Dvouvýběrové testy
Cvičení 10

Úvodní nastavení.
- Stáhněte si data Iq.txt a zdrojové kódy cviceni10.R a figKS.R, případně i cviceni10-obrazky.R.
- Otevřete si program R Studio, změňte si pracovní adresář a vyčistěte pracoviště.
- Do proměnné alpha is uložte testovací hladinu 0.05.

Data. Data Iq.txt obsahují informace o náhodně vybraných žácích osmé třídy základní školy. Pro každého máme k dispozici jeho IQ, pohlaví (kódování 1 dívky a 0 chlapci) a dále pak průměrnou známku na vysvědčení v pololetí sedmé třídy a v pololetí osmé třídy. Upravte svoje data tak, aby odpovídali následujícímu popisu:

`IQ`	hodnota IQ,
`zn7`	průměrná známka na vysvědčení v pololetí sedmé třídy,
`zn8`	průměrná známka na vysvědčení v pololetí osmé třídy,
`pohlavi`	pohlaví žáka (1- žena, 0 - muž)
`FPohlavi`	jinak kódované pohlaví žáka (faktor kategorie chlapec, divka).

Načtení a úpravu dat provedete následovně:

```
Iq = read.table("Iq.txt",header=T)
colnames(Iq)[2] = "IQ"
levels(Iq$Pohlavi)
Iq$FPohlavi = factor(Iq$pohlavi, levels = c(0,1), labels=c("chlapec","divka"))
```

Dvouvýběrový problém

1. Ujistěte se, že se Vám data dobře načetla.

   ```
   summary(Iq)
   attach(Iq)
   ```

   Prohlédněte si základní charakteristiky polohy jednotlivých veličin pomocí funkce summary. Všimněte si rozličného výstupu pro veličiny pohlavi a FPohlavi.
   Příkaz attach nám zpřístupní jednotlivé proměnné z dat, takže je budeme moci volat přímo jejich názvem (např. IQ namísto Iq$IQ).

2. Pomocí vhodného obrázku vizualizujte genderové složení našeho datového vzorku.

3. Budeme se zabývat otázkou, zda se nějak liší IQ u chlapců a dívek. Začneme nejprve prohlídkou popisných statistik.

```
IQchlapci=IQ[FPohlavi=="chlapec"]
IQdivky=IQ[FPohlavi=="divka"]

# charakteristiky polohy dle pohlavi
summary(IQchlapci)
summary(IQdivky)

# nebo rychleji pomoci 1 prikazu
tapply(IQ,FPohlavi,summary)

# podminene sd
tapply(IQ,FPohlavi,sd)
```

Co usuzujete na základě těchto číselných charakteristik?

4. Problém si ještě budeme vizualizovat:

```
boxplot(IQ~FPohlavi)

# totez jako
boxplot(IQchlapci,IQdivky,names=c("chlapci","divky"))

# muzeme vybarvit a doplnit popisky:
boxplot(IQ~FPohlavi,ylab="IQ",col=c("lightblue","pink"))
```

Co usuzujeme z tohoto obrázku?

5. Nyní se podíváme na to, zda je možné (resp. vhodné) předpokládat, že oba výběry pocházejí z normálního rozdělení.

```
par(mfrow=c(1,2))
hist(IQchlapci,prob=T,xlim=c(60,150))
hist(IQdivky,prob=T,xlim=c(60,150))

par(mfrow=c(1,2))
qqnorm(IQchlapci,main="Chlapci")
qqline(IQchlapci)
qqnorm(IQdivky,main="Divky")
qqline(IQdivky)
par(mfrow=c(1,1))
```

Dvouvýběrový Kolmogorovův-Smirnovův test

6. Pomocí Kolmogorovova-Smirnovova testu budeme testovat, zda je rozdělení IQ chlapců i dívek stejné.
   (a) Co předpokládáme za model?
   (b) Jak zní nulová a alternativní hypotéza?
   (c) Provedeme test:
   `ks.test(IQdivky, IQchlapci, exact = FALSE)`

Jaký je náš závěr?

7. Testová statistika tohoto testu má tvar[1]

$$K_{n,m} = \sup_{x\in\mathbb{R}} |\hat{F}_X(x) - \hat{F}_Y(x)|.$$

Ověříme, že R počítá podle stejného předpisu:

```
e1=ecdf(IQchlapci)
e2=ecdf(IQdivky)
max(abs(e1(IQ)-e2(IQ)))
```

Kdybychom chtěli zjistit, kde maximum nastává:

```
IQ[which.max(abs(e1(IQ)-e2(IQ)))]
```

Celou situaci si můžeme i vizualizovat pomocí následujícího obrázku:

```
source("figKS.R")
figKS(IQ, FPohlavi)
```

Dvouvýběrový t-test

8. Pomocí dvouvýběrového t-testu budeme testovat shodu středních hodnot IQ u chlapců a dívek. Nejprve použijeme verzi pro shodné rozptyly.
   (a) Jaký předpokládáme model? Jsou tyto předpoklady pro naše data reálné?
   (b) Napište, jaké hypotézy testujeme a porovnejte je s hypotézami K-S testu.
   (c) Připomeňte si tvar testové statistiky tohoto testu:

$$T_{n,m} = \sqrt{\frac{nm}{n+m}}\frac{\overline{X}_n - \overline{Y}_m}{S_{n,m}}, \quad S^2_{n,m} = \frac{1}{n+m-2}\left((n-1)S^2_X + (m-1)S^2_Y\right).$$

   Jaké je její přesné rozdělení za nulové hypotézy (v předpokládaném modelu)?

   (d) Provedeme test:

```
t.test(IQ ~ FPohlavi, var.equal = TRUE)
# nebo jiny zpusob zadani:
t.test(IQchlapci,IQdivky, var.equal = TRUE)
```

   Jaký je náš závěr ohledně středních hodnot IQ? Co odhaduje uvedený interval spolehlivosti?

9. Spočítáme hodnotu testové statistiky manuálně:

[1] Za nulové hypotézy pro $n, m \to \infty$ platí $\sqrt{\frac{nm}{n+m}}K_{n,m} \xrightarrow{D} Z$, kde Z je náhodná veličina s distribuční funkcí $G(z) = \left(1 - 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2z^2}\right)\mathrm{I}[z>0]$.

```
xmean=mean(IQchlapci)
ymean=mean(IQdivky)
nx=length(IQchlapci)
ny=length(IQdivky)
S2=1/(nx+ny-2)*((nx-1)*var(IQchlapci)+(ny-1)*var(IQdivky))
(T.stat=(xmean-ymean)/(sqrt(S2*(1/nx+1/ny))))

# kriticka hodnota
qt(1-alpha/2,df=nx+ny-2)
#p-hodnota
2 * pt(-abs(T.stat), df = nx+ny-2)
```

10. Který z předpokladů testu lze do určité míry ignorovat a test zůstane v platnosti alespoň asymptoticky? Nesplnění kterého předpokladu může naopak vést k nedodržení předepsané hladiny testu?

11. Nyní si vyzkoušíme jinou verzi t-testu, která předpoklad shody rozptylů neuvažuje. Jde o tzv. Welchův test.

    (a) Provedeme test:
    ```
    t.test(IQ ~ FPohlavi)
    #nebo
    t.test(IQchlapci,IQdivky)
    ```
    Jaký je náš závěr? Jaké rozdělení má asymptoticky uvedená testová statistika? A na základě jakého rozdělení je spočtená p-hodnota? Co stačí předpokládat pro rozdělení IQ dívek a chlapců?
    Jak vidíme, tato verze t-testu je v R nastavena defaultně. Jelikož předem většinou nevíme, zda je možné předpokládat shodu rozptylů zkoumaných dvou výběrů, je lepší rovnou použít tuto obecnější verzi testu. V případě, že rozptyly ve skutečnosti shodné jsou, má tento obecnější test jen zanedbatelně nižší sílu ve srovnání s testem odvozeným pro výběry se shodným rozptylem, viz simulace ve 13.

    (b) Opět provedeme manuální výpočet testové statistiky
    ```
    (T.W=(ymean-xmean)/(sqrt(var(IQdivky)/nx+var(IQchlapci)/ny)))
    ```

12. V rámci procvičení zkuste otestovat domněnku, zda je rozdíl IQ dívek a chlapců vyšší než 5.

13. Na základě následujících simulací porovnáme sílu t-testu se shodou rozptylů a Welchova testu pro situaci, kdy shoda rozptylů platí

```
opak=1000
n1=100
n2=150
p.s=numeric(opak)
p.w=numeric(opak)
for(i in 1:opak){
  x=rnorm(n1,0.2,sd=1)
  y=rnorm(n2,0,sd=1)
  p.s[i]=t.test(x,y,var.equal=TRUE)$p.val
```

```
  p.w[i]=t.test(x,y)$p.val
  }
mean(p.s<=0.05)
mean(p.w<=0.05)
```

V předchozích simulacích proveďte následující změnu: simulujte x za nulové hypotézy, tj. nastavte střední hodnotu na 0, ale změňte směrodatnou odchylku na 3. Tím se podíváme na to, jaká je hladina testů za nulové hypotézy v případě neshody rozptylů. Měli bychom vidět, že varianta t-testu, která shodu předpokládá, nedodržuje předepsanou hladinu testu.

Celou situaci si můžeme i graficky znázornit pomocí první simulace ze souboru cviceni10-obrazky.R. Současné nastavení nám ukazuje porovnání síly při shodě rozptylů. Změňte nastavení tak, abyste viděli, jak situace vypadá, pokud (a) zvýšíme směrodatnou odchylku x, (b) zvýšíme směrodatnou odchylku y.

DVOUVÝBĚROVÝ WILCOXONŮV TEST (MANNŮV-WHITNEYŮV TEST)

14. Opět se vrátíme k původní otázce, a to, zda se liší IQ u chlapců a dívek. Nyní použijeme dvouvýběrový Wilcoxonův test.
    (a) Jaký model předpokládáme pro data?
    (b) Jak zní nulová a alternativní hypotéza?
    (c) Testová statistika má tvar
    $$W_{n,m} = \sum_{i=1}^{n} R_i,$$
    kde R_i je pořadí X_i ve sdruženém výběru $X_1, \dots, X_n, Y_1, \dots, Y_m$.
    Pro jaké její hodnoty budeme zamítat?
    (d) Provedeme asymptotickou verzi testu (bez korekce pro spojitost):
    `wilcox.test(IQ~FPohlavi, correct = FALSE)`
    Jaký je náš závěr?
    (e) Přesnou verzi testu bychom zavolali pomocí nastavení exact=TRUE v předchozí funkci. Zkuste.

15. Zkusíme manuální výpočet jednotlivých položek testu. Spočítáme testovou statistiku podle vzorce z přednášky:

```
r=rank(IQ)
r=rank(IQ)
(W1=sum(r[FPohlavi=="chlapec"]))
(W2=sum(r[FPohlavi=="divka"]))
```

    Dostali jsme stejný výsledek jako funkce wilcox.test?
    P-hodnotu Wilcoxonova testu bychom dostali pomocí aproximace normálním rozdělením, viz Samostatná práce (iii).

16. Vypočteme Mannovu-Whitneyovu statistiku, která má tvar

$$W^*_{n,m} = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{I}[X_i < Y_j]$$

a v případě, kdy máme v datech shody, tak

$$W^*_{n,m} = \sum_{i=1}^{n}\sum_{j=1}^{m} \left(\mathrm{I}[X_i < Y_j] + \frac{1}{2}\mathrm{I}[X_i = Y_j]\right).$$

Výpočet provedeme buď pomocí for cyklu (který je ale obecně v R mnohem pomalejší než maticové výpočty)

```
U1=0;U2=0
for(i in 1:nx)
  for(j in 1:ny){
    U1=U1+I(IQchlapci[i]<IQdivky[j])+1/2*I(IQchlapci[i]==IQdivky[j])
    U2=U2+I(IQchlapci[i]>IQdivky[j])+1/2*I(IQchlapci[i]==IQdivky[j])
  }
U1; U2
```

nebo pomocí matic:

```
x1 <- matrix(rep(IQchlapci, ny), ncol = ny)             # ve sloupcich IQchlapci
x2 <- matrix(rep(IQdivky, nx), ncol = ny, byrow = TRUE) # v radcich IQdivky
head(x1)
head(x2)
#
(U1 <- sum(x1 < x2) + 0.5 * sum(x1 == x2))
(U2 <- sum(x1 > x2) + 0.5 * sum(x1 == x2))
```

Vidíme teď nějakou shodu s výstupem funkce wilcox.test?

17. Mezi Wilcoxonovou statistikou a M-W statistikou je jednoznačný vztah (viz přednáška):

```
(nx * ny + 0.5 * nx * (nx + 1) - W1)        # U1
(nx * ny + 0.5 * ny * (ny + 1) - W2)        # U2
```

18. V následujících simulacích si porovnáme sílu dnes uvažovaných testů pro několik vybraných situací. Využijeme k tomu funkci simuluj, kterou jsme si načetli spolu s obrázkem K-S testu. Do této funkce zadáme rozsahy uvažovaných výběrů n a m a který z následujících scénářů chceme:

- normal-posun: $X_i \sim \mathsf{N}(0,1)$, $Y_i \sim \mathsf{N}(1,1)$,
- normal-hetero $X_i \sim \mathsf{N}(0,1)$, $Y_i \sim \mathsf{N}(0,3)$,
- normal-exp: $X_i \sim \mathsf{N}(0,1)$ a $Y_i = W_i - 1$, kde $W_i \sim \mathsf{Exp}(1)$,
- normal-exp2: $X_i \sim \mathsf{N}(0,1)$ a $Y_i = W_i - \log 2$, kde $W_i \sim \mathsf{Exp}(1)$.

```
simuluj(n=10, m=30, scenar="normal-posun")
simuluj(n=50, m=80, scenar="normal-posun")

simuluj(n=10, m=30, scenar="normal-hetero")
simuluj(n=100, m=300, scenar="normal-hetero")

simuluj(n=10, m=30, scenar="normal-exp")
simuluj(n=100, m=300, scenar="normal-exp")

simuluj(n=10, m=30, scenar="normal-exp2")
simuluj(n=100, m=300, scenar="normal-exp2")
```

Spusťte si jednotlivé příkazy postupně a zamyslete se, co testujeme a zda generujeme data za nulové hypotézy nebo za alternativy příslušného testu. Co testujeme u Wilcoxonova testu?

Samostatná práce

(i) Zjistěte, zda mají dívky v osmé třídě lepší známky než chlapci. Vyberte test vhodný pro naše data a interpretujte výsledek.

(ii) Zjistěte, zda je rozdíl ve známkách z předchozího bodu větší než 0.5.

(iii) V předchozí práci jsme zjistili, že funkce wilcox.test počítá Mannovu-Whitneyovu testovou statistiku a nikoliv Wilcoxonovu. Z přednášky ale víme, že Mannův-Whitneův test je ekvivalentní Wilcoxonově testu. Tudíž bychom měli dostat stejnou p-hodnotu. Spočítáme ji pomocí aproximace normálním rozdělením, kde musíme navíc přidat i korekci pro shody v datech:

```
nx=length(IQchlapci)
ny=length(IQdivky)
EW=nx*(nx+ny+1)/2
# varW.bez=nx*ny*(nx+ny+1)/12

# vypocet korekce
t=table(r)
kor=sum(t^3-t)/((nx+ny)*(nx+ny-1))
varW.kor=nx*ny*(nx+ny+1-kor)/12

U=(W1-EW)/sqrt(varW.kor)

2*pnorm(-abs(U))
# porovname:
wilcox.test(IQ~FPohlavi,correct=FALSE)$p.val
```

Dostali jsme nyní stejnou p-hodnotu jako funkce wilcox.test?

(iv) Předpokládejme, že máme dva nezávislé výběry $X_1, \ldots, X_n$ a $Y_1, \ldots, Y_n$ (oba výběry mají stejný počet pozorování), kde $X_i \sim \mathsf{N}(\mu_1, \sigma_1^2)$ a $Y_i \sim \mathsf{N}(\mu_2, \sigma_2^2)$. Chceme testovat shodu středních hodnot, tj. $H_0 : \mu_1 = \mu_2$. Co když namísto Welchova t-testu použijeme párový test? Tj. spočteme rozdíly $Z_i = X_i - Y_i$ a pak test založíme na testové statistice

$$T_n = \sqrt{n}\frac{\overline{Z}_n}{S_n^{(Z)}}, \qquad \text{kde} \qquad S_n^{2(Z)} = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \overline{Z}_n)^2$$

a nulovou hypotézu zamítneme, pokud $|T_n| > t_{n-1}(1 - \alpha/2)$.

(v) Rozdělení testové statistiky Welchova testu je aproximováno pomocí t-rozdělení s ν stupni volnosti, kde ν se určí na základě výběrových rozptylů a počtů pozorování v jednotlivých výběrech. Ověřte, zda R využívá stejný vzorec jako je uvedený ve skriptech v kapitole 6.2.

(vi) Zajímá nás, zda v osmé třídě dochází k zlepšení prospěchu oproti sedmé třídě. Jaký test je vhodný pro tuto situaci? Proveďte jej a interpretujte závěr.