

Práce s chybějícími pozorováními v sociologických datech

Cílem tohoto dvoudenního semináře je představit problematiku chybějících dat v kvantitativních sociologických šetřeních a diskutovat vhodnost jednotlivých metod řešení tohoto všudypřítomného jevu. Největší prostor bude z metod věnován mnohonásobným imputacím chybějících hodnot. Oba dny semináře se skládají ze dvou kratších přednášek a dvou delších praktických cvičení v prostředí SPSS (s využitím speciálního modulu Missing Values 25). V rámci praktických cvičení budou demonstrovány analýzy na českých datech ze šetření European Social Survey (ESS). Seminář je určen pro studenty (zejména Mgr. a Ph.D.), akademické pracovníky a výzkumníky, kteří se chtějí detailněji seznámit s problémy spojenými s chybějícími daty a možnostmi řešení tohoto problému.

- Mechanismy chybějících hodnot, jednoduché metody práce s chybějícími hodnotami (listwise/pairwise deletion, nahrazování průměrem a regresně)
- Mnohonásobné imputace chybějících hodnot, nepoužívanější algoritmy mnohonásobných imputací, Rubinova pravidla slučování výsledků
- Realizace reálných analýz (t-testy, korelační koeficienty, lineární regrese) na mnohonásobně imputovaných datech
- Výběr proměnných do modelu imputace, odvozené proměnné, interakce mezi proměnnými, praktická doporučení, workflow práce s chybějícími pozorováními
- EM algoritmus, možnosti řešení problému v situacích, když data nechybí náhodně (NMAR), možnosti práce s chybějícími daty v jiných statistických programech (zejména R, STATA)

Doporučená literatura:

Baraldi & Enders (2010). An introduction to modern missing data analyses. *Journal of School Psychology* 48: 5–37.

Enders (2010). *Applied Missing Data Analysis*. New York: Guilford Press.

IBM Software Group. 2017. *IBM SPSS Missing Values 25*. Chicago: IBM Corporation.

Petrůšek (2015). *Analýza chybějících hodnot*. Praha: Sociologický ústav AV ČR.

Schafer (1999). *Multiple imputation: A primer*. *Statistical methods in medical research* 8(1): 3–15.

Cílová skupina: zejména Ph.D. studenti a akademičtí pracovníci z oblasti sociálních věd

Předpokládaný termín: čtvrtek a pátek 10.-11. 6. 2021, čas: 9-16, **online**

Předpokládané znalosti: základní statistické postupy – popisná statistika, korelace, lineární regrese

Cena: 2000 Kč

Akci spoluorganizuje Institut sociologických studií FSV UK a Česká sociologická společnost.

Lektor: Ivan Petrášek