

ANALÝZA DAT V R

Datové struktury (datová tabulka, seznam)

2 Datové struktury

2.5 Datová tabulka

- datová tabulka je tabulka dat, kde sloupce označují sledované znaky a řádky označují jednotlivá pozorování
- vhodná pro ukládání dat a jejich další zpracování

- `data.frame(sloupec1=data1, sloupec2=data2, ...)` vytvoří datovou tabulku s definovanými názvy sloupců (`sloupec1`, `sloupec2`, ...) a jejich hodnotami (`data1`, `data2`, ...)

```
> studenti <- data.frame(fakulta=c("FT","FP","FS","FS","FT"), vek=c(23,28,25,24,24))
```

	fakulta	vek
1	FT	23
2	FP	28
3	FS	25
4	FS	24
5	FT	24

- přidání argumentu `row.names=c("", "", ...)` udává názvy řádků, tj. zkoumaných jednotek

```
> studenti <- data.frame(fakulta=c("FT","FP","FS","FS","FT"), vek=c(23,28,25,24,24),  
row.names=c("Dana","Klara","Jan","Petr","Martina"))
```

	fakulta	vek
Dana	FT	23
Klara	FP	28
Jan	FS	25
Petr	FS	24
Martina	FT	24

- `names()`, kde v závorce je uveden název datové tabulky, vypíše názvy sloupců, tj. sledovaných znaků

```
> names(studenti)  
[1] "fakulta" "vek"
```

- chceme-li k datové tabulce přidat další sloupec, tj. další sledovaný znak, napíšeme za název datové tabulky znak dolaru \$ a název nového sloupce (vše bez mezer) a do něj uložíme vektor s příslušnými hodnotami

```
> studenti$rocnik <- c(2,3,3,1,2)
```

```
> studenti
```

	fakulta	vek	rocnik
Dana	FT	23	2
Klara	FP	28	3
Jan	FS	25	3
Petr	FS	24	1
Martina	FT	24	2

- chceme-li naopak sloupec z datové tabulky vymazat, přiřadíme do tohoto sloupce NULL

```
> studenti$vek <- NULL
```

```
> studenti
```

	fakulta	rocnik
Dana	FT	2
Klara	FP	3
Jan	FS	3
Petr	FS	1
Martina	FT	2

- na sloupce datové tabulky lze opět aplikovat vhodné funkce (pokud to dává smysl)

2.6 Seznam

- nejobecnější datová struktura
- sdružuje několik různých datových typů/struktur do jednoho velkého objektu, jednotlivé části se obvykle nazývají složky
- seznam je výstupem mnoha funkcí v R, např. histogramu

```
> histogram <- hist(1:10)
```

```
> histogram
```

Příklad 1

Následující tabulka shrnuje, kolik bodů dostali jednotliví studenti z 1. a 2. zápočtové písemky na předmětu Statistika v loňském akademickém roce.

Počet bodů z 1. pís.	6	8	7	4	5	7	2	3	8	4	6
Počet bodů z 2. pís.	5	5	3	9	1	9	2	0	7	5	6

- Definujte datovou tabulku odpovídající uvedené tabulce a nazvěte ji *body* (nezapomeňte na názvy sloupců, jednotlivé řádky nepojmenovávejte).
Uvědomte si, že v R je orientace tabulky jiná než zde! Nesnažte se proto nutně mít tabulku stejně orientovanou.
- Přidejte do tabulky sloupec, který udává součet bodů z obou písemek a vhodně jej pojmenujte.
- Určete průměrný počet bodů z 1., resp. 2. písemky.
- Určete rozptyl počtu bodů z 1., resp. 2. písemky.
- Sledované objekty, tj. studenty, vhodně pojmenujte, např. *Student 1*, *Student 2*, atd.
- Odstraňte z datové tabulky sloupec, který udává počet bodů z 2. písemky.