

Analýza chybějících hodnot: Srovnání metod při zkoumání determinantů politické znalosti a příjmu

Elektronická příloha obsahuje skript sloužící k reprodukovatelnosti a kontrole simulačních studií realizovaných v knize. Skript je dostupný vždy ve dvou formátech: **R** (RStudio script) a **txt** (textový dokument). Celková délka skriptu (včetně komentářů a vynechaných oddělovacích řádků) je 9 595 řádků. Pro jednodušší orientaci byl skript rozdělen do tří složek. První složka obsahuje kompletní skript k první simulační studii (které je v knize věnována pátá kapitola). Druhá složka obsahuje kompletní skript k druhé simulační studii (které v knize patří šestá kapitola). Třetí složka obsahuje skript pro konstrukci grafů a schémat vytvořených na základě obou simulačních studií. Jedná se o grafy 2.1; 3.1; 5.1; 5.2; 5.3 a schéma 1.2.

Skript je ve složkách dále rozdělen do několika postupně očíslovaných dokumentů (s příponou .R a .txt). Každý z těchto skriptů obsahuje komentáře, které popisují a vysvětlují jednotlivé části kódu. Komentáře jsou psány bez diakritiky. Všechny komentáře začínají mřížkou (#) a ve skriptu z RStudia (dokumenty s příponou .R) jsou automaticky zvýrazněny zelenou barvou.

Celkovou strukturu příloh znázorňuje následující schéma. Tři zmíněné složky jsou znázorněny černou odrážkou a tučným písmem. Samotné skripty jsou znázorněny prázdnou odrážkou a příponou .R. Číslo na konci každého řádku označuje počet řádků, které daný skript obsahuje:

• 1 První simulační studie	
○ 1 Generování kompletních dat.R	75
○ 2 Generování chybějících hodnot podle 3 mechanismu.R	241
○ 3a MCAR analýzy.R	919
○ 3b MAR analýzy.R	918
○ 3c NMAR analýzy.R	916
○ 4a Standardizované vychylení.R	286
○ 4b RMSE.R	316
○ 4c Interval spolehlivosti.R	303
○ 5a Smerodatné chyby - mechanismus MCAR.R	69
○ 5b Smerodatné chyby - mechanismus MAR.R	69
• 2 Druhá simulační studie	
○ 1 Generování kompletních dat.R	183
○ 2 Generování chybějících hodnot podle 3 mechanismu.R	669
○ 3a MCAR analýzy.R	752
○ 3b MAR analýzy.R	739
○ 3c NMAR analýzy.R	767
○ 4a Standardizované vychylení.R	427
○ 4b RMSE.R	361
○ 4c Interval spolehlivosti.R	1223
• 3 Grafy a schéma	
○ 1 Všechny grafy a schéma.R	362

Skripty musí být spouštěny ve výše uvedeném pořadí. Jinými slovy, u každé simulační studie je vždy nutné nejdříve vytvořit kompletní data (spuštěním skriptu *1 Generování kompletních*

dat.R), a až potom vytvářet chybějící hodnoty podle jednotlivých mechanismů (spuštěním skriptu *2 Generovani chybejicich hodnot podle 3 mechnismu.R*) atd.

Technické specifikace:

- Verze R používané při realizaci simulačních studií: **3.1.0 (2014-04-10) -- "Spring Dance"** a **3.1.2 (2014-10-31) -- "Pumpkin Helmet"**
- Verze RStudio používaná při realizaci simulačních studií: **0.98.507**
- Počítač: **Dell Latitude E6330**
- Procesor: **Intel(R) Core(TM) i7-3540M CPU @ 3.00GHz**
- Nainstalovaná paměť: **8,00 GB**
- Operační systém: **Windows 7 Professional (64bitový operační systém)**

Všechny výpočty v rámci **první simulační studie** byly realizovány **v červenci 2014** (verze R: 3.1.0 - "Spring Dance"). **Druhá simulační studie** byla realizována **v prosinci 2014** (verze R: 3.1.2 - "Pumpkin Helmet").

Pracovní adresáře byly při provádění simulačních studií vytvořeny na ploše. Například u první simulační studie byl vytvořen adresář: *C:/Users/ivan.petrusek/Desktop/Prvni simulacni studie*. Skript tedy vždy ukládá průběžné soubory do adresáře, jehož vlastníkem je *ivan.petrusek*. Při reprodukování je tedy nutné v adresáři změnit uživatele (*C:/Users/UŽIVATEL/Desktop/Prvni simulacni studie*) nebo vytvořit vlastní pracovní adresář (a ten vždy změnit ve skriptu).

Kromě tzv. core packages byly při provádění výpočtů používány následující balíky (seřazeny abecedně):

- **BinNor** (*Simultaneous generation of multivariate binary and normal variates*) - Anup Amatya a Hakan Demirtas
- **hydroGOF** (*Goodness-of-fit functions for comparison of simulated and observed hydrological time series*) - Mauricio Zambrano Bigiarini
- **mice** (*Multivariate Imputation by Chained Equations*) - Stef van Buuren
- **mvtnorm** (*Multivariate Normal and t Distributions*) - Torsten Hothorn (et al.)
- **psych** (*Procedures for Psychological, Psychometric, and Personality Research*) - William Revelle
- **VIM** (*Visualization and Imputation of Missing Values*) - Matthias Templ (et al.)

V případě jakýchkoli dotazů k uvedeným přílohám nebo samotné knize mě můžete kontaktovat na emailové adrese **ivan.petrusek@soc.cas.cz** nebo **ipetrusek@gmail.com**

Ivan Petrušek

20.10.2015, Praha