

# Klasická statistika

- výběrový průměr, výběrový rozptyl, výběrová měřidlová odchylka
- klasický model lineární regrese
- t-test, ANOVA
- Spearmanův korelační koeficient
- ⋮

# Statistické modely

$X = (X_1, \dots, X_n)'$  ... vektor pozorování; typicky  $X_1, \dots, X_n$  tvoří náhodný vektor

• parametrický model -  $X$  má sdruženou distribuční funkci  $F(x, \theta)$ ,  $F$  známe až na hodnotu neznámého parametru  $\theta \in \mathbb{R}^p$ .

výhody: větší přehled, jednoduchost, lepší interpretace, výpočty méně náročné

nevýhody: citlivé na porušení předpokladů modelu

• neparametrický model - nepředpokládá žádný specifický tvar rozdělení vektoru  $X$ ; neznámý parametr je nekonečněrozměrný

- histogram, jádrový odhad hustoty, neparametrická regrese

- pořadí, pořádkové testy, pořádková statistika

• semiparametrický model - má parametrickou a neparametrickou složku  $\rightarrow$  konečné i nekonečně rozměrným parametrem

- model Coxova proporcionálního rizika

$T = \text{čas do události}$   $F(t) = 1 - \exp \left\{ - \int_0^t \lambda_0(u) e^{\beta'x} du \right\}$

$x$  - vektor kovariát  
 $\beta$  - měrný parametr  
 $\lambda_0(u)$  - základní riziková funkce (neznámý parametr)

## Robustní postupy

- nejmenší cílíme na porušení předpokladů modelu (normalita, odlehká pozorování, ...), ale zachovávat si efektivitu, pokud předpoklady porušené nejsou.

## Odlehká pozorování (outliers)

Proč je hned neodstráníme R dal?

- jsme líní a neodhalíme je
- nejme líní, ale neodhalíme je (ve větších dimenzích)
- máme málo pozorování (málo informace)
- odstraněním podhodnotíme odhad rozptylu

"Vědci věří v normální rozdělení chyb. Experimentátoři, protože je pořádají za matematický teorém,  
a matematická, protože je pořádají za experimentální fakt."

(Poincaré, 1912)