

Statistika: Opakování

V tomto dokumentu si zopakujeme základní typy objektů a jak se stahují k typu proměnných, které se ve statistice používají.

Prvním typem proměnné, který si ukážeme je **nominální** proměnná. Tento typ proměnné se vyznačuje tím, že jeho hodnoty nemůžeme seřadit (nemůžeme říci, jaká hodnota je větší než jiná) nebo nemůžeme říci o kolik je jedna hodnota větší než druhá. Příkladem takové proměnné je například pohlaví nebo národnost. V R je pak takováto proměnná typicky uložena jako objekt třídy **character**. V R sice můžeme takovýto objekt seřadit, ale toto seřazení je podle abecedy, ne podle toho jaká je hodnota je větší.

```
jmena <- c("Jan", "Monika", "Jakub", "Anna", "Lucie")
print(jmena)
```

```
## [1] "Jan"      "Monika" "Jakub"   "Anna"    "Lucie"
```

```
class(jmena)
```

```
## [1] "character"
```

```
sort(jmena)
```

```
## [1] "Anna"     "Jakub"    "Jan"      "Lucie"    "Monika"
```

Dalším typem proměnné, který si představíme je **ordinální** proměnná. U ordinální proměnné (na rozdíl od nominální) můžeme hodnoty seřadit, stále ale nemůžeme říci, o kolik je jedna hodnota větší než druhá. Mezi ordinální proměnné patří například vzdělání (základní/střední/vysokoškolské) nebo odpovědi na škály (spokojený/nespokojený). V R ordinální proměnné vytvoříme pomocí funkce **factor**. V této funkci specifikujete pořadí hodnot v argumentu **levels**.

```
skala <- factor(x = c("Velmi spokojeny", "Spokojeny", "Nespokojeny", "Velmi nespokojeny",
                      "Spokojeny", "Velmi nespokojeny", "Velmi spokojeny", "Nespokojeny"),
               levels = c("Velmi spokojeny", "Spokojeny",
                           "Nespokojeny", "Velmi nespokojeny"))
print(skala)
```

```
## [1] Velmi spokojeny   Spokojeny          Nespokojeny        Velmi nespokojeny
```

```
## [5] Spokojeny         Velmi nespokojeny  Velmi spokojeny     Nespokojeny
```

```
## Levels: Velmi spokojeny Spokojeny Nespokojeny Velmi nespokojeny
```

```
class(skala)
```

```
## [1] "factor"
```

```
sort(skala)
```

```
## [1] Velmi spokojeny   Velmi spokojeny    Spokojeny          Spokojeny
```

```
## [5] Nespokojeny       Nespokojeny        Velmi nespokojeny  Velmi nespokojeny
```

```
## Levels: Velmi spokojeny Spokojeny Nespokojeny Velmi nespokojeny
```

Přestože vektor třídy **factor** vypadá jako textový vektor, ve skutečnosti je blíže číselnému vektoru (**numeric**). Aby R vědělo, jak má hodnoty seřadit, jsou jednotlivé hodnoty reprezentovány čísly. Na tuto číselnou reprezentaci se můžete podívat pokud převedete **factor** do číselné (**numeric**) proměnné.

```
as.numeric(skala)
```

```
## [1] 1 2 3 4 2 4 1 3
```

Poslední proměnnou, kterou si představíme **kardinální**. U tohoto typu proměnné můžeme určit jak pořadí, tak i to, o kolik je jedna hodnota větší než druhá. Typicky jsou kardinální proměnné číselné proměnné. Příkladem kardinální proměnné je například výška. V R jsou kardinální proměnné typicky reprezentovány objektem třídy `numeric`.

```
vyska <- c(165, 186, 176, 182, 191)
print(vyska)
```

```
## [1] 165 186 176 182 191
```

```
class(vyska)
```

```
## [1] "numeric"
```

```
sort(vyska, decreasing = TRUE)
```

```
## [1] 191 186 182 176 165
```

```
vyska[1] - vyska[1:4]
```

```
## [1] 0 -21 -11 -17
```

Shrnutí:

Typ	Můžeme seřadit	O kolik větší
Nominální		
Ordinální	x	
Kardinální	x	x

Nakonec si připomeneme, že pokud chceme proměnné jednotlivých tříd vložit do matice (řádky x sloupce), máme dvě možnosti: objekt `matrix` nebo `data.frame`. `matrix` se vyznačuje tím, že všechny proměnné (sloupce) musejí mít stejnou třídu. U objektu `data.frame` můžeme použít různé typy proměnných jako sloupce.

```
cbind(jmena, vyska)
```

```
##      jmena      vyska
## [1,] "Jan"      "165"
## [2,] "Monika"   "186"
## [3,] "Jakub"    "176"
## [4,] "Anna"     "182"
## [5,] "Lucie"    "191"
```

```
data.frame(jmena = jmena, spokojenost = skala[1:5], vyska = vyska)
```

```
##      jmena      spokojenost      vyska
## 1   Jan      Velmi spokojeny      165
## 2 Monika      Spokojeny      186
## 3  Jakub      Nespokojeny      176
## 4  Anna Velmi nespokojeny      182
## 5  Lucie      Spokojeny      191
```