

Reiczigel Jenő – Harnos Andrea – Solymosi Norbert

BIOSTATISZTIKA
nem statisztikusoknak

Pars Kft., Nagykovácsi
2019

A könyv 2007-es kiadása az Oktatási és Kulturális Minisztérium támogatásával,
a Felsőoktatási Tankönyv- és Szakkönyv-támogatási Pályázat keretében jelent meg.

Lektorok

LANG ZSOLT
KIS JÁNOS

Borító: Németh János

HU ISBN 978-963-88339-8-3
HU ISSN 1788-4349

© Reiczigel Jenő – Harnos Andrea – Solymosi Norbert, 2019
© Pars Kft., Nagykovácsi, 2019

A 2018-as kiadás javított utánnomása.

Minden jog fenntartva, beleértve a mű bővített, illetve rövidített változatainak kiadási jogát is. A kiadó írásos engedélye nélkül sem elektronikus, sem hagyományos úton nem sokszorosítható, tárolható, illetve terjeszthető.

Nyomdai munkálatok: Mondat Kft., Vác

Tartalomjegyzék

Előszó	1
Köszönetnyilvánítás	5
Hogyan olvassuk ezt a könyvet?	6
Szükséges előismeretek	7
Jelölések, írásmód	8
Ismerkedés az R-rel	9
Hogyan olvassuk az R-kódokat?	10
1. Bevezetés	13
1.1. Miért tanuljunk statisztikát?	13
1.2. Megjegyzések a példákhoz	16
1.3. Hétköznapi valószínűségszámítás és statisztika	20
2. A statisztika alapfogalmai	23
2.1. Populáció és minta	23
2.2. Leíró és induktív statisztika	27
2.3. Mintavételi módszerek	29
2.4. Az adatok	32
2.4.1. Adatmátrix	33
2.4.2. Adattípusok, mérési skálák	34
2.4.3. Transzformációk, származtatott változók	39
2.4.4. Hiányzó értékek	45
2.4.5. Kiugró értékek	48
3. Egy kis valószínűségszámítás	51
3.1. Események, valószínűség	51
3.2. Odds és logit	56
3.3. Relatív kockázat és esélyhányados	58
3.4. Valószínűségi változók	60
3.4.1. Valószínűségi változók függetlensége	67

3.5.	A statisztikában leggyakrabban használt eloszlások	68
3.5.1.	A hipergeometrikus és a binomiális eloszlás	68
3.5.2.	A Poisson-eloszlás	75
3.5.3.	A normális eloszlás	78
3.5.4.	További folytonos eloszlások	80
3.6.	A valószínűségszámítás és a statisztika kapcsolata	82
4.	Leíró statisztika	87
4.1.	Táblázatok és ábrák	87
4.1.1.	Egy változó ábrázolása	88
4.1.2.	Két változó együttesének ábrázolása	98
4.2.	Mérőszámok, statisztikák	103
4.2.1.	Egy változó jellemzése	104
4.2.2.	Két változó közötti összefüggés jellemzése	115
4.2.3.	Asszociációs mértékek	118
4.2.4.	Adattranszformációk hatása a statisztikai mérőszámokra	121
5.	Becslés	123
5.1.	Alapfogalmak	124
5.1.1.	Pontbecslés	124
5.1.2.	Intervallumbecslés	126
5.1.3.	Matematikai formalizmus	129
5.1.4.	A mintaátlag néhány fontos tulajdonsága	131
5.1.5.	Becslés pontossága	132
5.2.	Pontbecslések jósága	135
5.2.1.	Torzítatlanság	135
5.2.2.	Konzisztencia	139
5.3.	Eljárások pontbecslések készítésére	139
5.3.1.	Behelyettesítéses becslés	139
5.3.2.	Maximum likelihood (ML) becslés	140
5.4.	Eljárások konfidencia-intervallumok szerkesztésére	142
5.5.	Több paraméter szimultán becslése	145
5.6.	A szükséges mintaelemszám meghatározása becsléshez	147
6.	Hipotézisvizsgálat	151
6.1.	A statisztikai hipotézisvizsgálat alap gondolata	153
6.1.1.	Az indirekt bizonyítás	154
6.1.2.	A tudomány fejlődése	155
6.1.3.	Nullhipotézis és alternatíva	156
6.1.4.	Döntés a nullhipotézisről	159

6.2.	A hipotézisvizsgálat technikai kérdései	163
6.2.1.	Próbastatisztika	164
6.2.2.	A p -érték meghatározása	168
6.2.3.	Döntés a H_0 -ról p -érték nélkül	170
6.2.4.	Egyszerű és összetett hipotézisek	174
6.2.5.	Próba ereje	177
6.3.	További témák	182
6.3.1.	Többszörös összehasonlítások	182
6.3.2.	Tesztek és konfidencia-intervallumok	184
6.3.3.	A szükséges mintaelemszám meghatározása	185
6.3.4.	Paraméteres és nemparaméteres eljárások	187
7.	Gyakran használt statisztikai próbák	193
7.1.	Várható értékekre (populációátlagokra) vonatkozó próbák	194
7.1.1.	Egy várható érték	194
7.1.2.	Két várható érték, független minták	196
7.1.3.	Két várható érték, párosított minták	200
7.1.4.	Kettőnél több várható érték	202
7.2.	Variációkra vonatkozó próbák	202
7.2.1.	Egy variancia	203
7.2.2.	Két variancia, független minták	203
7.2.3.	Kettőnél több variancia, független minták	205
7.3.	Eloszlásokra vonatkozó próbák	206
7.3.1.	Egy eloszlás: illeszkedésvizsgálat	206
7.3.2.	Két változó együttes eloszlása: függetlenség vizsgálat	213
7.3.3.	Két vagy több eloszlás: homogenitásvizsgálat	218
7.4.	Valószínűségekre (populációbeli arányokra) vonatkozó próbák	220
7.4.1.	Egy valószínűség	220
7.4.2.	Két valószínűség, független minták	223
7.4.3.	Két valószínűség, párosított minták	226
7.4.4.	Kettőnél több valószínűség, független minták	227
7.5.	Mediánokra vonatkozó próbák	227
7.5.1.	Egy medián	227
7.5.2.	Két vagy több medián	229
7.6.	Rangpróbák	231
7.6.1.	Wilcoxon-féle előjeles rangpróba	231
7.6.2.	Mann–Whitney-féle U-próba	235
7.6.3.	Kruskal–Wallis-féle H-próba	238
8.	Korrelációs számítás	243

8.1.	A Pearson-féle korrelációs együttható	244
8.1.1.	Hipotézisvizsgálat a Pearson-féle korrelációs együtthatóra vonatkozóan	246
8.2.	Együtthatók monoton kapcsolatokra	247
8.2.1.	A monoton korrelációs együtthatókra vonatkozó próba .	249
9.	Regressziószámítás	251
9.1.	A regressziószámítás szokásos kérdésfeltevései	252
9.2.	Véletlenség a magyarázó és a függő változóban	253
9.3.	Mikor használjunk korreláció-, illetve regressziószámítást? . .	254
9.4.	Egyszerű lineáris regresszió: I-es modell	255
9.4.1.	Hipotézisvizsgálatok	257
9.4.2.	A determinációs együttható	259
9.4.3.	Predikció a modellben	260
9.5.	Origón átmenő regresszió	263
9.6.	Egyszerű lineáris regresszió: II-es modell	265
9.6.1.	MA-regresszió	265
9.6.2.	SMA-regresszió	265
9.7.	Többszörös lineáris regresszió	268
9.7.1.	Hipotézisvizsgálatok	271
9.8.	További korrelációs mérőszámok	273
9.8.1.	A többszörös korreláció és a determinációs együttható .	273
9.8.2.	A parciális korreláció	274
9.9.	Multikollinearitás	276
9.10.	Regressziós diagnosztika	279
9.10.1.	Az illesztett modell jóságának vizsgálata	279
9.10.2.	Alkalmazhatósági feltételek vizsgálata	280
9.10.3.	Kiugró értékek és torzító pontok	285
9.10.4.	Diagnosztikus ábrák	292
9.11.	Nemlineáris kapcsolatok	293
9.11.1.	Lineárisra visszavezethető regressziók	294
9.11.2.	Példák változók transzformálásával végzett regressziókra	296
9.11.3.	Lineárisra nem visszavezethető regressziók	303
10.	Varianciaelemzés (ANOVA)	313
10.1.	A számítások	315
10.1.1.	Varianciatabla (szórásfelbontás)	319
10.2.	Csoportok páronkénti összehasonlítása	322
10.3.	Többtenyezős varianciaelemzés	325
10.4.	Kísérleti elrendezések	329

10.4.1.	Véletlen blokkos elrendezés	329
10.4.2.	Latinnégyzet-elrendezés	331
10.5.	Az ANOVA diagnosztikája	332
10.6.	Kontrasztok	334
11.	Az általános lineáris modell	337
11.1.	A fejezet példája	337
11.1.1.	A kísérlet rövid leírása	337
11.1.2.	Exploratív elemzések	339
11.2.	Statisztikai modellek	344
11.3.	A modell felírása	345
11.3.1.	Példák különböző modellekre	347
11.3.2.	Faktorok a lineáris modellben	348
11.4.	A lineáris modell paramétereinek becslése	358
11.4.1.	A becsült értékek és a vetítő mátrix	360
11.5.	Hipotézisvizsgálat	360
11.5.1.	A null- és a telített modell	360
11.5.2.	Modell és részmodell összehasonlítása	361
11.5.3.	Az összes magyarázó változó együttes tesztelése	362
11.5.4.	Több változó szimultán tesztelése	364
11.5.5.	Megjegyzések a modellek tesztelésével kapcsolatban	365
11.6.	A lineáris modellek alkalmazhatóságának feltételei	367
11.6.1.	Linearitás	367
11.6.2.	Kiugró és torzító pontok	370
11.7.	Modellválasztás	370
11.7.1.	Mit értsünk a „legjobb” modellen?	371
11.7.2.	A legszűkebb modell, amely nem különbözik szignifikánsan a teljes modelltől	372
11.7.3.	Információs kritériumok	372
11.8.	Modellszelekciós eljárások	374
11.8.1.	Egyenkénti beléptetés	374
11.8.2.	Egyenkénti kihagyás	374
11.8.3.	Váltakozó beléptetés-kihagyás	375
11.9.	Mikor használjuk az aov(), és mikor az lm() függvényt?	376
11.9.1.	Négyzetösszegetípusok	376
11.10.	Többszörös összehasonlítások	378
11.11.	Kontrasztok az általános lineáris modellben	380
11.11.1.	Kontrasztok (általános lineáris hipotézisek) becslése és tesztelése	381

12. Az R-nyelv és -környezet	395
12.1. Telepítés	395
12.2. RGui	397
12.3. Első lépések az R-rel	398
12.3.1. Függvények	400
12.3.2. Csomagok	401
12.3.3. Súlyó	402
12.4. R-munkafolyamat	405
12.5. Adatok olvasása és írása	406
12.5.1. Munkakönyvtár	406
12.5.2. Adatok olvasása	406
12.5.3. Adatok kiírása	407
12.6. Adattároló objektumok	409
12.6.1. Vektor	409
12.6.2. Mátrix	410
12.6.3. Data frame-ek	411
12.6.4. Lista	412
12.6.5. Hivatkozás az objektumok elemeire	413
Függelék	
A. Konfidencia-intervallumok képletei	421
A.1. Normális eloszlású változó átlaga	421
A.2. Két normális eloszlású változó átlaga közötti különbség	422
A.3. Normális eloszlású változó varianciája, illetve szórása	425
A.4. Valószínűség (populációbeli arány)	426
A.4.1. Wald-féle intervallum	427
A.5. Két valószínűség különbsége	429
A.6. Relatív kockázat	430
A.7. Esélyhányados	431
B. Statisztikai táblázatok	433
Irodalomjegyzék	443
Tárgymutató	447