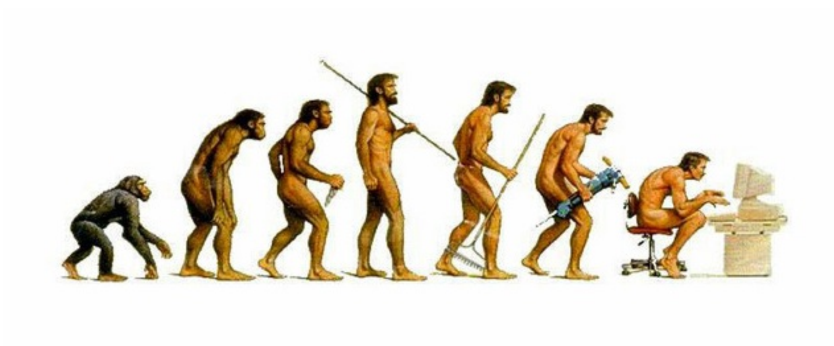
Masarykova univerzita v Brně
Přírodovědecká fakulta

SBÍRKA PŘÍKLADŮ K PŘEDMĚTU APLIKOVANÁ STATISTIKA I



Brno, 2015

1 - Základní práce se softwarem R - Příkazy

V úvodní hodině se seznámíme se statistickým softwarem R a naučíme se používat základní příkazy, načítat datové soubory, kreslit grafy a exportovat je do pdf souborů.

Příklad č.1: Vyzkoušejte si následující příkazy a do následující hodiny si zautomatizujte používanou syntaxi.

```
#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)
(x2<-x^2)
(A2<-A^2)

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

# min a max
min(a)
max(a)
min(x)
max(x)
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')

#export grafu do pdf souboru
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.

1. Vytvořte variační řadu (tabulku rozložení četností)
   (a) známek z matematiky (znak X);
   (b) známek z angličtiny (znak Y).
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;
   (b) pouze pro muže.
5. Vytvořte kontingenční tabulku simultánních absolutních četností znaků X a Y.
6. Vytvořte kontingenční tabulku sloupcově a řádkově podmíněných relativních četností znaků X a Y.
   Zamyslete se nad odpověďmi na následující otázky:
   - Kolik procent studentů, kteří prospěli z angličtiny, neudělalo zkoušku z matematiky?
   - Jaký je podíl studentů, kteří neudělali zkoušku z angličtiny a neprospěli ani z matematiky?
     Kolik je to studentů?
   - Kolik procent studentů, kteří prospěli z matematiky, neudělalo zkoušku z angličtiny?
   - Jaký je podíl studentů, kteří neudělali zkoušku z matematiky a neprospěli ani z angličtiny?
     Kolik je to studentů?

Příklad č.2: Načtěte soubor ocel.txt.

1. Podle Sturgersova pravidla najděte optimální počet třídicích intervalů pro znaky *plasticita* a *pevnost* a vhodně stanovte meze třídicích intervalů pro každý znak. Dále určete středy těchto intervalů a příslušné variační řady.
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*).
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.

(b) Určete, zda mezi znaky *vlasy* a *oci* existuje nějaká závislot (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

Příklad č.2 Otevřete datový soubor `znamky.txt`.

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

(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.

Příklad č.3 Otevřete datový soubor ocel.txt.

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

(b) Vypočtěte a interpretujte 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

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:* Data k tomuto

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

příkladu lze nalézt v souboru odpovedi.txt. Vypočtěte průměr, rozptyl, šikmost a špičatost počtu správně zodpovězených otázek ve všech třech skupinách. Nakreslete sloupkové diagramy absolutních četností.

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?

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.

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.

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 $\vartheta = 0.5$.

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

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

Geometrické rozložení pravděpodobností:

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

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?

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?

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?

Diagnostické testy - Nepovinné

Příklad č.10: Provádělo se ověřování kvality nového testu pro diagnostikování jisté poruchy sluchu, 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?

Rovnoměrné rozložení

Příklad č.2: Na automatické lince se plní láhve 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 láhvi bude aspoň 1010 ml mléka?

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?

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?

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?

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?

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?

6 - Výpočet číselných charakteristik náhodných veličin pomocí softwaru 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 $F_{0.05}(2,10)$.

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.

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

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.

Příklad 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 $R(X_1, X_2)$.

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.

Příklad č.2: Přírůstky cen akcií v % na burze v New Yorku u 10-ti náhodně vybraných společností dosáhly těchto hodnot: $10, 16, 5, 10, 12, 8, 4, 6, 5, 4$. Odhadněte střední hodnotu μ a směrodatnou odchylku σ růstu cen akcií a dále odhadněte pravdě-podobnost růstu cen akcií aspoň o 8.5 %.

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.

Poznámka: Interpretace hodnot koeficientů korelace $|\rho|$:

(a) v přírodních vědách:

hodnota $\|\rho\|$	interpretace
$\langle 0\,; 0.4)$	žádná/téměř žádná závislost
$\langle 0.4\,; 0.6)$	slabá závislost
$\langle 0.6\,; 0.8)$	mírná závislost
$\langle 0.8\,; 1\rangle$	silná závislost

(b) v sociálních vědách:

Nesmíme zapomenout, že kromě míry závislosti můžeme pomocí koeficientu korelace určit, zda jde o závislost přímou (koef.korelace je kladný) nebo nepřímou (koef.korelace je záporný).

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

Najděte nestranný bodový odhad rozptylu hmotnosti, rozptylu výšky a kovariance hmotnosti a výšky. Vypočtěte rovněž realizaci výběrového koeficientu korelace hmotnosti a výšky.

hodnota $\|\rho\|$	interpretace
$\langle 0\,;\,0.15)$	žádná/téměř žádná závislost
$\langle 0.15\,;\,0.3)$	slabá závislost
$\langle 0.3\,;\,0.5)$	mírná závislost
$\langle 0.5\,;\,0.6)$	celkem silná závislost
$\langle 0.6\,;\,0.8)$	silná závislost
$\langle 0.8\,;\,1\rangle$	velmi silná závislost

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

Výslednou hodnotu koeficientu korelace interpretujte. Dále vytvořte histogramy pro hmotnost a výšku.

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.

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?

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

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.

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.

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.

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.

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.

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.

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 $N(\mu_2, \sigma_2^2)$.

(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.

(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 $\mu_1 - \mu_2$.

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.

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;

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

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.

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

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

(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** (Haladovu podmínku), jejíž splnění je nezbytné pro relevantnost závěru.

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 %.

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 $\alpha = 0.05$.

10 - Analýza rozptylu jednoduchého třídění

Příklad č.1: V jisté továrně se měřil čas, který potřeboval každý ze tří dělníků k uskutečnění téhož pracovního úkonu. Čas v minutách:

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

Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že výkony těchto tří dělníků jsou stejné. Zamítnete-li nulovou hypotézu, určete, výkony kterých dělníků se liší na dané 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**. 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?

Příklad č.2 Na střední škole byl uskutečněn experminet zjišťující efektvitu 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.

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.

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.

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.

Příklad č.3: Dvouvýběrový Wilcoxonův 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.

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:

1.typ:	50	35	43	30	62	52	43	57	33	70	64	58	53	65	39
2.typ:	31	37	59	67	44	49	54	62	34	42	40				
3.typ:	27	19	32	20	18	23									
4.typ:	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.

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.

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:

1.podnik:	420	560	600	490	550	570	340	480	510	460		
2.podnik:	400	420	580	470	470	500	520	530				
3.podnik:	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.

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. Výsledky zkoumání jsou uvedeny v následující tabulce a v souboru vlasy_oci.txt.

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.

Příklad č.2: Otevřete si soubor ped_hodnost.txt. 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. Data v souboru mají následující tvar:

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

Poznámka: Nezapomeňte před samotným testováním ověřit podmínky dobré aproximace.

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.

Příklad č.4: Podíl šancí Pro údaje z příkladu č.3 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.

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.

13 - Jednoduchá korelační analýza

Příklad č.1: Testování nezávislosti ordinálních veličin 12 různých softwarových firem nabízí speciální programové vybavení pro vedení účetnictví. Jednotlivé programy byly posouzeny odbornou komisí složenou z počítačových odborníků a komisí složenou z profesionálních účetních. Úkolem bylo doporučit vhodný program na základě stanovení pořadí jednotlivých programů. Výsledky posouzení:

Produkt firmy číslo	1	2	3	4	5	6	7	8	9	10	11	12
Pořadí dle odborníků	6	7	1	8	4	2.5	9	12	10	2.5	5	11
Pořadí dle účetních	4	5	2	10	6	1	7	11	8	3	12	9

Vypočtěte Spearmanův koeficient pořadové korelace a na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že hodnocení obou komisí jsou nezávislá.

Příklad č.2: Bylo sledováno 10 žáků. Na základě psychologického vyšetření byli tito žáci seřazeni podle nervové lability (čím byl žák labilnější, tím dostal vyšší pořadí R_i). Kromě toho sledování žáci dostali pořadí Q_i na základě svých výsledků v matematice (nejlepší žák v matematice dostal pořadí 1). Výsledky jsou uvedeny v tabulce:

Pořadí R_i	1	2	3	4	5	6	7	8	9	10
Pořadí Q_i	9	3	8	5	4	2	10	1	7	6

Vypočtěte Spearmanův koeficient pořadové korelace a na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že nervová labilita a výsledky v matematice jsou nezávislé.

Příklad č.3: Testování nezávislosti intervalových a poměrových veličin Zjišťovalo se, kolik mg kyseliny mléčné je ve 100 ml krve matek prvorodiček (veličina X) a u jejich novorozenců (veličina Y) těsně po porodu. Byly získány tyto výsledky:

Číslo matky	1	2	3	4	5	6
x_i	40	64	34	15	57	45
y_i	33	46	23	12	56	40

Nakreslete dvourozměrný tečkový diagram, vypočtěte výběrový korelační koeficient, sestrojte 95 % interval spolehlivosti pro korelační koeficient a na hladině významnosti $\alpha = 0.05$ testujte hypotézu o nezávislosti výsledků obou měření.

Příklad č.4: V náhodném výběru 10-ti dvoučlenných domácností byl zjišťován měsíční příjem (veličina X, v tisících Kč) a vydání za potraviny (veličina Y, v tisících Kč).

x_i	15	21	34	35	39	42	58	64	75	90
y_i	3	4.5	6.5	6	7	8	9	8	9.5	10.5

Vypočtěte výběrový koeficient korelace. Na hladině významnosti 0.05 testujte hypotézu o nezávislosti veličin X, Y. Sestrojte 95 % asymptotický interval spolehlivosti pro ρ. (Data jsou uložena v souboru prijem_vydani.sta).

Příklad č.5: Porovnání dvou korelačních koeficientů V psychologickém výzkumu bylo vyšetřeno 426 hochů a 430 dívek. Ve skupině hochů činil výběrový koeficient korelace mezi verbální a performační složkou IQ 0.6033, ve skupině dívek činil 0.5833. Za předpokladu dvourozměrné normality dat testujte na hladině významnosti $\alpha = 0.05$ hypotézu, že korelační koeficienty se neliší.

Příkald č.6: Načtěte datový soubor IQ.sta. Za předpokladu dvourozměrné normality dat (orientačně ověřte pomocí dvourozměrného tečkového diagramu) testujte na hladině významnosti $\alpha = 0.1$ hypotézu, že korelační koeficienty mezi verbální a performační složkou IQ jsou stejné u dětí z města a venkova.

14 - Porovnání empirického a teoretického rozložení

Příklad č.1: Ze souboru rodin s pěti dětmi bylo náhodně vybráno 84 rodin a byl zjišťován počet chlapců:

Počet chlapců	0	1	2	3	4	5
Počet rodin	3	10	22	31	14	4

Na asymptotické hladině významnosti $\alpha = 0.05$ testujte hypotézu, že rozložení počtu chlapců se řídí binomickým rozložením $Bi(5; 0.5)$.

Příklad č.2: Jsou známy počty občanů města Brna podle měsíce narození (stav k 31.12.2001).

měsíc narození	počet osob
leden	32309
únor	30126
březen	35010
duben	34761
květen	34955
červen	32883
červenec	33255
srpen	31604
září	31173
říjen	30536
listopad	28571
prosinec	29467
celkem	384650

Na asymptotické hladině významnosti $\alpha = 0.05$ ověřte hypotézu, že rozložení porodnosti je během roku rovnoměrné. Počty narozených lidí v jednotlivých měsících roku rovněž znázorněte graficky.

Příklad č.3: Firma, která vlastní několik supermarketů, se zajímá, zda zákazníci dávají přednost některému dnu v týdnu pro nákup. Náhodně bylo vybráno 300 zákazníků, kteří měli říci, který den v týdnu nejčastěji nakupují v supermarketu. Výsledky jsou uvedeny v následující tabulce:

Den	pondělí	úterý	středa	čtvrtek	pátek	sobota	neděle
Počet	10	20	40	40	80	60	50

Na asymptotické hladině významnosti $\alpha = 0.05$ testujte hypotézu, že žádný den v týdnu nemá při nakupování v supermarketu přednost před jinými dny.

Příklad č.4: Do rybníka bylo umístěno 5 pastí, přičemž každá past svítila jiným světlem (bílým, žlutým, modrým, zeleným, červeným). Do těchto pastí se chytilo 56, 72, 41, 53 a 38 jedinců. Na asymptotické hladině významnosti $\alpha = 0.05$ testujte hypotézu, že barva světla v pasti nemá vliv na počet chycených jedinců.