

**Bi7491 Regresní modelování**

---

# Úvod

# Úvod

---

## Obecné informace

# Cíle předmětu

Na konci tohoto kurzu bude student schopen:

- definovat různé typy regresních modelů,
- navrhnout a vytvořit regresní model vhodný pro řešení zadaného problému,
- posoudit kvalitu sestaveného modelu,
- interpretovat výsledky regresní analýzy.

# Předběžný plán předmětu

- Úvod, Motivace, Opakování (21.2.2019)
- Regrese s normálním modelem I (7.3.2019)
- Regrese s normálním modelem II (14.3.2019)
- Případová studie – normální lineární model (21.3.2019)
- Logistická regrese (28.3.2019)
- **Písemka (zřejmě 4.4.2019)**; Jiné zobecněné lineární modely
- Případová studie – logistický regresní model (18.4.2019)
- Smíšený model a longitudinální modely (25.4.2019)
- Prediktivní modelování, prostor pro **vypracování projektu**/konzultace, předtermín
  
- zpravidla 2 hodiny přednáška + 1 hodina cvičení (počítače, balík R)
- konzultace dle potřeby po předchozí domluvě

# Požadavky na ukončení

- docházka na cvičení (lze 2 absence)
- písemka v průběhu semestru
- vypracování praktického úkolu
  - zvolíte si datový soubor a podrobíte jej regresní analýze
- zkoušková písemka
- ústní část zkoušky
  - diskuse nad praktickým úkolem a písemnou částí
- aktivita v hodinách – PTEJTE SE 😊

# Literatura

- Andersen P. K., Skovgaard, L.T. (2010) Regression with Linear Predictors, Springer.
- Harrel F. E., Jr. (2001): Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression and Survival Analysis. Springer, Springer Series in Statistics, New York.
- Forbelská, M: Materiály k výuce lineárních statistických modelů a zobecněných lineárních modelů
- Dalgaard, P (2008) Introductory statistics with R, sec. ed., Springer.
- Vittinghoff, E., Glidden, DV., Shiboski, SC., McCulloch, CE. (2005) Regression methods in biostatistics, Springer.
  
- Šmilauer, P: Moderní regresní metody  
<http://regent.jcu.cz/MRM.pdf>
- Pekár, S., Brabec, M. (2009) Moderní analýza biologických dat -- Zobecněné lineární modely v prostředí R

# Elektronické texty

- **Biostatistika pro matematickou biologii**
- <http://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickyh-a-biologickyh-dat--biostatistika-pro-matematickou-biologii>
- **Regresní modelování**
- <http://portal.matematickabiologie.cz/index.php?pg=analiza-a-hodnoceni-biologickyh-dat--regresni-modelovani>
- **Statistické modelování**
- <http://portal.matematickabiologie.cz/index.php?pg=analiza-a-hodnoceni-biologickyh-dat--statisticke-modelovani>

# Úvod

---

**Čím se budeme zabývat  
a k čemu to bude?**



# Čím se budeme zabývat

- **(Víceprediktorové) regresní modely**
- Hledání vztahu mezi několika prediktory (vysvětlujícími proměnnými) a nějakým výsledkem (vysvětlovanou proměnnou)

# Účel regresního modelování

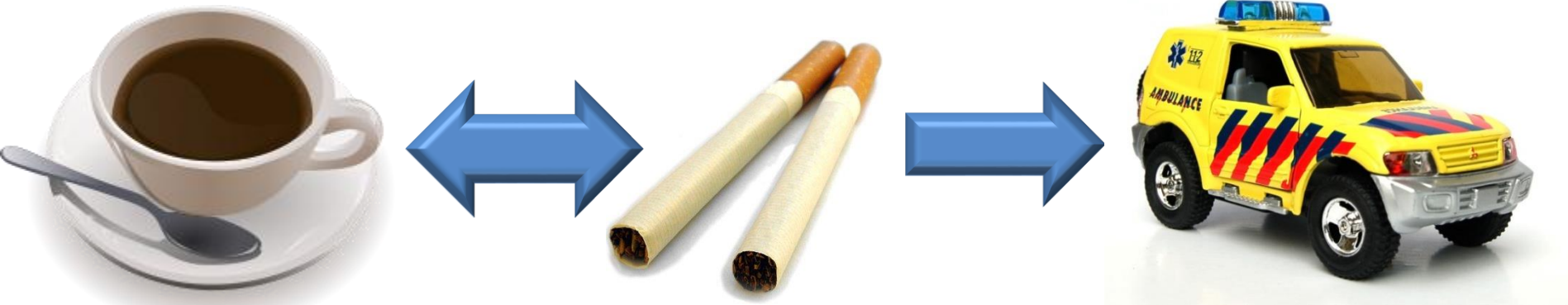
1. Zjištění „čistého“ vlivu konkrétního prediktoru na výsledek  
Protože ostatní proměnné nás mohou mást (**zavádějící faktory**)
2. Pochopení vztahu mezi více vysvětlujícími proměnnými a nějakým výsledkem  
Vliv proměnných se pro různé subjekty může lišit (**interakce**)
3. Předpověď výsledku pro nově pozorované subjekty

# 1. Zjištění „čistého“ vlivu



**Konzumace kávy se zdá být spojena s vyšším rizikem ischemické choroby srdeční**

# 1. Zjištění „čistého“ vlivu

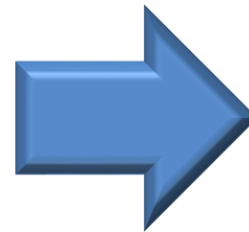


**Důvodem je však spojení konzumace kávy s kouřením, které má vliv na riziko ischemické choroby srdeční. Konzumace kávy samotná proti chorobě spíše chrání.**

# 2. Vztahy mezi prediktory

Interakce - modifikace vlivu

Mladší:

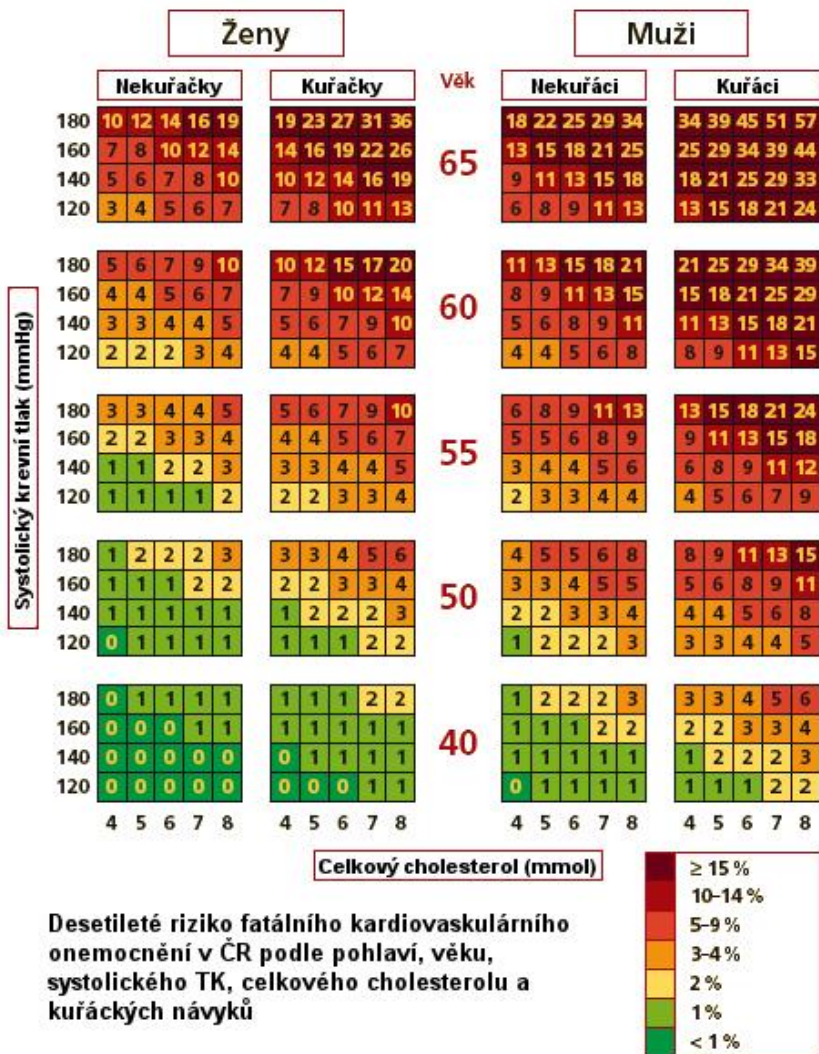


Starší:



Slábnutí negativního vlivu u starších osob (kouření na srdeční choroby)

# 3. Předpověď výsledku



Tabulka rizika úmrtí v důsledku srdečně-čevního onemocnění v následujících 10 letech (pro českou populaci)

**věk + pohlaví + kouření +  
systolický krevní tlak + celkový  
cholesterol → riziko úmrtí**

Zdroj: <http://www.szu.cz/tema/prevence/score>

# Diskuse



**Lze zadání vašich diplomových projektů  
formulovat jako regresní problém?  
(prediktory, výsledek, zkreslující proměnné, interakce)**