

14. Regrese



Lineární regrese
Neparametrické varianty
Nelineární regrese?

Fisherův exaktní (přímý) test



Typ regrese	Počet nezávislých proměnných	Počet závislých proměnných	Předpoklady	Funkce
Jednoduchá lineární regrese (MNČ)	1	1	Gauss-Markovova věta	lineární
Vícenásobná lineární regrese	n	1	-	lineární
Neparametrické metody (Theil-Sen, Siegel Repeated Medians)	1	1	Gauss-Markovova věta	lineární
?	n	1	-	lineární
Zobecněné lineární modely (GLM)	n	1	Gauss-Markovova věta po transformaci	logistická, logaritmická
Zobecněné aditivní modely (GAM)	n	1		
Lasso regrese, Ridge regrese				
Bayesovské metody				

Jednoduchá lineární regrese



- V případě existence vzájemného vztahu (korelace) lze tento vztah podrobněji popsat.
- Cíl regresní analýzy: popsat závislost hodnot proměnné Y na hodnotách proměnné X.
- V případě lineární regrese je tento popis dán lineárním modelem tvaru $y = ax + b$.
- Existují i techniky nelineární regrese.
- Nemáme-li dostatek informací k teoretickému souboru, snažíme se odhadnout typ funkce pomocí dvourozměrného diagramu.

Předpoklady lineární regrese

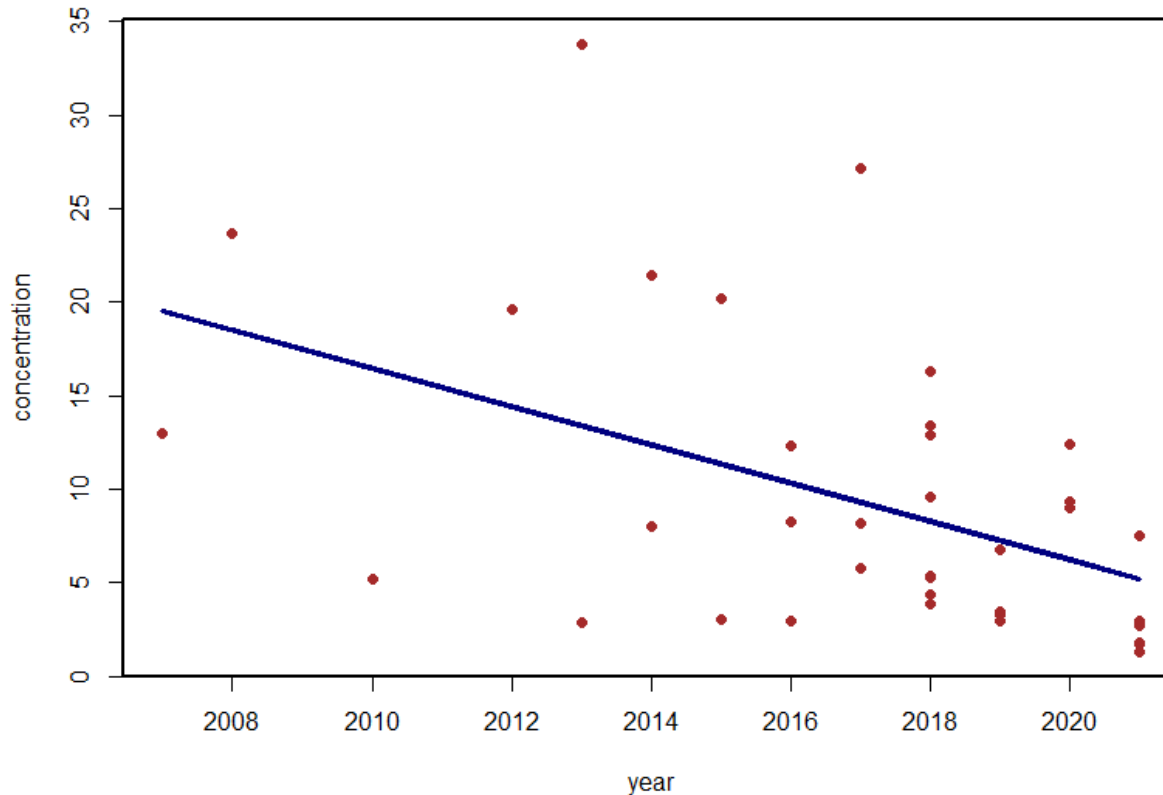


- Hlavním předpokladem je splnění požadavků Gauss-Markovovy věty:
 1. $E(\varepsilon_i) = 0,$
 2. $V(\varepsilon_i) = \sigma^2 < \infty,$
 3. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$
- Splnění těchto předpokladů je zajištěno v případě, kdy jsou rezidua normálně rozdělena, nezávislá na prediktorech (které jsou nezávislé).

Jednoduchá lineární regrese



$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \hat{Y} = \beta_0 + \beta_1 X \quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$



- Varianta bez interceptu (abs. členu) $Y = \beta_0 + \beta_1 X + \varepsilon$

Vyjádření rizik ve čtyřpolní tabulce - motivace



- Sledujeme souvislost věku matky a výskytu náhlého úmrtí kojence (SIDS). Výsledky dány v tabulce.
- Pomocí Pearsonova chí-kvadrátu nebo Fisherova exaktního testu můžeme rozhodovat o závislosti/nezávislosti dvou sledovaných veličin. Testy ale neumožňují tento vztah kvantifikovat.
- Má-li to smysl a chceme-li kvantifikovat (rozhodovat o těsnosti této závislosti) můžeme použít tzv. relativní riziko a poměr šancí.

SIDS	Věk matky		Celkem
	Do 25 let	25 a více let	
Ano	29	15	44
Ne	7301	11241	18542
Celkem	7330	11256	18586