Seznámení s programem R a popisné statistiky
19.2.2013

Úvodní nastavení.

- Ve svém domovském adresáři si založte speciální adresář Statistika na toto cvičení.
- Z internetové stránky `www.karlin.mff.cuni.cz/~hudecova/education/` si stáhněte datový soubor studenti.csv a uložte si jej do adresáře Statistika.
- Otevřete si program R (např. pomocí ikonky "R"na ploše nebo přes nabídku Start→...).
- Změňte si pracovní adresář pomocí File→Change working directory na Váš právě založený adresář statistika.

1. **R Commander.**
   (a) V R si spusťte R Commander pomocí Packages→Load Package, zde vyberte package Rcmdr. Mělo by se Vám otevřít nové okno s názvem R Commander.
   Jiný způsob: library(Rcmdr).
   (b) Společně si prohlédneme a vysvětlíme, jakým způsobem probíhá práce v R a R Commander .
   (c) 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 do Script Window a potvrdíte submit. Stejným způsobem si lze zavolat nápovědu ke každé funkci v R.
   (d) Do vektoru nazvaného N si uložíme pět měření obsahu dusíku v ovzduší ($\mu g/m^3$) v oblasti slévárny, které jsme naměřili:

```
N=c(10.53, 22.40, 16.34, 13.07, 18.31)
```

   Pomocí funkcí min, max, mean si spočítejte minimální, maximální a průměrný obsah dusíku, který jsme naměřili. Vysvětlíme si, jak se v R s vektory a jinými objekty pracuje.

2. **Data.**
   (a) Otevřte si data studenti.csv nejprve v Excelu a prohlédněte si jednotlivé proměnné a jejich hodnoty. Význam jednotlivých proměnných (sloupců) je uveden v tabulce 1.
   Poté soubor zavřete a vraťte se zpět do R.
   (b) Načtení dat: Pomocí Data→Import Data →from text file, clipboard or URL si načteme data studenti.csv. Do jednotlivých políček vyplňte:

Enter name for data set:	studenti
Variable names in file:	ano
Missing data indicator:	
Location of data file:	Local system file
Fields separeter:	Commas
Decimal point character:	Period [.]

   Poté vylistujte soubor studenti.csv a potvrďte.

ID	identifikační číslo studenta
Rok	rok přednášky
mesic.naroz	měsíc narozní
rok.naroz	rok narození
pohlavi	pohlaví (0 žena, 1 muž)
vyska	výška v cm
vaha	hmotnost v kg
boty	velikost bot
pocet.souroz	počet sourozenců
vek.otce	věk otce
vek.matky	věk matky
kraj	bydliště (kraj)

Tabulka 1: Datový soubor studenti.csv — popis proměnných

(c) Pomocí View data set se podívejte na načtená data a zkontrolujte, že se všechno načetlo tak, jak má.

3. **Veličiny měřené na různých měřítkách..**

(a) Rozmyslete si, na jakém měřítku jsou jednotlivé proměnné měřeny. Jakými charakteristikami a obrázky by bylo vhodné je popsat?

(b) Pomocí Statistics →Summaries →Active data set si nechte vypsat základní popisné statistiky všech veličin zahrnutých v datech.
Všimněte si rozdílu mezi tím, co R vypsalo pro výšku a kraje. Je popis veličiny udávající pohlaví smysluplný?

(c) Jsou-li některé veličiny kategoriální a jsou kódovány pomocí čísel, musíte R-ku sdělit, že je má chápat jako tzv. faktory. To provedeme pomocí
Data→Manage variables in active data set →Convert numeric variables to factors.
Proveďte toto pro veličinu pohlaví. Poté zopakujte bod (b) a podívejte se, co se změnilo.

4. **Popis kvalitativních (kategoriálních) veličin.**

(a) Pomocí Statistics →Summaries →Frequency distributions zjistěte, jaké bylo procentuální zastoupení mužů a žen na přednáškách v minulých letech.

(b) Předchozí výsledek si znázorníme také graficky: Pomocí Graphs →Pie chart si vykreslíme koláčový diagram a pomocí Graphs →Bar graph sloupcový graf.

(c) Podívejte se, které příkazy R volá v předchozích bodech.

(d) Uložte si jeden z předchozích obrázků do svého adresáře Statistika.

(e) Zjistěte, jaké bylo procentuální zastoupení Pražáků na přednáškách v minulých letech?

(f) Vykreslete si sloupcový graf, který znázorní, jak se měnil počet studentů přítomných na přednášce pro jednotlivé roky.

5. **Uložení práce.**

(a) Uložte si skript (příkazy do R, které se objevovaly v horní části R Commander) do svého adresáře pomocí File→Save script. Ukládat můžete i průběžně.

(b) Podobně si můžete uložit i output (výsledky v dolní části R Commander), ale to není nutné.

(c) Ukončete práci v R Commander .