# Experimental Design and Statistics - AGA46E

M. Maciak (Czech University of Life Sciences, Prague)

Lab Session 3 - Summer Term 2015

## 1 Prehled nekterych prikazu

- `read.table()` a `read.csv()` pro nacitani vlastniho datoveho suboru do Rka; Prvni prikaz je vseobecny pro libovolny datovy soubor, druhy prikaz (csv) slouzi specialne pro nacitani datovych souboru s koncovkou *csv* (doporuceny typ datoveho souboru);
- data je take mozne nacist primo z "clipboardu" pocitace a to pomocou kombinace tlacitek CTRL + C a prikazu:

```
read.table(file = "clipboard", sep="\t", header=TRUE)
```

- dodatecne parametre, ktere podrobneji specifikuju datovu soubor, ktery chceme nacist, jsou napr. `sep` pro definovani oddelovace sloupcu, `dec` pro definovani desetinnych cisel (symbol pro desetinou carku), `header` pro upresneni jesli data obsahuju hlavicku ci nikoli a pripadne parametr `skip`, ak je za potreby nacitavat data pozdeji nez hned v prvnim radku;
- funkce pro vykreslovani obazku: `plot()`, `boxplot()`, `barblot`, `pie()`, `dotchart()`, `hist()` - tyhle funkce zaroven oteviraji v Rku graficke okno, do ktereho sa postupne zakresluje obrazek;
- dodatecne graficke funkce v Rku: `lines()`, `abline()`, `segments()`, `points()`, `text()`, nebo `legend()` - tyhle funkce lze pouzit pouze v pripade, ze graficke okno je uz otevrene; Funkce slouzi k dokresleni dodatecnych objektu jiz do existujiciho obrazku.

## 2 Nahodne Generatory v Rku

- pouzite google a vyhledejte prikazy, ktore se v Rku pouzivaji ke generovani nahodnych cisel z Binomickeho, Poissonoveho, ci Geometrickeho rozdeleni;
- dodatecne parametre jsou nutne k spravnemu fungovani. Jake parametre to jsou a jak funguji? (vyzkousejte ruzne nastaveni pro dodatecne parametre a pomoci graficky nastroju - napr. funkce `hist()` zjistite, jaky vlyv maji na generovane hodnoty);
- vygenerujte dostatecne dlouhu posloupnost pro nejake rozdeleni a pomoci obrazku overte, jestli tvar distribucni funkce zodpoveda s distribucni funkci daneho rozdeleni. (pouzite napr. prikazy `hist()`, `ecdf()`, nebo `stepfun()`);
- zopakujte to same, ale zakazdym pro delsi posloupnost generovanu z daneho rozdeleni; (napr. pro rozsah nahodniho vyberu $n = 10, 20, 50, 100, 1000, 10000, \ldots$)

# 3 Diskretni nahodni veliciny a faktory v Rku

- nektere diskretni nahodni veliciny muzu byt take reprezentovany jako faktory - faktorove promene; Kazdy faktor ma urcity pocet levelu (hladin) a kazda hladina ma svoje pojmenovani - label; (napr. namisto hodnot 0 a 1 pro muze a zeny, muzeme stejne pouzit kodovani 'male' a 'female', nebo 'M' a 'F')
- pojmenovani faktoru a jejich prislusnych levelu by melo byt intuitivni a primocare;
- jaky je hlavni rozdil medzi diskretni interpretaci hodnot a faktorovou interpretaci hodnot? (napr. mezi hodnotami 0 a 1 a oznacenim 'M' a 'F', v obou pripadech pro rozlyseni muzu a zen)

```
> genderVector <- rbinom(20, 1, 0.5) # random vector of zeros and ones
> genderFactor <- as.factor(genderVector)
```

- porovnajte diskretni interpretaci `genderVector` a faktorovou interpretaci `genderFactor` a take `class(genderVector)` s `class(genderFactor)`;
- muzeme zmenit pojmenovani faktorovych hladin (labels) z numerickych znaku 0 a 1 na cokolvek jine, napr."male" a "female", nebo "M" and "F", atd. (povsimnete si, ze pak uz se nejedna o numericke hodnoty);

```
> levels(genderFactor) <- c("male", "female")
> genderFactor
```

- vyuzijte nahodne generatory v Rku a nagenerujte nejake sekvence nahodnych diskretnych hodnot; Nasledne z hodnot udelejte faktory s urcitym poctem levelu a nakonec tyto levely prejmenujte (pouzite nejake intuitivni nazvy prislusnych hladin);

  Napriklad:

```
> accidents <- rpois(31, 2)
> accidents[accidents >= 3] <- 3
> accidents[accidents >=1 & accidents <=2 ] <- 1
> accidents[accidents == 0] <- 0
> accidentsFactor <- factor(accidents, labels = c("None", "Minor", "> 3"))
```

- pouzijte datovy soubor Orange (k dispozici primo v instalaci Rka) a overte, ktere promenne jsou faktory a jake jsou jejich prislusne hladiny;

```
> Orange
> levels(Orange[,1])
```

- pomoci prikazu `levels()` zmente jmena prislusnych levelu;
- vytvorte vlastni datovy soubor (tabulku) napr. pomoci prikazu `matrix`, nebo `data.frame()`, ktory bude obsahovat alespon tri faktorove promene, kazdou s alespon dvema hladinami;
- proc je dulezite (az kriticke) spravne rozoznavat diskretni promenne a faktorove promenne?

# 4 Realne data s diskretnymi a faktorovymi promennymi

- stahnite a nactete datovy soubor passengerData.csv;
- vyuzijte zakladni popisne charakteristiky, abyste ste dozvedeli neco o datech;
  (jaka je jejich struktura, jake hodnoty nabyvaji a pod.)
- ktere promenne jsou faktory a ktere ne?
- vyuzijte graficke nastroje v Rku a udelejte nekolik obrazku;
- pomoci dodatecnym funkci udelejte obrazky hezci;
  (napr. dodatecne parametre `pch`, `xlab` a `ylab`, `main`, `col`, prikaz `legent()`, a dalsi)