

# Metody ve fyzické geografii II - klimatologie

## Cvičení č. 2 – Základy práce v prostředí R

---

**Zadání:** S využitím dodaného souboru s vybranými klimatickými charakteristikami procvičte základy práce v programovém prostředí R a základní úlohy popisné statistiky:

- základní orientace v prostředí R, základy ovládání, nastavení prostředí
- vybrané základní příkazy, instalace packages a nahrávání knihoven
- popisná statistika, základní testy, regrese a korelace
- základy práce s grafy (histogram, proložení rozdělání, časová řada)
- přehled typů proměnných a datových objektů
- základy skriptování

### 1. Vybrané základní příkazy – analýza jednorozměrných dat

**# nastavení pracovního adresáře (příklad)**

```
> setwd("c:/data/R_data/cviceni")
```

**# nahrát tabulku csv - průměrná teplotní řada Brna, 1800–2010**

```
> bt <- read.table("brno_t.csv", header=TRUE, sep=";")
> names(bt)
> summary(bt)
```

**# vybrané statistiky lednových teplot, šipka nahoru volá historii příkazů**

**# příkaz „attach“ raději nepoužívat**

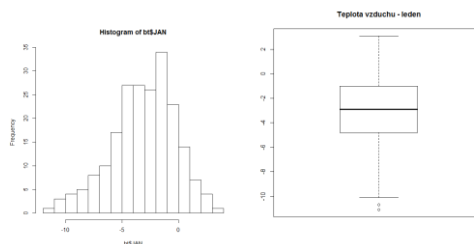
```
> mean(bt$JAN)
[1] -3.034123
> sd(bt$JAN)
[1] 2.884156
> quantile(bt$JAN, c(.25, .75))
 25%  75%
-4.8 -1.0
```

**# utřídění a uložení do nové proměnné**

```
> s_JAN <- sort(bt$JAN)
> s_JAN
```

**# vykreslit histogram a krabicový graf, další úpravy grafu - help(hist)**

```
> hist(bt$JAN)
> hist(bt$JAN, breaks=20)
> boxplot(bt$JAN, main="Teplota vzduchu - leden")
```



**# testování normality - test a vykreslení q-q grafu**

```
> shapiro.test(bt$JAN)
```

Shapiro-Wilk normality test

```
data: bt$JAN  
W = 0.98268, p-value = 0.01078
```

```
> qqnorm(bt$JAN)
```

## 2. Analýza dvourozměrných dat

```
# nahrát tabulku csv - měsíční hodnoty indexu NAO, 1825-2010
```

```
> nao <- read.table("NAOI.csv", header=TRUE, sep=";")
```

```
# upravit teplotní řadu Brna na stejné období jako je řada NAO
```

```
> bts <- subset(bt, YEAR > 1824)
```

```
# kontrola délky dvou proměnných (lednové hodnoty z obou souborů), které se budou porovnávat
```

```
> length(bts$JAN)
```

```
[1] 186
```

```
> length(nao$I)
```

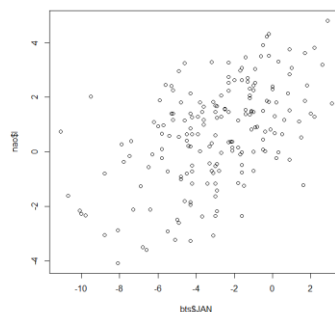
```
[1] 186
```

```
# vykreslení korelačního pole a výpočet korelačního koeficientu
```

```
> plot(bts$JAN, nao$I)
```

```
> cor(bts$JAN, nao$I)
```

```
[1] 0.511255
```



```
# je korelace statisticky významná?
```

```
> cor.test(bts$JAN, nao$I)
```

Pearson's product-moment correlation

```
data: bts$JAN and nao$I
```

```
t = 8.0693, df = 184, p-value = 8.882e-14
```

```
alternative hypothesis: true correlation is not equal to 0
```

```
95 percent confidence interval:
```

```
0.3965451 0.6102453
```

```
sample estimates:
```

```
cor
```

```
0.511255
```

```
# lineární regrese
```

```
> lm(bts$JAN ~ nao$I)
```

```
Call:
lm(formula = bts$JAN ~ nao$I)
```

```
Coefficients:
(Intercept)      nao$I
   -3.4349      0.7991
```

```
# lineární regrese - uložit vše pro další použití
```

```
> leden <- lm(bts$JAN ~ nao$I)
> summary(leden)
```

```
Call:
lm(formula = bts$JAN ~ nao$I)
```

```
Residuals:
    Min     1Q   Median     3Q     Max
-8.2645 -1.4667  0.0933  1.7220  6.0098
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.43488    0.19092  -17.991 < 2e-16 ***
nao$I        0.79914    0.09904   8.069 8.88e-14 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.45 on 184 degrees of freedom
Multiple R-squared:  0.2614,    Adjusted R-squared:  0.2574
F-statistic: 65.11 on 1 and 184 DF,  p-value: 8.884e-14
```

```
# vykreslit korelační pole a do grafu doplnit rovnici lineární regrese
```

```
> plot(bts$JAN, nao$I)
> abline(leden)
```

### 3. Vybrané testy

```
# f test - liší se významně variabilita hodnot indexu NAO v lednu a v červenci?
```

```
> var.test(nao$I, nao$VII)
```

```
F test to compare two variances
```

```
data: nao$I and nao$VII
F = 1.2737, num df = 185, denom df = 185, p-value = 0.1008
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.9539535 1.7006957
sample estimates:
ratio of variances
 1.273729
```

```
# t-test - porovnání dvou nezávislých souborů
```

```
> t.test(nao$I, nao$VII)
```

```
Welch Two Sample t-test
```

```
data: nao$I and nao$VII
```

```
t = 3.4795, df = 364.71, p-value = 0.0005631
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
0.2695746 0.9703179
```

```
sample estimates:
```

```
mean of x mean of y
```

```
0.65317204 0.03322581
```

```
# t-test pro párované hodnoty - porovnání dvou nezávislých souborů
```

```
> t.test(nao$I, nao$VII, paired=T)
```

```
Paired t-test
```

```
data: nao$I and nao$VII
```

```
t = 3.3629, df = 185, p-value = 0.0009373
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
0.2562511 0.9836414
```

```
sample estimates:
```

```
mean of the differences
```

```
0.6199462
```

#### 4. Práce s časovou řadou

```
# převede červencové teploty na časovou řadu a vypíše
```

```
> cervenec <- ts(bts$JUL, start=bts$YEAR[1])
```

```
> cervenec
```

```
# vykreslí graf časové řady a dodá nízkofrekvenční filtr
```

```
> plot.ts(cervenec)
```

```
> lines(ksmooth(bts$YEAR, bts$JUL, "normal", bandwidth = 12), col = "blue",  
lwd=3)
```

```
# vykreslí průběh autokorelační funkce
```

```
> acf(cervenec)
```