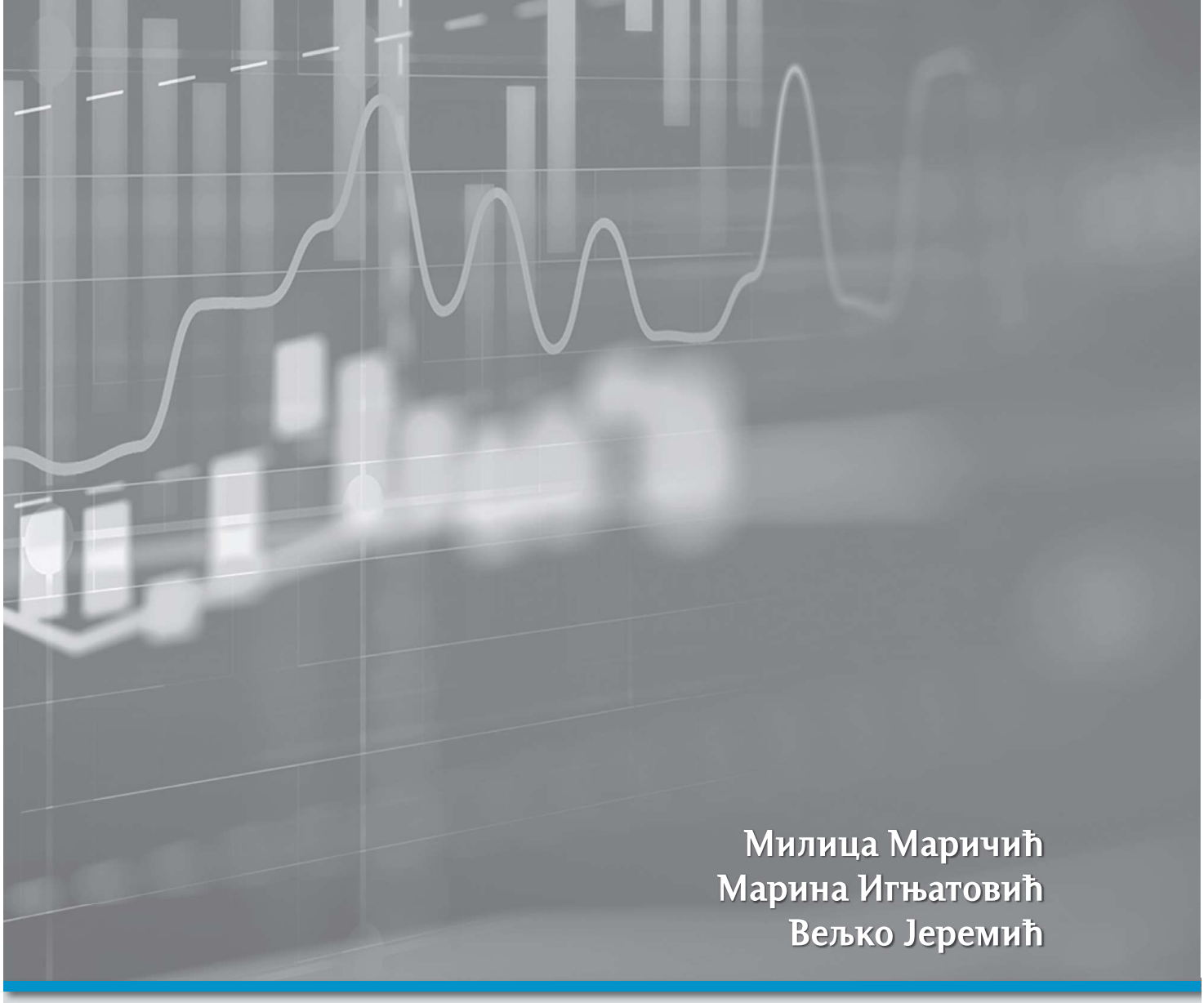


Милица Маричић
Марина Игњатовић
Вељко Јеремић

Модели статистичког учења

Академска мисао

Садржај поглавља

1 Увод у моделе статистичког учења и софтверско окружење R 23

2 Основе статистичког закључивања 55

3 Визуелизација података 123

4 Прост линеарни регресиони модел 183

5 Вишеструки линеарни регресиони модел 245

6 Додатне теме у регресионој анализи 295

7 Регуларизација линеарних регресионих модела 339

8 Логистичка регресија 393

9 Кластеровање 443

10 Бикластеровање 513

11 Моделовање структурних једначина 573

12 Симултане једначине 661

13 Решења задатака за самосталан рад 709

Садржај

Предговор **15**
Захвалница **16**
О ауторима **17**
Како користити ову књигу? **19**
Како су поглавља конципирана? **21**

1 Увод у моделе статистичког учења и софтверско окружење R 23

1.1 Модели статистичког учења **26**
1.2 Статистичке методе **31**
1.3 Увод у софтверско окружење R **33**
1.3.1 Зашто баш софтверско окружење R? 34
1.3.2 Инсталација софтверског окружења R, интегрисане развојне средине RStudio и основна подешавања 36
1.3.3 Основе R синтаксе и инсталирање R пакета 39
1.3.4 Шта ако R код не функционише? 42
1.4 Матрице података коришћене у овој књизи **43**
1.4.1 Матрица WEF 44
1.4.2 Матрица CFI2021 45
1.4.3 Матрица Capitals 45
1.4.4 Матрица SPI2019 45
1.4.5 Матрица SSI2016 46
1.4.6 Матрица QS2020 46
1.4.7 Матрице ARWU2016 и ARWU2021 47
1.4.8 Матрица GFSI2021 47
1.4.9 Матрица Sponsorship 47
1.4.10 Матрица SpectralData 49
1.4.11 Матрице PonudaTraznja1 и PonudaTraznja2 49
1.5 Закључна разматрања **50**
1.5.1 Кључне речи и појмови 51
1.5.2 Верзије R пакета коришћене у овој књизи 52
1.5.3 Списак слика и табела 52
1.5.4 Литература 52

2 Основе статистичког закључивања 55

2.1 Дескриптивна статистика 58
2.1.1 Статистичка обележја 60
2.1.2 Анализа фреквенција 62
2.1.3 Мере централне тенденције 63
2.1.4 Мере варијабилитета 65
2.1.5 Дескриптивна статистика у софтверском окружењу R 70

2.2 Аналитичка статистика 79
2.2.1 Теоријске основе 79
2.2.2 Параметарско тестирање хипотеза 86
2.2.3 Параметарско тестирање хипотеза у софтверском окружењу R 90
2.2.4 Непараметарско тестирање хипотеза 102
2.2.5 Непараметарско тестирање хипотеза у софтверском окружењу R 106

2.3 Задаци за самосталан рад 115
2.3.1 Дескриптивна статистика 115
2.3.2 Параметарско тестирање хипотеза 115
2.3.3 Непараметарско тестирање хипотеза 116

2.4 Закључна разматрања 116
2.4.1 Кључне речи и појмови 119
2.4.2 Коришћени R пакети и функције 119
2.4.3 Списак слика и табела 120
2.4.4 Литература 120

3 Визуелизација података 123

3.1 Визуелизација података у софтверском окружењу R 126

3.2 Униваријантни графикони 130
3.2.1 Графикони континуалне променљиве 130
3.2.2 Графикони категоријске променљиве 138

3.3 Мултиваријантни графикони 145
3.3.1 Графикони две континуалне променљиве 145
3.3.2 Графикони једне континуалне и једне категоријске променљиве 151
3.3.3 Графикони две категоријске променљиве 156
3.3.4 Графикони три и више променљивих 160

3.4 Специфични графикони 164
3.4.1 Статичке мапе 165
3.4.2 График корелационе матрице 169

3.4.3 Динамичке мапе 171
3.4.4 Интерактивни дијаграм зависности 172
3.5 Начини експортовања графика из софтверског окружења R 174
3.6 Задаци за самосталан рад 176
3.7 Закључна разматрања 177
3.7.1 Кључне речи и појмови 178
3.7.2 Коришћени R пакети и функције 178
3.7.3 Списак слика и табела 179
3.7.4 Литература 181

4 Прост линеарни регресиони модел 183

4.1 Основни теоријски концепти и анализе 186
4.1.1 Коефицијент корелације 186
4.1.2 Детерминистичка и стохастичка веза 189
4.1.3 Дијаграм зависности 192
4.2 Теоријске основе простог линеарног регресионог модела 193
4.3 Претпоставке простог линеарног регресионог модела 198
4.4 Метода најмањих квадрата 200
4.4.1 Особине оцена добијених методом најмањих квадрата 205
4.5 Случајне грешке, варијанса регресије и варијансе оцена регресионих параметара 207
4.6 Мере репрезентативности простог линеарног регресионог модела 210
4.6.1 Стандардна грешка регресије 210
4.6.2 Коефицијент детерминације 212
4.7 Тестирање значајности простог линеарног регресионог модела 215
4.7.1 Т-тест 215
4.7.2 Интервали поверења за оцене регресионих параметара 218
4.7.3 Ф-тест (анализа варијансе) 219
4.8 Коришћење простог линеарног регресионог модела за оцењивање и предвиђање 222
4.8.1 Оцењивање просечних вредности 223
4.8.2 Предвиђање индивидуалних вредности 223
4.9 Спровођење просте линеарне регресионе анализе у софтверском окружењу R 225
4.10 Задаци за самосталан рад 239

4.11 Закључна разматрања 240
4.11.1 Кључне речи и појмови 242
4.11.2 Коришћени R пакети и функције 243
4.11.3 Списак слика и табела 243
4.11.4 Литература 244

5 Вишеструки линеарни регресиони модел 245

5.1 Претпоставке вишеструког линеарног регресионог модела 248
5.2 Линеарни регресиони модел са две независне променљиве 250
5.2.1 Метода најмањих квадрата 252
5.3 Мере репрезентативности вишеструког линеарног регресионог модела 254
5.3.1 Стандардна грешка регресије 254
5.3.2 Коефицијент вишеструке детерминације 255
5.3.3 Подешени коефицијент детерминације 257
5.4 Матрична формулација вишеструког линеарног регресионог модела 258
5.5 Тестирање значајности вишеструког линеарног регресионог модела 259
5.5.1 Т-тест 259
5.5.2 Интервали поверења за оцене регресионих параметара 262
5.5.3 Ф-тест 263
5.6 Коришћење вишеструког линеарног регресионог модела за оцењивање и предвиђање 265
5.7 Методе селекције објашњавајућих променљивих које чине вишеструки линеарни регресиони модел 267
5.8 Одређивање релативног доприноса објашњавајућих променљивих у моделу 269
5.9 Спровођење вишеструке линеарне регресије у софтверском окружењу R 270
5.10 Задаци за самосталан рад 290
5.11 Закључна разматрања 291
5.11.1 Кључне речи и појмови 292
5.11.2 Коришћени R пакети и функције 292
5.11.3 Списак слика и табела 293
5.11.4 Литература 293

6 Додатне теме у регресионој анализи 295

6.1 Мултиколинеарност 298
6.1.1 Последице мултиколинеарности 301
6.1.2 Методе одређивања мултиколинеарности 302
6.1.3 Решење проблема мултиколинеарности 304

6.2 Аутокорелација 305
6.2.1 Последице аутокорелације 307
6.2.2 Методе одређивања аутокорелације 308
6.2.3 Решење проблема аутокорелације 310

6.3 Хетероскедастичност 310
6.3.1 Последице хетероскедастичности 314
6.3.2 Методе одређивања хетероскедастичности 314
6.3.3 Решење проблема хетероскедастичности 317

6.4 Лажне променљиве 318
6.4.1 Лажне променљиве које мењају одсечак (*intercept*) 319
6.4.2 Лажне променљиве које мењају нагиб (*slope*) 320

6.5 Спровођење додатних анализа у линеарној регресији у софтверском окружењу R 322

6.6 Задаци за самосталан рад 335

6.7 Закључна разматрања 335
6.7.1 Кључне речи и појмови 336
6.7.2 Коришћени R пакети и функције 337
6.7.3 Списак слика 337
6.7.4 Литература 337

7 Регуларизација линеарних регресионих модела 339

7.1 Принцип регуларизације 342

7.2 Риџ регресија 345

7.3 Ласо регресија 348

7.4 Компарација риџ и ласо регресије 351

7.5 Одређивање вредности параметра ламбда 352

7.6 Спровођење риџ и ласо регресије у софтверском окружењу R 354

7.7 Задаци за самосталан рад 387

7.8 Закључна разматрања 388
7.8.1 Кључне речи и појмови 389
7.8.2 Коришћени R пакети и функције 390
7.8.3 Списак слика и табела 390
7.8.4 Литература 391

8 Логистичка регресија 393

8.1 Модел логистичке регресије 397
8.1.1 Поређење логистичког и линеарног модела 400
8.2 Логит модел 401
8.2.1 Однос шанси *OR* (*odds ratio*) 402
8.2.2 Особине логит модела 404
8.3 Оцена и евалуација логит модела 406
8.3.1 Метод максималне веродостојности 406
8.3.2 Валд-ов тест (*Wald test*) и Z-тест 407
8.3.3 *Likelihood Ratio* (*LR*) или *Deviance* тест 408
8.3.4 Псеудо коефицијент детерминације 409
8.3.5 Анализа резидуала 410
8.3.6 Евалуација класификације 410
8.4 Спровођење логистичке регресије у софтверском окружењу R 413
8.5 Задаци за самосталан рад 437
8.6 Закључна разматрања 438
8.6.1 Кључне речи и појмови 440
8.6.2 Коришћени R пакети и функције 441
8.6.3 Списак слика и табела 441
8.6.4 Литература 442

9 Кластеровање 443

9.1 Теоријске основе 449
9.1.1 Скалирање података 449
9.1.2 Мере удаљености 451
9.1.3 Матрица удаљености 453
9.1.4 Нестандардне опсервације и њихова детекција 454
9.2 Хијерархијско кластеровање 458
9.2.1 Метода удруживања 460
9.2.2 Метода раздвајања 462
9.3 Партиционо кластеровање 464
9.3.1 Алгоритам к-средина (*k-means*) 465
9.3.2 Алгоритам к-медоида (*k-medoids*) 467
9.3.3 Алгоритам кластеровања
великих скупова података (*CLARA*) 469
9.4 Валидација кластеровања 470
9.4.1 Испитивање тенденције кластеровања 471
9.4.2 Метрике квалитета кластеровања 473
9.4.3 Одређивање броја кластера 475

9.5 Спровођење кластеровања у софтверском окружењу R 476
9.6 Задаци за самосталан рад 507
9.7 Закључна разматрања 508
9.7.1 Кључне речи и појмови 509
9.7.2 Коришћени R пакети и функције 510
9.7.3 Списак слика и табела 510
9.7.4 Литература 512

10 Бикластеровање 513

10.1 Теоријске основе 517
10.1.1 Типови бикластера 517
10.1.2 Конфигурације бикластера 521
10.1.3 Алгоритамски приступи 524
10.2 Одабрани алгоритми бикластеровања 525
10.2.1 Алгоритам δ-бикластеровања 526
10.2.2 Алгоритам FLOC 528
10.2.3 Алгоритам xMotif 529
10.2.4 Алгоритам Bimax 530
10.2.5 Спектрално бикластеровање 531
10.2.6 Алгоритам FABIA 534
10.3 Евалуација бикластер структуре 536
10.4 Визуелизација резултата бикластеровања. 539
10.5 Спровођење бикластеровања у софтверском окружењу R 540
10.6 Задаци за самосталан рад 565
10.7 Закључна разматрања 566
10.7.1 Кључне речи и појмови 568
10.7.2 Коришћени R пакети и функције 569
10.7.3 Списак слика и табела 569
10.7.4 Литература 570

11 Моделовање структурних једначина 573

11.1 Теоријске основе 577
11.1.1 Анализа путање 578
11.1.2 Конфирматорна факторска анализа 582
11.1.3 Модел структурних једначина – CEM модел 584
11.1.4 Метрике поузданости и валидности скала 592
11.1.5 Метрике квалитета модела 595
11.1.6 Правила моделовања – Често постављена питања 599

11.2 Моделовање над континуалним и ординалним променљивим у софтверском окружењу R 602
11.2.1 R пакети и синтакса 602
11.2.2 Модел анализе путање 604
11.2.3 Конфирматорна факторска анализа 610
11.2.4 CEM модел са латентним променљивим 616
11.2.5 Модел са латентним и измереним променљивим 623
11.3 Моделовање над номиналним променљивим 634
11.3.1 Моделовање над номиналним променљивим у софтверском окружењу R 635
11.4 Хијерархијско моделовање 639
11.4.1 Хијерархијско моделовање у софтверском окружењу R 641
11.5 Моделовање засновано на методи делимичних најмањих квадрата 645
11.5.1 Моделовање засновано на методи делимичних најмањих квадрата у софтверском окружењу R 646
11.6 Задаци за самосталан рад 654
11.7 Закључна разматрања 655
11.7.1 Кључне речи и појмови 657
11.7.2 Коришћени R пакети и функције 657
11.7.3 Списак слика и табела 658
11.7.4 Литература 659

12 Симултане једначине 661

12.1 Модели симултаних једначина 665
12.1.1 Структурни облик модела 665
12.1.2 Редуковани облик модела 666
12.2 Испитивање претпоставки модела 667
12.2.1 Математичка комплетност модела 667
12.2.2 Идентификација модела 667
12.2.3 Поступак идентификације модела 668
12.3 Методе за оцењивање симултаних једначина 674
12.3.1 Индиректна метода најмањих квадрата 674
12.3.2 Двостепена метода најмањих квадрата 677
12.4 Оцењивање симултаних једначина у софтверском окружењу R 678
12.5 Задаци за самосталан рад 704
12.6 Закључна разматрања 704
12.6.1 Кључне речи и појмови 706
12.6.2 Коришћени R пакети и функције 706

12.6.3 Списак табела 707
12.6.4 Литература 707

13 Решења задатака за самосталан рад 709

13.1 Основе статистичког закључивања 711
13.1.1 Дескриптивна статистика 711
13.1.2 Параметарско тестирање хипотеза 713
13.1.3 Непараметарско тестирање хипотеза 714
13.2 Визуелизација података 715
13.3 Прост линеарни регресиони модел 718
13.4 Вишеструки линеарни регресиони модел 720
13.5 Додатне теме у регресионој анализи 722
13.6 Регуларизација линеарних регресионих модела 723
13.7 Логистичка регресија 727
13.8 Кластеровање 730
13.9 Бикластеровање 732
13.10 Моделовање структурних једначина 735
13.11 Симултане једначине 739

Предговор

Ова књига је намењена како студентима различитих нивоа студија, тако и читаоцима који су заинтересовани да унапреде своја знања из области статистичког закључивања и модела статистичког учења. Може се користити као уџбеник за појединачне предмете и наставне јединице које се изучавају у области Рачунарске статистике на високошколским установама. Такође, ова књига може да послужи аналитичарима и доносиоцима одлука у организацијама који желе да самостално савладају одређене моделе статистичког учења и њихово спровођење у софтверском окружењу R.

Књига има за циљ да читаоцу помогне да стекне и продуби знања о моделима статистичког учења, да га оспособи да препозна и примени адекватни модел статистичког учења, као и да, на крају, правилно интерпретира и представи добијене резултате. Осим тога, наш циљ је био да читаоцу на једноставан и јасан начин представимо, са једне стране теоријске основе модела статистичког учења, а са друге њихову практичну примену у софтверском окружењу R и да при томе направимо добар баланс између теоријског и практичног дела.

Основна идеја књиге је да читаоцу на разумљив начин уз велики број примера пружи увод у основе већег броја модела статистичког учења. Правилно разумевање основа модела статистичког учења може послужити као одличан увод за разумевање и коришћење напреднијих модела и метода које се у континуитету развијају. Радујемо се повратним информацијама читалаца које ће нам послужити као смернице приликом израде новог издања.

„Бројеви имају причу коју желе да нам испричају, док статистика може да нам помогне да ту причу чујемо и разумемо".

Београд 2022.

Аутори

Захвалница

Професорки Радојки Малетић и професору Владану Девецићу на детаљним и конструктивним коментарима који су нам значајно помогли приликом писања ове књиге и допринели њеном квалитету.

Нашим менторима, професорки Милици Булајић и професору Зорану Радојичићу, који су у нама препознали интересовање за рачунарску статистику, несебично са нама поделили своје знање у области и од нас начинили предаваче и истраживаче какви смо данас.

Нашој матичној академској институцији, Факултету организационих наука, Универзитета у Београду, што нам је пружио прилику да се под његовим окриљем усавршавамо и напредујемо.

Студентима и експертима из привреде чијим смо питањима и коментарима вођени приликом писања ове књиге.

Колегама и пријатељима на речима охрабрења, саветима, размени искустава и разумевању које су нам пружили током процеса писања ове књиге.

Хвала нашим родитељима, Александри и Милану, Душки и Петру, Љубинки и Милојку на љубави и подршци коју су нам увек пружали и пружају је и даље свакодневно.

Књигу посвећујемо нашим породицама и свима које волимо и који нас воле.

О ауторима

Милица Маричић је доцент Универзитета у Београду, Факултета организационих наука, Катедре за операциона истраживања и статистику. Основне, мастер и докторске студије је завршила на Факултету организационих наука специјализирајући се у области рачунарске статистике, анализе података, пословне статистике, као и у области креирања и евалуације композитних индикатора. Од 2015. године је ангажована у настави за ужу научну област Рачунарска статистика. Ангажована је на предметима који се изводе на основним, мастер и докторским студијама. Стручно се усавршавала на Универзитету *Radboud*, Холандија, Универзитету *Justus-Liebig*, Немачка и Берлинској школи економије и права (*Berlin School of Economics and Law*), Немачка. Награђена је за најбољи рад студената докторских студија (*Young Scholars for Better Statistics Award*) на конференцији *Better Statistics for Better Lives* коју су организовали OECD (*Organisation for Economic Co-operation and Development*) и IAOS (*International Association for Official Statistics*). Ангажована је на неколико пројеката од националног значаја. Објавила је преко 60 научних радова од којих се истиче 13 радова објављених у индексираним часописима.

~

Марина Игњатовић је ванредни професор Универзитета у Београду, Факултета организационих наука, Катедре за операциона истраживања и статистику. Основне, мастер и докторске студије завршила је на Факултету организационих наука специјализирајући се у области анализе података, рачунарске статистике, економетрије, анализе временских серија, као и у области креирања и евалуације композитних индикатора. Од 2008. године ангажована је у настави за ужу научну област Рачунарска статистика. Ангажована је на предметима који се изводе на основним, мастер и докторским студијама. Стручно се усавршавала на Берлинској школи економије и права (*Berlin School of Economics and Law*), Немачка. Ангажована је на неколико пројеката од националног значаја. Објавила је преко 40 научних радова у домаћим часописима, као и у часописима међународног значаја.

~

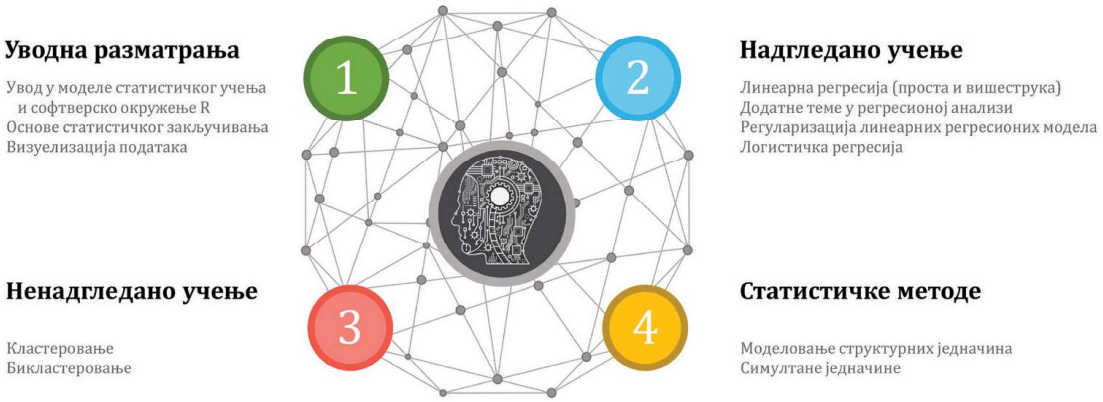
Вељко Јеремић је ванредни професор Универзитета у Београду, Факултета организационих наука, Катедре за операциона истраживања и статистику. Основне, мастер и докторске студије је завршио на Факултету организационих наука специјализирајући се у области рачунарске статистике и анализе података, мапирања научно-истраживачке изврсности појединаца и институција, као и

у области креирања и евалуације композитних индикатора. Од 2009. године је ангажован у настави за ужу научну област Рачунарска статистика на предметима који се изводе на основним, мастер, специјалистичким и докторским студијама. Гостујући професор на Универзитету *Roma Tre*, Италија, Универзитету *Federico II*, Италија, Универзитету *Vilnius*, Литванија, Универзитету *Kozminski*, Пољска, Универзитету *Catholique de Lille*, Француска, као и Универзитету *Heilbronn*, Немачка. Цитиран је више од 1500 пута у *Google Scholar* индексној бази. Ангажован је на неколико пројеката од националног значаја. Руководилац је пројекта "*Setting foundation for caPAcity buildiNg of shAring Community in SErbiA*" – PANACEA који је финансиран од стране Фонда за науку Републике Србије у оквиру програма ИДЕЈЕ.

Како користити ову књигу?

Научна истраживања су показала да међу студентима постоји страх и видна анксиозност од статистике и повезаних научних дисциплина (*fear of statistics*). Полазећи од тога, покушали смо да на једноставан, приступачан и разумљив начин читаоцима приближимо област рачунарске статистике, посебно моделе статистичког учења. Из тог разлога је књига подељена у логичке целине и поглавља, при чему свако поглавље има јасну структуру. Такође, књига је у великој мери интерактивна и на тај начин олакшава читаоцима упознавање са изложеном материјом.

Ова књига се може поделити у четири целине које су графички представљене на Слици 0.1:



Слика 0.1 Целине књиге Модели статистичког учења

Прва целина се односи на уводна разматрања (Поглавља 1-3) у којима су приказани основни концепти потребни за разумевање модела статистичког учења. У првом поглављу су представљени основни типови модела статистичког учења, као и основе коришћења софтверског окружења R у коме је спроведен практични део књиге. У другом поглављу су изложене основе статистичког закључивања које обухватају дескриптивну и аналитичку статистику. Треће поглавље има за циљ да представи неке од метода визуелизације података које су доступне у софтверском окружењу R.

Друга целина се односи на алгоритме надгледаног учења. Приказане су основе просте и вишеструке линеарне регресије као и додатне теме које се у ове две области јављају (Поглавља 4, 5 и 6). Затим је у Поглављу 7 дат осврт на приступе регуларизацији линеарних регресионих модела. У последњем поглављу ове целине, у Поглављу 8, дате су основе логистичке регресије.

Трећа целина се односи на методе и алгоритме ненадгледаног учења. У Поглављу 9 је представљен већи број алгоритама кластеровања, док су у Поглављу 10 представљени у пракси често коришћени алгоритми бикластеровања.

У четвртој целини (Поглавља 11 и 12) су приказане теоријске основе и практични примери решавања моделовања структурних једначина и симултаних једначина.

Последње поглавље књиге, Поглавље 13, чине решења задатака за самосталан рад.

Као што се са Слике 0.1 може видети, иако одвојене, целине ове књиге су међусобно повезане, допуњују се и омогућавају систематично и постепено упознавање са моделима статистичког учења. Књига је конципирана тако да су у почетним деловима поглавља изнете теоријске основе посматране методе или алгоритма, док су у наставку дати практични примери у софтверском окружењу R и задаци за самосталан рад.

Свака од метода која је теоријски представљена у књизи додатно је објашњена кроз практични пример који је спроведен у софтверском окружењу R. Матрице над којима су спроведени примери, као и матрице над којима се могу спровести задаци за самосталан рад су јавно доступне на GitHub страници књиге: https://github.com/msu-fon. Осим матрица, на GitHub страници књиге налазе се и додатни материјали у виду статистичких таблица, као и одређених R кодова.

Сугеришемо читаоцима да преузму матрице и покушају да сами спроведу и реше захтеве који су приказани у књизи јер ће на тај начин додатно савладати код и спровођење модела статистичког учења у софтверском окружењу R.

Како су поглавља конципирана?

Свако поглавље почиње са кратким уводним разматрањима која имају за циљ да читаоца упознају са моделом статистичког учења који ће у поглављу бити приказан.

Затим су приказане теоријске основе модела статистичког учења којима се детаљно објашњавају концепти, појмови и терминологија која се користи у области.

Закључак сваког теоријског поглавља се налази на крају поглавља и јасно је обележен плавом бојом.

Нулта и алтернативна хипотеза сваког статистичког теста који је приказан у књизи су такође јасно обележене:

H_0: Нулта хипотеза теста
H_1: Алтернативна хипотеза теста

Коришћени R код је детаљно објашњен, са посебним акцентом на параметризацију функција. R код је обележен као:

```
# Opis funkcije
funkcija
```

Резултати покретања кода, аутпут, су такође приказани. У зависности од модела статистичког учења приказан је или цео аутпут или само најважнији делови. Аутпут је обележен као:

```
Autput
```

Главни резултати и закључци који се доносе на основу добијеног аутпута су детаљно интерпретирани.

Закључак сваког задатка је обележен светло плавом бојом.

У сваком поглављу се налази неколико задатака за самосталан рад који читаоцу омогућују да провери и усаврши своје стечено знање.

На крају сваког поглавља се налази поглавље у коме су дата закључна разматрања и у коме се налази сумарни приказ кључних речи, појмова, као и формула које се у посматраном поглављу први пут појављују.

Како би читаоци имали јасан преглед R пакета и функција који су коришћени у поглављу, на крају сваког поглавља се налази листа коришћених пакета и функција.

Уколико читаоци желе да употпуне или прошире своје знање о одређеном моделу статистичког учења, на самом крају поглавља налази се списак литературе која је коришћена приликом писања поглавља и која читаоцу може пружити додатне информације.

Непараметарско тестирање хипотеза подразумева тестирање хипотеза које се односе на целокупну расподелу над променљивим које не подлежу Нормалној расподели.

2.2.5 Непараметарско тестирање хипотеза у софтверском окружењу R

R пакети коришћени у овом поглављу су *readr* (Wickham et al., 2021), *coin* (Hothorn et al., 2021), *dunn.test* (Dinno, 2017), *gmodels* (Warnes et al., 2018) и *vcd* (Meyer et al., 2020). Већина функција које су потребне за спровођење непараметарских тестова су имплементиране у основно окружење R. Пакет *readr* се користи за учитавање матрице података, пакет *coin* омогућава спровођење Ман-Витни теста, пакет *dunn.test* омогућава спровођење Дановог теста, пакет *gmodels* омогућава спровођење Табеле контигенције, док пакет *vcd* омогућава спровођење детаљније анализе Табеле контигенције.

```
# Instaliranje paketa potrebnih za sprovođenje neparametarskih testova
install.packages("readr")
install.packages("coin")
install.packages("dunn.test")
install.packages("gmodels")
install.packages("vcd")

# Pokretanje paketa potrebnih za sprovođenje neparametarskih testova
library(readr)
library(coin)
library(dunn.test)
library(gmodels)
library(vcd)
```

Матрица података над којом ће бити приказано спровођење непараметарских тестова је матрица *WEF.csv* (Поглавље 1.4.1). Само учитавање матрице је приказано у Поглављу 2.1.5.

Задатак 20. Испитати да ли је медијана популације индикатора *Institutions_Score_2011* једнака 3.1.

Индикатор *Institutions_Score_2011* је променљива која је континуална, а поред тога је потребно да испитамо да је ли је њена медијана популације једнака претпостављеној. Имајући у виду да испитујемо вредност медијане, а не средње вредности, користићемо Вилкоксонов тест медијане.

Свакако је први корак да се испита да ли је посматрана променљива нормално распоређена и у ту сврху ћемо користити непараметарски Шапиро-Вилков тест.

```
# Shapiro-Wilk test
shapiro.test(WEF$Institutions_Score_2011)
```

```
Shapiro-Wilk normality test

data:  WEF$Institutions_Score_2011
W = 0.94523, p-value = 3.116e-05
```

Нулта хипотеза теста гласи да је променљива нормално распоређена, а алтернативна да није. Вредност статистике је W = 0.945. Пошто је добијена *p*-вредност мања од 0.05, прихватамо алтернативну хипотезу и тврдимо да на посматраном узорку променљива *Institutions_Score_2011* није нормално распоређена што указује да над њом можемо спроводити само непараметарске тестове.

Следећи корак у анализи је спровођење Вилкоксоновог теста медијане. Функција којом се позива тест је *wilcox.test*(), док је од параметара функције потребно дефинисати променљиву над којом се тест спроводи, као и претпостављену вредност медијане (*mu*).

```
# Wilcoxon signed rank test for a median
wilcox.test(WEF$Institutions_Score_2011, mu = 3.1)
```

```
Wilcoxon signed rank test with continuity correction

data:  WEF$Institutions_Score_2011
V = 9079, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 3.1
```

Нулта хипотеза је да је медијана индикатора 3.1, док је алтернативна хипотеза назначена у самом аутпуту и гласи да се медијана посматраног индикатора разликује од 3.1. V представља суму рангова и није потребно тумачити је. Пошто је добијена *p*-вредност мања од 0.05, одбацујемо нулту хипотезу и тврдимо да се медијана индикатора *Institutions_Score_2011* разликује од 3.1.

На нивоу значајности од 5% тврдимо да се медијана индикатора *Institutions_Score_2011* разликује од 3.1.

Задатак 21. Испитати да ли постоји разлика у вредностима индикатора *Institutions_Score_2011* између земаља које припадају факторској фази и фази T12.

Индикатор *Institutions_Score_2011* је променљива која је континуална, а у задатку 20 смо видели да није нормално распоређена. Имајући у виду да поредимо вредности две групе (групе земаља које су у факторској фази и фази T12) користићемо Ман-Витни тест.

Као и у случају код Студентовог т-теста за разлике просечних вредности потребно је формирати матрицу у којој ће се наћи само оне земље које су у групама које поредимо, у овом случају у факторској фази или фази T12. Користићемо функцију *subset*().

```
# Formiranje matrice sa odgovarajućim entitetima
WEF_FT <- subset(WEF, (Faza == "Faktorska" | Faza == "T12"))
```

Следећи корак је спровођење Ман-Витни теста користећи функцију *wilcox_test*(). Од параметара функције потребно је дефинисати континуалну променљиву и категоријску променљиву.

```
#Mann-Whitney test
wilcox_test(WEF_FT$Institutions_Score_2011 ~ as.factor(WEF_FT$Faza))
```

```
Asymptotic Wilcoxon-Mann-Whitney Test

data:  WEF_FT$Institutions_Score_2011 by WEF_FT$Faza (Faktorska, T12)
Z = -2.0101, p-value = 0.04442
alternative hypothesis: true mu is not equal to 0
```

Нулта хипотеза је да су узорци из исте популације, док је алтернативна хипотеза да су узорци нису из исте популације. Вредност статистике је Z = -2.010. Пошто је добијена *p*-вредност мања од 0.05, одбацујемо нулту хипотезу и тврдимо да постоји статистички значајна разлика између група.

На нивоу од 5% тврдимо да постоји статистички значајна разлика у вредностима индикатора *Institutions_Score_2011* које остварују земље у факторској фази и фази T12.

Задатак 22. Испитати да ли постоји разлика у вредностима индикатора *Institutions_Score_2011* између земаља које припадају факторској фази, фази T12 и фази T23.

Индикатор *Institutions_Score_2011* је променљива која је континуална, а у задатку 20 смо видели да није нормално распоређена. Имајући у виду да поре-

димо вредности три групе (групе земаља које су у факторској фази, фази T12 и фази T23) користићемо Крускал-Валис тест.

У првом кораку је потребно је формирати матрицу у којој ће се наћи само оне земље које су у групама које поредимо, у овом случају у факторској фази, фази T12 или фази T23. Користићемо функцију *subset*().

```
# Formiranje matrice sa odgovarajućim entitetima
WEF_FTT <- subset(WEF, (Faza == "Faktorska" | Faza == "T12" | Faza=="T23"))
```

Следећи корак је спровођење Крускал-Волис теста користећи функцију *kruskal.test*(). Потребно је дефинисати континуалну променљиву, категоријску променљиву и матрицу података.

```
# Kruskal-Wallis test
kruskal.test(Institutions_Score_2011 ~ Faza, data = WEF_FTT)
```

```
Kruskal-Wallis rank sum test

data:  Institutions_Score_2011 by Faza
Kruskal-Wallis chi-squared = 9.4335, df = 2, p-value = 0.008944
```

Нулта хипотеза је да су узорци из исте популације, док је алтернативна хипотеза да узорци нису из исте популације. Вредност статистике је $\chi^2 = 9.434$. Пошто је добијена *p*-вредност мања од 0.05, одбацујемо нулту хипотезу и тврдимо да постоји статистички значајна разлика између група.

Као што се из резултата може видети, спровођењем Крускал-Валис теста добијамо информацију о томе да ли разлике између група постоје или не. Како би се одредило између којих тачно група постоје разлике користи се Данов пост хок тест.

Данов тест спроводимо користећи функцију *dunn.test*(). Потребно је дефинисати континуалну променљиву, категоријску променљиву, да ли примењујемо неку корекцију *p*-вредности и да ли желимо да се спроведе Крускал-Валис тест (*kw=TRUE*).

```
# Dunn's test
dunn.test(WEF_FTT$Institutions_Score_2011, WEF_FTT$Faza,
          method = "bonferroni", kw = TRUE)
```

```
Kruskal-Wallis rank sum test

data: x and group
Kruskal-Wallis chi-squared = 9.4335, df = 2, p-value = 0.01

                    Comparison of x by group
                          (Bonferroni)
Col Mean-|
Row Mean | Faktorska        T12
---------+----------------------
     T12 |-2.027084
         |  0.0640
         |
     T23 |-2.856542     -0.563404
         |  0.0064*        0.8597

alpha = 0.05
Reject Ho if p <= alpha/2
```

У првом делу аутпута је поново дата вредност Крускал-Валис теста, која је, као што се види, иста као и вредност коју смо добили позивањем Крускал-Валис теста функцијом *kruskal.test*().

У наставку је дата табела унакрсног поређења група са одговарајућим вредностима статистика и *p*-вредностима. Нулта хипотеза је да су узорци из исте популације, а алтернативна да нису из исте популације. На основу добијених *p*-вредности закључујемо да постоји статистички значајна разлика између вредности које остварују земље у факторској фази и фази T23.

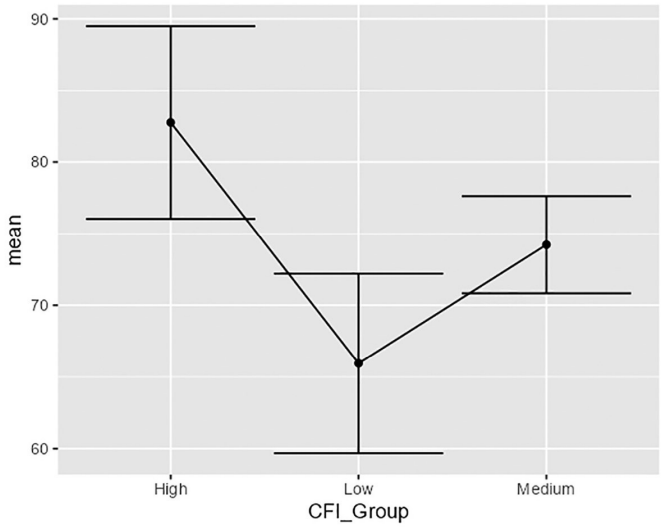
На нивоу од 5% тврдимо да постоји статистички значајна разлика у вредностима индикатора *Institutions_Score_2011* које остварују земље у факторској фази, фази T12 и фази T23. Пост хок тест је показао да постоји статистички значајна разлика између земаља у факторској фази и фази T23.

Задатак 23. Испитати да ли је дошло до промене вредности индикатора *Innovation_Score* између 2011. и 2013. године.

Индикатор *Innovation_Score* је променљива која је континуална, а поред тога је потребно да испитамо да је ли је дошло до промене вредности између 2011. и 2013. године.

Први корак је да се испита да ли су посматране променљиве нормално расподељене и користићемо непараметарски Шапиро-Вилков тест.

```
# Shapiro-Wilk test
shapiro.test(WEF$Innovation_Score_2011)
shapiro.test(WEF$Innovation_Score_2013)
```



Слика 3.25 Дијаграм средњих вредности по групама
са интервалима поверења – Задатак 10

Најчешће коришћени мултиваријантни графикони једне континуалне и једне категоријске променљиве су хистограм по групама, боксплот по групама, као и дијаграм сумарних статистика по групама.

3.3.3 Графикони две категоријске променљиве

Врло често у пракси је потребно је извршити анализу распореда по категоријама једне променљиве у односу на другу. У овом поглављу су представљене теоријске основе и примери мултиваријантних графикона који комбинују две категоријске променљиве. Овим поглављем је обухваћено креирање наслаганог стубичастог дијаграма, груписаног стубичастог дијаграма, као и сегментираног стубичастог дијаграма.

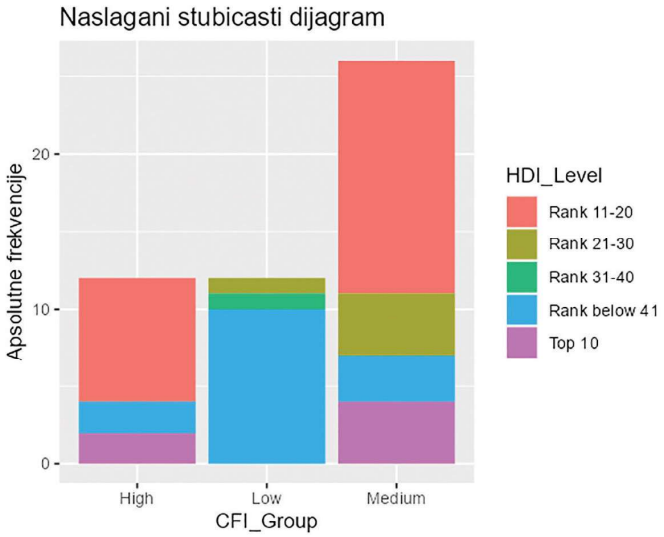
Наслагани стубичасти дијаграм - *Stacked bar chart*

Наслагани стубичасти дијаграм пружа информације о томе каква је расподела фреквенција једне категоријске променљиве у оквиру сваке појединачне категорије друге категоријске променљиве. Оно што је карактеристично за овај тип дијаграма је да је расподела фреквенција једне категоријске променљиве у оквиру сваке појединачне категорије друге категоријске променљиве приказана у оквиру једног стубића.

Задатак 11. Потребно је скицирати наслагани стубичасти дијаграм индикатора *HDI_Level* у оквиру индикатора *CFI_Group*. Индикатори се налазе у матрици *CFI2021.csv* (Поглавље 1.4.2).

Пошто је унос података приказан у задатку 1, одмах прелазимо на скицирање наслаганог стубичастог дијаграма. Скицирање наслаганог стубичастог дијаграма започињемо функцијом *ggplot*() у оквиру које је потребно да дефинишемо матрицу података као и промењљиве које желимо да нанесемо на осе графика. У оквиру функције *aes*() додатно дефинишемо параметар *fill* којим назначујемо да је стубиће дијаграма потребно испунити на основу вредности промењљиве *HDI_Level*. Користимо слој *geom_bar*() у оквиру кога дефинишемо да расподелу фреквенција промењљиве *HDI_Level* уносимо у стубиће сваке категорије промењљиве *CFI_Group*. Вредност параметра *position* је у овом случају *stack*. Функцијом *labs*() додатно дефинишемо назив графикона и назив Y осе.

```
# Naslagani stubičasti dijagram
ggplot(CFI2021, aes(x = CFI_Group, fill = HDI_Level)) +
  geom_bar(position = "stack") +
  labs(title = "Naslagani stubičasti dijagram",
       y = "Apsolutne frekvencije")
```



Слика 3.26 Наслагани стубичасти дијаграм – Задатак 11

Као што сам назив графикона имплицира, на њему се апсолутне фреквенције једне категоријске промењљиве „слажу“ у оквиру категорија друге категоријске промењљиве. На основу креираног графика можемо закључити да се у групи *low* налази највише градова из земаља којима је ранг на *HDI* индексу испод 41. места, да се у групи *medium* налази велики број градова из земаља којима је ранг на *HDI* индексу између 11. и 20. места, као и да се у групи *high* налази неколико градова из земаља којима је ранг на *HDI* индексу испод 41. места.

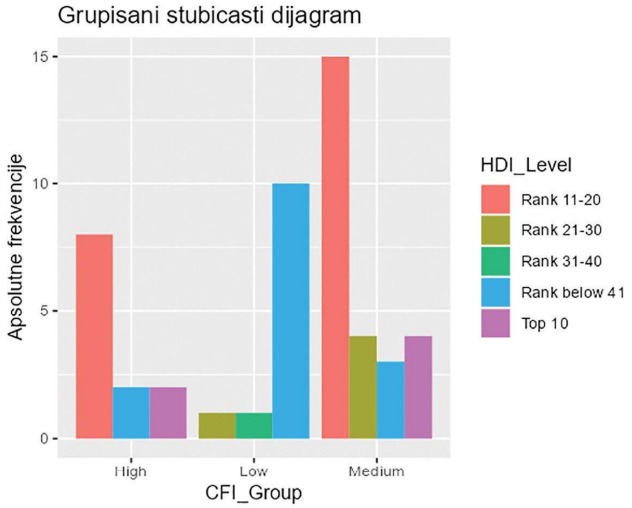
Груписани стубичасти дијаграм - *Grouped bar chart*

Слично као и наслагани стубичасти дијаграм, груписани стубичасти дијаграм пружа информације о томе каква је расподела фреквенција једне категоријске променљиве у оквиру сваке појединачне категорије друге категоријске променљиве. Основна разлика у односу на наслагани стубичасти дијаграм је у начину приказивања. Код груписаног стубичастог дијаграма се расподела фреквенција једне категоријске променљиве у оквиру сваке појединачне категорије друге категоријске променљиве приказује појединачним стубићима.

Задатак 12. Потребно је скицирати груписани стубичасти дијаграм индикатора *HDI_Level* у оквиру индикатора *CFI_Group*. Индикатори се налазе у матрици *CFI2021.csv* (Поглавље 1.4.2).

Пошто је унос података приказан у задатку 1, одмах прелазимо на скицирање груписаног стубичастог дијаграма. Скицирање груписаног стубичастог дијаграма започињемо функцијом *ggplot*() у оквиру које је потребно да дефинишемо матрицу података као и променљиве које желимо да нанесемо на осе графика. У оквиру функције *aes*() додатно дефинишемо параметар *fill* којим назначујемо да је стубиће дијаграма потребно испунити на основу вредности променљиве *HDI_Level*. Користимо слој *geom_bar*() у оквиру кога дефинишемо да расподелу фреквенција променљиве *HDI_Level* уносимо као појединачне стубиће сваке категорије променљиве *CFI_Group*. Вредност параметра *position* је у овом случају *dodge*. Функцијом *labs*() додатно дефинишемо назив графикона и назив Y осе.

```
# Grupisani stubičasti dijagram
ggplot(CFI2021, aes(x = CFI_Group, fill = HDI_Level)) +
  geom_bar(position = "dodge") +
  labs(title = "Grupisani stubičasti dijagram",
       y = "Apsolutne frekvencije")
```



Слика 3.27 Груписани стубичасти дијаграм – Задатак 12

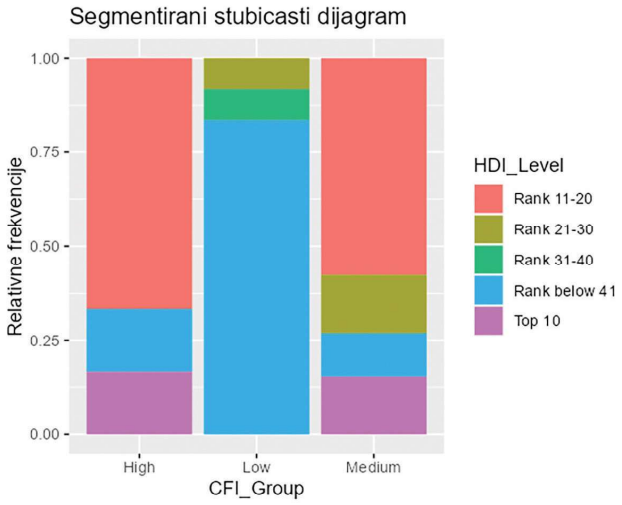
Сегментирани стубичасти дијаграм - *Segmented bar chart*

Сегментирани стубичасти дијаграм представља наслагани стубичасти дијаграм на коме су приказане релативне фреквенције. Овај график пружа информације о томе каква је расподела релативних фреквенција једне категоријске променљиве у оквиру сваке појединачне категорије друге категоријске променљиве.

Задатак 13. Потребно је скицирати сегментирани стубичасти дијаграм индикатора *HDI_Level* у оквиру индикатора *CFI_Group*. Индикатори се налазе у матрици *CFI2021.csv* (Поглавље 1.4.2).

Пошто је унос података приказан у задатку 1, одмах прелазимо на скицирање сегментираног стубичастог дијаграма. Скицирање сегментираног стубичастог дијаграма започињемо функцијом *ggplot*() у оквиру које је потребно да дефинишемо матрицу података као и променљиве које желимо да нанесемо на осе графика. У оквиру функције *aes*() додатно дефинишемо параметар *fill* којим назначујемо да је стубиће дијаграма потребно испунити на основу вредности променљиве *HDI_Level*. Користимо слој *geom_bar*() у оквиру кога дефинишемо да расподелу релативних фреквенција променљиве *HDI_Level* уносимо у стубиће сваке категорије променљиве *CFI_Group*. Вредност параметра *position* је у овом случају *fill*. Функцијом *labs*() додатно дефинишемо назив графикона и назив Y осе.

```
# Segmentirani stubičasti dijagram
ggplot(CFI2021, aes(x = CFI_Group, fill = HDI_Level)) +
  geom_bar(position = "fill") +
  labs(title = "Segmentirani stubičasti dijagram",
       y = "Relativne frekvencije")
```



Слика 3.28 Сегментирани стубичасти дијаграм – Задатак 13

тати како латентне променљиве *Препознатљивост* која настаје као линеарна комбинација *Prep1-Prep3* и *Став о спонзорима* која настаје као линеарна комбинација *Ss1-Ss8* утичу на латентну променљиву *Намера о куповини* која настаје као линеарна комбинација *Nk1-Nk5*.

11.7 Закључна разматрања

Моделовање структурних једначина или *Structural Equation Modelling* (СЕМ) представља скуп комплексних статистичких мултиваријационих анализа која се заснивају на комбинацији експлораторне факторске анализе и вишеструке линеарне регресионе анализе која омогућава испитивање каузалних веза.

- Анализа путање представља специфичан облик СЕМ анализе који испитује како је *n* независних променљивих статистички повезано са зависном променљивом. Оно што је карактеристично за моделе анализе путање је то што модел чине само измерене променљиве.
- Каузална веза може бити директна каузална веза – која води од узрока ка ефекту и корелациона каузална веза – која показује да између променљивих постоји узајамни однос.
- Каузална веза која се испитује коришћењем СЕМ анализе може бити директна – која показује како једна променљива утиче на другу променљиву и индиректна – која показује како једна променљива посредством једне или више променљивих утиче на неку другу променљиву.
- Конфирматорна факторска анализа представља облик факторске анализе у којој истраживач има могућност да одреди које ће измерене променљиве чинити латентну променљиву, као и то које ће латентне променљиве бити корелисане.
- Модел структурних једначина је модел који се састоји од измерених и латентних променљивих које по својој природи могу бити независне (егзогене) или зависне (ендогене, латентне).
- Параметарски приступ СЕМ анализи је базиран на анализи коваријанси (*Covariance Based SEM – CB-SEM*), док се непараметарски приступ базира на методи делимичних најмањих квадрата (*Partial Least Squares SEM – PLS-SEM*).
- Пре испитивања квалитета СЕМ модела препоручује се испитивање интерне конзистентности и валидности предложених скала. Метрике за евалуацију скала се деле на две велике групе метрика, на метрике поузданости (*Reliability*) и валидности (*Validity*).
- За евалуацију квалитета СЕМ модела користе се индекси апсолутног фитовања, индекси компаративног фитовања, као и индекси штедљивог фитовања.

- Приликом СЕМ моделовања јавља се низ питања на која је потребно да истраживач одговори и она су везана за комплексност модела, величину узорка, статистичку значајност оцењених регресионих коефицијената, јачину оцењених регресионих коефицијената, као и примену индекса модификације.
- У СЕМ анализи могуће је моделовање над подацима који су номинални. Приликом креирања модела у којима фигуришу номиналне променљиве потребно је за сваку категоријску променљиву која има k нивоа дефинисати $k-1$ бинарних променљивих, док је као методу оцењивања регресионих коефицијената потребно користити неку од метода најмањих квадрата (*LS*).
- Хијерархијско моделовање представља напредну СЕМ анализу која подразумева креирање и евалуацију модела у којим се формирају латентне променљиве вишег реда као линеарне комбинације латентних променљивих нижег реда.
- Моделовање засновано на методи делимичних најмањих квадрата је вид непараметарске СЕМ анализе који се спроводи у случајевима када нису испуњени услови параметарске СЕМ анализе (мали узорак, комплексан модел, подаци који су веома асиметрични и нису нормално распоређени).

Као и свака статистичка анализа СЕМ анализа има одређене предности:

- Веома практична статистичка анализа којом се могу верификовати различити концептуални модели;
- Широкопримењива анализа;
- Заснована је на јасном теоријском оквиру;
- Омогућава испитивање сложених, вишедимензионалних феномена и појава;
- Постоји већи број софтверских решења за њено спровођење и
- Ґрафичка је и њени резултати се могу лако интерпретирати и представити доносиоцима одлука.

Са друге стране, потенцијални недостаци СЕМ анализе су:

- Ако је креиран веома комплексан модел на малом узорку, постоји могућност да добијени резултати не буду валидни;
- Како се у пракси најчешће користи приступ генерисања модела, истраживач мора бити веома пажљив да „не претера“ у унапређењу квалитета модела;
- У принципу, за креирање модела средње комплексности потребан је велики узорак, бар преко 200 испитаника и
- Резултати анализе као и квалитет модела зависе од истраживача који спроводи анализу.

11.7.1 Кључне речи и појмови

Појам	Појам
Анализа путање	Метода пондерисаних најмањих квадрата (*Weighted Least Squares – WLS*)
Директна каузална веза	Метода дијагоналних најмањих квадрата (*Diagonal Weighted Least Squares – DWLS*)
Корелациона каузална веза	Спецификација модела
Директни утицаји	Идентификација модела
Индиректни утицаји	Респецификација модела
Конфирматорна факторска анализа	Метрике поузданости (*Reliability*)
Једнофакторски модел	Метрике валидности (*Validity*)
Измерене променљиве	Кронбахова алфа
Латентне променљиве	Индекси апсолутног фитовања
Стандардизовани регресиони коефицијент	Индекси компаративног фитовања
Нестандардизовани регресиони коефицијент	Индекси штедљивог фитовања
Ограничења регресионих коефицијената	Индекси модификације
Параметарска СЕМ анализа	Хејвудов случај
Непараметарска СЕМ анализа	Хијерархијско моделовање
Метод максималне веродостојности (*Maximum Likelihood Estimation – MLE*)	Променљиве првог реда (*First-order variables*)
Генерализована метода најмањих квадрата (*Generalized Least Squares – GLS*)	Променљиве другог реда (*Second-order variables*)

11.7.2 Коришћени R пакети и функције

Пакети			
lavaan	readr	semPLS	
psych	semPlot	semTools	
Функције			
alpha()	inspect()	plsLoadings()	sempls()
bootsempls()	install.packages()	plsm()	set.seed()
c()	library()	read_csv()	str()
cbind()	matrix()	round()	summary()
cfa()	optionsmax()	rSquared()	title()
fitMeasures()	parallelplot()	sem()	View()
ifelse()	pathCoeff()	semPaths()	

11.7.3 Списак слика и табела

Списак слика

Слика 11.1 „Породично стабло“ моделовања
структурних једначина и повезаних метода (Henseler, 2010) 576
Слика 11.2 Пример вишеструког линеарног регресионог модела 578
Слика 11.3 Пример модела анализе путање 579
Слика 11.4 Директни и индиректни утицаји
у моделу анализе путање - Пример 579
Слика 11.5 Зависне и независне променљиве
у моделу анализе путање - Пример 580
Слика 11.6 Резултат спровођења анализе путање - Пример 581
Слика 11.7 Експлораторна (а) и конфирматорна (б)
факторска анализа - Поређење 583
Слика 11.8 Елементи CEM модела: Латентне
и измерене променљиве и случајне грешке 585
Слика 11.9 Структурни и мерни CEM модел 585
Слика 11.10 Алгоритам спровођења CEM анализе (према Kline, 2016) 590
Слика 11.11 Категоризација скала
у зависности од нивоа поузданости и валидности 593
Слика 11.12 Графички приказ резултата анализе путање – Задатак 1 609
Слика 11.13 Графички приказ резултата модела
конфирматорне факторске анализе – Задатак 2 616
Слика 11.14 Графички приказ резултата CEM модела – Задатак 3 622
Слика 11.15 Графички приказ резултата CEM модела
са латентним и измереним променљивим – Задатак 4 628
Слика 11.16 Хијерархијски модел - Пример 640
Слика 11.17 Графички приказ резултата
хијерархијског CEM модела – Задатак 6 644
Слика 11.18 Графички приказ резултата
непараметарске CEM анализе са реузорковањем – Задатак 7 653

Списак табела

Табела 11.1 Једначине и елементи спољашњег
и унутрашњег CEM модела 586
Табела 11.2 Израчунавање стандардизованог
регресионог коефицијента – Пример 587
Табела 11.3 Компаративни приказ особина
и претпоставки параметарске и непараметарске CEM анализе 589
Табела 11.4 Индекси за мерење квалитета CEM модела
и генерално правило прихватљивости 598

11.7.4 Литература

Allen, J., Eboli, L., Mazzulla, G., & de Dios Ortúzar, J. (2020). Effect of critical incidents on public transport satisfaction and loyalty: an Ordinal Probit SEM-MIMIC approach. *Transportation, 47*(2), 827-863. https://doi.org/10.1007/s11116-018-9921-4

Bagozzi, R. P. (2010). Structural equation models are modelling tools with many ambiguities: Comments acknowledging the need for caution and humility in their use. *Journal of Consumer Psychology, 20*(2), 208-214. https://doi.org/10.1016/j.jcps.2010.03.001

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Epskamp, S., Stuber, S., Nak, J., Veenman, M., & Jorgensen, T. (2019). Package 'semPlot'. Доступно на: https://cran.r-project.org/web/packages/semPlot/semPlot.pdf

Garner, W.R. (1960). Rating scales, discriminability and information transmission. *Psychological Review*, 67, 343-352. https://doi.org/10.1037/h0043047

Hair, J.F.; Hult, G.T.M. Ringle, C.M.; Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd Edition). Thousand Oaks, CA: Sage.

Henseler, J. (2010). *Covariance-based Structural Equation Modeling: Foundations and Applications.* Доступно на: http://www.sensometric.org/Resources/Documents/2010/Meeting/Presentations/013-000-Henseler_2010.pdf

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods, 3*(4), 424-453. https://doi.org/10.1037/1082-989X.3.4.424

Hutchinson, S, R., & Olmos, A (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling: A Multidisciplinary Journal*, 5, 344-364. https://doi.org/10.1080/10705519809540111

Jorgensen, T., et al. (2021). Package 'semTools'. Доступно на: https://cran.r-project.org/web/packages/semTools/semTools.pdf

Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances: Is a Heywood case a symptom of misspecification?. *Sociological Methods & Research, 41*(1), 124-167. https://doi.org/10.1177/0049124112442138

Liu, Y., Wu, A. D., & Zumbo, B. D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Ordinal/rating scale item responses. *Educational and Psychological Measurement*, *70*(1), 5-21. https://doi.org/10.1177/0013164409344548

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490-504. https://doi.org/10.1037/0033-2909.111.3.490

Maricic, M., Kostic-Stankovic, M., Bulajic, M., & Jeremic, V. (2019). See it and believe it? Conceptual model for exploring the recall and recognition of embedded advertisements of sponsors. *International Journal of Sports Marketing and Sponsorship. 20*(2), 333-352. https://doi.org/10.1108/IJSMS-02-2018-0020

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness of fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410. https://doi.org/10.1007/BF01102761

Milenković, M., Glavić, D., & Maričić, M. (2019). Determining factors affecting congestion pricing acceptability. *Transport Policy*, 82, 58-74. https://doi.org/10.1016/j.tranpol.2019.08.004

Monecke, A. (2015). Package 'semPLS'. Доступно на: https://cran.r-project.org/web/packages/semPLS/semPLS.pdf

Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. *Best practices in quantitative methods*.

Revelle, W. (2021). Package 'psych'. Доступно на: https://cran.r-project.org/web/packages/psych/psych.pdf

Rosseel, Y. (2021). Package 'lavaan'. Доступно на: https://cran.r-project.org/web/packages/lavaan/lavaan.pdf

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research, 69*(10), 3998–4010. https://doi.org/10.1016/j.jbusres.2016.06.007

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). Utrecht work engagement scale-9. *APA PsycTests*. https://doi.org/10.1037/t05561-000

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, *2*, 53. https://doi.org/10.5116/ijme.4dfb.8dfd

Wickham, H. et al. (2021). Package 'readr'. Доступно на: https://cran.r-project.org/web/packages/readr/readr.pdf