

# Praktická zápočtová práce | Horses

Dátum: 18.12.2024

Méno a priezvisko študenta/študentky (*doplniť*)

Podkladový [Rmd súbor](#) pre vypracovanie pomocou knižnice Knitr (*UTF 8 kódovanie*)

## Návod

Vypracovanie zápočtovej práce môžete písať priamo do podkladového [Rmd súboru](#), ktorý na záver stačí v programe R Studio zkompilovať stlačením tlačítka "Knit HTML". Vytvorí sa výsledný HTML súbor s vašimi komentármi, zdrojovými kódmi aj R výstupmi. Tento HTML súbor po skončení práce odošlete emailom na adresu [maciak/hlavka\[AT\]karlin.mff.cuni.cz](mailto:maciak/hlavka[AT]karlin.mff.cuni.cz).

Je možné využiť aj iný formát výstupu (t.j., typ súboru, ktorý ako prílohu odošlete emailom), ale samotný zdrojový kód v textovom formáte nebude akceptovaný ako regulérny výstup.

Zdrojový kód pre program R vkladajte do bloku vyznačenom nasledujúcimi symbolmi:

```
```{r}

### zdrojovy kod programu R

```
```

Nasledujúce dva príkazy načítajú niektoré užitočné funkcie (R funkcie, ktoré sme bežne využívali na cvičení, ale nie sú štandardnou súčasťou programu R) a podkladové data.

```
rm(list = ls())

load(url("http://msekc.e.karlin.mff.cuni.cz/~maciak/NMFM301/zadanie2024horses.RData"))
load(url("http://msekc.e.karlin.mff.cuni.cz/~maciak/NMFM301/functions.RData"))
```

Data sú uložené v premennej `horses`.

Medzi príkazy a R výstupy doplňujte svoje zmysluplné odpovede a plnohodnotné komentáre k jednotlivým problémom a otázkam.

**Zápočtová práca obsahuje štyri úlohy a maximálne bodové ohodnotenie je 40 bodov. K úspešnému vypracovaniu zápočtovej práce je potrebné získať aspoň 20 bodov.**

## Data

K dispozícii je dátový súbor `horses`, ktorý predsavuje namerané hodnoty tepovej frekvencie (HR - heart rate) u niekoľika koní. Jedná sa o experiment, kde u 38 koní bola zameraná tepová frekvencia potom, ako bol na daného koňa aplikovaný konkrétny rušivý podnet. Rušivý podnet je v datach očíslovaný pomocou číslíc 1 až 7 a predpokladá sa, že jednotlivé impulzy sú postupne usporiadané vzhľadom k narastajúcemu rušivému efektu (t.j., intenzite hluku, ktorý daný impulz/podnet vyvolá). Okrem tepovej frekvencie a konkrétneho rušivého stimulu, ktorý bol na koňa aplikovaný, sú zaznamenané aj dodatočné popisné premenné (viď nižšie).

```
head(horses)
```

Celkovo má dátový súbor 38 nezávislých pozorovaní a zaznamenáva 11 premenných. Podrobný popis jednotlivých premenných je v nasledujúcej tabuľke.

| Název       |  | Význam |
|-------------|--|--------|
| gender      | pohlavie koňa (kobyľa, valach a hřebec)  |        |
| type        | typ koňa (teplokrevný, chladnokrevný, poník)   |        |
| utilization | využitie koňa (pretekársky, tréningový, rekreačný)   |        |
| outside     | čas, ktorý trávi koň vo výbehu, resp. v prírode: 1 = 1 až 3 hodiny denne; 2 = 4 až 9 hodín denne; 3 = viac ako 10 hodín;   |        |
| x1 až x7    | zaznamenaná tepová frekvencia (t.j., počet úderov za minútu) pro sedm stimulů (X1 = tichý zvuk; X2 = náhla hudba; X3 = rozmačkání plastové láhve; X4 = rozmačkání plechovky; X5 = použití spreje; X6 = hodení lopty; X7 = náhle otevření dáždníka) |        |

## Úloha 1

**Otestujte nulovú hypotézu, že strední tepová frekvencia valacha po náhlém otevření deštníku je 100 úderů za minútu. Doplníte vhodné popisné charakteristiky a obrázek.**

Zvolte pouze jeden test, který je pro daný problém podle vašeho názoru nejvhodnější. Svoju volbu náležite zdůvodnite. Do odpovědi napište, co je nulová hypotéza, alternativa, vypočtená hodnota testové statistiky a p-hodnota. Uveďte kritický obor testu. s správné interpretujte výsledek testu. Diskutujte, zdali jsou splněny předpoklady použitého testu.

```
x7_valach <- horses[X7[horses$gender == "valach"]]
par(mfrow = c(2, 1))
hist(x7_valach, main = "Histogram tepovej frekvencie valacha po otvorení dáždníka", xlab = "Tepová frekvencia")
boxplot(x7_valach, main = "Boxplot tepovej frekvencie valacha po otvorení dáždníka")
```

Pro začátek vybereme data x7\_valach z datového souboru a udeláme si základní vizualizaci. Budeme tedy testovat následující hypotézy: Nulová hypotéza - střední hodnota dat x7\_valach je 100 Alternativní hypotéza - střední hodnota dat x7\_valach je různá od 100 Budeme potřebovat použít jednovýberový test pro střední hodnotu. K tomu se nabízí použít jednovýberový t-test/z-test.

```
t.test(x7_valach, mu = 100)
```

Výsledek testu je následující:

One Sample t-test

data: x7\_valach t = -0.072945, df = 19, p-value = 0.9426 alternative hypothesis: true mean is not equal to 100 95 percent confidence interval: 83.66884 115.23116 sample estimates: mean of x 99.45

Výsledek testu není nijak významný, p-hodnota je 0.9426, což je výrazně větší než standardní hladina významnosti 0.05. Nemáme tedy důvod zamítnout nulovou hypotézu. Střední hodnota 99.45 je také velmi blízká 100.

Použití t-testu je podmíněno normalitou dat. Z vizualizace je vidět, že data nejsou úplně normální, na druhou stranu box plot vypadá poměrně symetricky. Navíc máme 32 pozorování, což je v praxi dostatečné pro validní výsledek asymptotického z-testu. Výsledek testu je navíc tak přesvědčivý, že můžeme prohlásit nulovou hypotézu za nezamítnutelnou.

Maximální bodové ohodnocení: 10 bodů

## Úloha 2

Je pravda, že tepová frekvence koní je stejná, když aplikujeme vzruchový stimul č.5 (t.j., použití spreje) a stimul č.4 (t.j., rozmačkání plechovky)? Spočítajte interval spoľahlivosti pre vhodnú charakteristiku rozdielu a tento interval interpretujte.

Zvolte pouze jeden test, který je pro daný problém podle vašeho názoru nejvhodnější. Svoji volbu náležitě zdůvodněte. Do odpovědi napište, co je nulová hypotéza, alternativa, vypočtená hodnota testové statistiky a p-hodnota. Uveďte kritický obor testu. s správně interpretujte výsledek testu. Diskutujte, zdali jsou splněny předpoklady použitého testu.

```
x5 <- horses$X5
x4 <- horses$X4
par(mfrow = c(2, 1))
hist(x5, col = rgb(1, 0, 0, 0.5), xlim = c(min(x5, x4), max(x5, x4)), main = "Histogram tepovej frekvencie", xlab = "Tepová frekvencia")
hist(x4, col = rgb(0, 0, 1, 0.5), add = TRUE)
boxplot(x5, x4,
        col = c(rgb(1, 0, 0, 0.5), rgb(0, 0, 1, 0.5)),
        names = c("Použitie spreju", "Rozmačkání plechovky"),
        main = "Boxplot tepovej frekvencie")
)
```

Opět začneme vizualizací dat. Sestavíme konfidenční interval pro rozdíl dat x5 a x4.

```
ci.t(x5 - x4)

hist(x5 - x4, main = "Histogram rozdílu tepovej frekvencie", xlab = "Rozdíl tepovej frekvencie")
boxplot(x5 - x4, main = "Boxplot rozdílu tepovej frekvencie")

qqnorm(x5 - x4)
qqline(x5 - x4)

> ci.t(x5-x4) CI.low Mean CI.up 2.329199 14.552632 26.776064
```

Je vidět, že konfidenční interval pro rozdíl je (2.33, 26.78). A střední hodnota rozdílu rozhodně není nulová.

Dal budeme testovat následující hypotézu: Nulová hypotéza - střední hodnoty dat x5 a x4 jsou stejné Alternativní hypotéza - střední hodnoty dat x5 a x4 jsou různé

Budeme potřebovat použít parový test. Z vizualizace můžeme soudit, že rozdíly dat jsou symetrické, nemají příliš odlehle hodnoty. Z vizualizace normality, můžeme spíše soudit, že data jsou oproti principu  $x=y$  rostoucí, tudíž data nejsou normální, ale není to dramatické. Raději se spolehneme na symetrii než normalitu a použijeme Wilcoxonův test.

```
wilcox.test(x5, x4, paired = TRUE, correct = FALSE)
```

Výsledek testu je:

Wilcoxon signed rank test

data: x5 and x4 V = 489.5, p-value = 0.004246 alternative hypothesis: true location shift is not equal to 0

Jelikož p-hodnota je radově menší než 0.05, test zamítá nulovou hypotézu s velkou rezervou. Můžeme tvrdit, že střední hodnoty tepu při použití spreje a rozmačkání plechovky nejsou stejné.

Maximální bodové ohodnocení: 10 bodů

## Úloha 3

Lze ze statistického pohledu říct, že tepová frekvence po šestém zvukovém stimulu (hození míče) je stejná pro valachy i hřebce?

Zvolte pouze jeden test, který je pro daný problém podle vašeho názoru nejvhodnější. Svoji volbu náležitě zdůvodněte. Do odpovědi napište, co je nulová hypotéza, alternativa, vypočtená hodnota testové statistiky a p-hodnota. Uveďte kritický obor testu. s správně interpretujte výsledek testu.

```
summary(horses$gender)
x6_valach <- horses$X6[horses$gender == "valach"]
x6_hrebec <- horses$X6[horses$gender == "hrebec"]
par(mfrow = c(3, 1))
length(x6_valach)
length(x6_hrebec)
hist(x6_valach, col = rgb(1, 0, 0, 0.5), main = "Histogram tepovej frekvencie valachu", xlab = "Tepová frekvencia")
hist(x6_hrebec, col = rgb(0, 0, 1, 0.5), main = "Histogram tepovej frekvencie hřebců", xlab = "Tepová frekvencia")
boxplot(x6_valach, x6_hrebec, col = c(rgb(1, 0, 0, 0.5), rgb(0, 0, 1, 0.5)), names = c("Valach", "Hřebec"), main = "Boxplot tepovej frekvencie")
```

Vybereme data o tepových frekvencích po hození míčem do proměnných x6\_valach a x6\_hrebec. A data vizualizujeme.

*Budeme testovat nasledujici hypotezy: Nulova hypoteza - stredni hodnoty dat x6\_valach a x6\_hrebec jsou stejne Alternativni hypoteza - stredni hodnoty dat x6\_valach a x6\_hrebec jsou ruzne*

*Budeme potrebovat pouzit dvouvzorkovy test. Z vizualizace muzeme soudit, ze data nejsou symetricka ani normalni. Tepova frekvence je naopak pomerne spojita velicina, zaroven ale nemame moc mereni, pro hrebce pouze 6. Pouzijeme Kolmogorovův-Smirnovův test, který je robustni vuci normalite dat.*

```
ks.test(x6_valach, x6_hrebec)
```

```
Exact two-sample Kolmogorov-Smirnov test
```

*data: x6\_valach and x6\_hrebec D = 0.26667, p-value = 0.7842 alternative hypothesis: two-sided*

*D hodnota není příliš vysoká, ale p hodnota testu je 0.7842, což je výrazně větší než standardní hladina významnosti 0.05. Nemáme tedy důvod zamítnout nulovou hypotézu. Můžeme prohlásit, že střední hodnoty tepu po hození míčem pro valachy a hrebce jsou stejné.*

*Maximálne bodové ohodnotenie: 10 bodov*

## Úloha 4

**Pomocou metody analýzy rozptylu (ANOVA) rozhodněte, či je rozdíl v tepové frekvenci po prvním zvukovém stimulu v závislosti na využití koně (pretekářský, tréninkový, rekreační).**

Uveďte vhodné popisné charakteristiky a doplňte ich vhodným obrázkem. Diskutujte předpoklady testu (není nutné statisticky overovat' a testovat' předpoklady, ale můžete nakreslit a okomentovat vhodné grafy). Výsledek testu podrobně interpretujte.

*Maximálne bodové ohodnotenie: 10 bodov*

```
### Tu môžete doplniť potrebný zdrojový kód pre program R
install.packages("pandoc")
rmarkdown::render("your_file.Rmd", output_format = "html_document")
```

*Tu môžete doplniť potrebné komentáre k Vášmu riešeniu*

**Vypracované riešenia (napr. zkompilovaný PDF nebo HTML súbor) odošlite ako prílohu emailom **najneskôr do konce cvičení** na adresu [hlavka\[AT\]karlin.mff.cuni.cz](mailto:hlavka[AT]karlin.mff.cuni.cz)**