

# PROGRAM R

12.10.2012

---

## ÚVODNÍ NASTAVENÍ.

- Ve svém domovském adresáři si založte speciální adresář na tento předmět, např. **Statistika**.
- Z internetové stránky [www.karlin.mff.cuni.cz/~hudecova/education/](http://www.karlin.mff.cuni.cz/~hudecova/education/) si stáhněte datový soubor **wine1.csv** a uložte si jej do adresáře **Statistika**.
- Otevřete si program **R Studio**.

## PROGRAM R A POPISNÉ STATISTIKY

1. Společně si prohlédneme a vysvětlíme, jakým způsobem probíhá práce v R.
2. Otevřeme si nový skriptový soubor pomocí **File**→**New** →**R Script**. Sem budeme nadále psát jednotlivé příkazy. Výsledek se bude zobrazovat v dolním okénku **Console**.

Jednotlivé příkazy se píší na samostatné řádky, nebo je třeba oddělit je středníkem. Spuštění příkazu provedeme pomocí **CTRL+R** (spustí se tak řádem, na němž je aktuálně kurzor). Chceme-li spustit více příkazů najednou, označíme je myší a provedeme **CTRL+R**. Příkazy je také možné psát přímo na **Console** a spouštět **Enter**.

3. Použijte R jako kalkulačku a spočítejte následující výrazy:

$$1 + 1, \quad 3 - 2, \quad \frac{2}{3}, \quad 2 \cdot 3, \quad 3^2, \quad \sqrt{3}, \quad \log(10), \quad \exp(10), \quad \sin\left(\frac{\pi}{2}\right).$$

Nechte si vypsat nápovědu k funkci **log** tak, že zadáte **?log**. Stejným způsobem si lze zavolat nápovědu ke každé funkci.

4. Do vektoru **N** si uložíme měření obsahu dusíku v ovzduší ( $\mu\text{g}/\text{m}^3$ ) v oblasti slévárny

```
N=c(4.61, 10.53, 22.40, 16.34, 13.07, 18.31, 8.10, 9.29, 24.49, 11.90)
```

Jaké popisné statistiky by bylo vhodné uvést pro tato měření? Jaké obrázky?

5. Postupně použijte následující příkazy:

```
summary(N)
mean(N)
min(N)
max(N)
median(N)
quantile(N,0.25)
```

```
sd(N)
var(N)
quantile(N,3/4)-quantile(N,1/4)
```

6. Společně si vysvětlíme význam a interpretaci následujících obrázků:

```
plot(N)
boxplot(N)
hist(N)
```

```
hist(N,prob=T)
lines(density(N))
```

7. Do proměnné `N2` si uložíme hodnoty dusíku v miligramech.

```
N2=N/1000
```

Znovu se podíváme na jednotlivé popisné statistiky. Jak se změnilo?

8. Pomocí `Import Data Set → From text file..` si načtete data `wine1.csv`. Data sestávají ze 177 vzorků vín, na nichž bylo měřeno 11 proměnných (m.j. koncentrace alkoholu, kyseliny jablečné, popelovin, magnesia, fenolů, flavonoidů, dále intenzita a odstín barvy a odrůda).

9. Zkontrolujeme rozsah (dimenzi) dat: `dim(wine1)`

10. Příkaz `summary` si prohlédneme souhrnné statistiky všech proměnných. Všimneme si odlišného výstupu pro proměnnou `vintage`.

11. Příkaz `names(wine1)` nám vypíše názvy proměnných v datech.

12. K jednotlivých proměnným v souboru můžeme přistupovat pomocí výrazu `wine1$alcohol` atd, nebo si je zpřístupníme pomocí

```
attach(data)
```

Pak stačí volat `mean(alcohol)` atd.

13. Pro kontrolu spočítejte maximum, minimum, medián a průměr pro koncentraci alkoholu „ručně“ a srovnajte výsledek s výstupem funkce `summary`.

14. Nakreslete si krabicový graf koncentrace alkoholu a opět jej porovnejte s výstupem funkce `summary`.

15. Vykreslete si histogram koncentrace alkoholu. Lze považovat rozdělení alkoholu za normální?

16. Popíšeme zastoupení jednotlivých odrůd v datech:

```
table(vintage)
prop.table(table(vintage))
```

```
plot(vintage)
pie(table(vintage))
```

17. Jeden z dnešních obrázků si zkuste uložit (v libovolném formátu).

18. Uložte si také dnešní Script a Workspace.