Robusztus megközelítés a figyelmi torzítás kutatásában (folyamatban lévő kézirat)

Smahajcsik-Szabó Tamás

# Bevezető

##### Winsorizálás vagy trimmelés?

A klasszikus parametrikus módszerek null-hipotézis tesztelésénél az előfeltételek sérülése hamis következtetésekre, így a null-hipotézis téves elvetéséhez vezet (**Erceg-Hurn & Mirosevich, 2008; Wilcox & Keselman, 2003; Wilcox et al, 2008**). A normalitás sérülése továbbá hamis hatásméret mutatókat, illetve pontatlan konfidencia intervallumokat eredményez. A kutatásokba bevont adatok nagy része ritkán mutat normális eloszlást, gyakoriak a többszörös modalitást, bizonyos irányú ferdeséget mutató, aszimmetrikus megoszlások. A reakció-idő méréséből származó adatok tendenciaszerűen nem-normális megoszlásúak. A szóráshomogenitás variancia-arányokkal (*variance ratio*, VR) történő összevetése rávilágított, hogy különböző minták szórásai ritkán egyeznek, azaz az 1:1 arány kívánalma ritkán teljesül. A heteroszkedasztikusság jelenléte ugyanakkor nem feltétlenül mérési feltételekből, kísérleti elrendezésekből, vagy mintavételi sajátosságokból ered; bizonyos létező faktorok (nem, életkori csoportok) alapján képzett csoportok között természetszerűen különbség van a variancia tekintetében. Ugyanígy, random csoportok esetében is adódhat szórás szintjén leképezhető eltérés a kísérleti hatás révén (például depresszió kezelésének új módszere).

A klasszikus előfeltevések sértése jelentősen megnöveli, vagy adott esetben csökkentheti az I. fajú hiba elkövetésének kockázatát a ténylegesen számolt p-értékhez képest (melyet módszertanilag 5%-on vagy alatta várunk el). A nem-normális eloszlású és eltérő szórású mintákon nyert p-értékek ezért nagyon pontatlanok lehetnek. Még a normalitás teljesülése mellett is, a szóráshomogenitás sérülése jelentősen csökkenti a parametikus módszerek erejét, mely orvosolható lett volna olyan modern módszerek használatával, melyekre **Wilcox** mint „*robust statistics*” utal (**2003, 2008, 2017**). A gyűjtőfogalom olyan módszereket fog át, melyek az I. fajú hiba rizikóját annak elvárt értékénél képesek tartani, és megőrzik az elemzés erejét, nem-normális és heteroszkedasztikus adatokon is. A klasszikus, parametrikus próbák robusztussága korlátozott, és inkább feltételek között érvényesülő, érzékeny kivétel, mintsem szabály. Noha adottak már klasszikus, non-parametrikus módszerek, ezek jórészt a parametrikus módszerekkel együtt születtek; robusztusságuk gyenge, ha heteroszkedasztikus adatokon végzik őket; és többnyire főhatások vizsgálatára alkalmasak (interakciók feltérképezésére modern utódaik inkább használhatóak). A modern robusztus módszerek célja egyaránt jól teljesíteni a klasszikus előfeltevések megléte és sérülése esetén is. E módszerek alapvetően abban mások, hogy új alternatívát kínálnak a regressziós együtthatók, a lokációs számítások (pl. átlag) és a kapcsolati mutatók (pl. *Pearson*-féle korrelációs együttható) helyett.

A heteroszkedasztikus megoszlások és a normalitás sérülése esetére kínál megoldást a **trimmelt átlagokkal** és a **winsorizált varianciával** való munka. Trimmelt adatokkal végzett statisztikai elemzések esetén, a hipotézis szerinti populációs átlag is trimmelt formában értendő. A winsorizálás során a trimmelt alsó és felső értékeket a még nem trimmelt legalacsonyabb és legmagasabb érték helyettesíti, mely un. winsorizált adateloszláshoz vezet, melyből winsorizált átlag és winsorizált variancia már a hagyományos képletekkel számítható. Alacsony elemszámnál a **bootstrap resampling** értékbehelyettesítésen alapuló módszere is ajánlott. Fontos, és a jelen munkában központi kérdés az outlier-ek szerepe, hiszen már egészen kis számú outlier is befolyásolni képes az átlagot, és a minta szórását, ezáltal hatva a későbbi statisztikai számításokra.

##### Az ABV-index

A figyelmi folyamatok vizsgálatánál, így a jelen kutatásban is használt *dot probe* tesztek révén nyert reakcióidő adatok azt mutatják, egyes csoportok, így *ADHD-s gyerekek* (**Epstein et al, 2011**), *PTSD-vel diagnosztizált veteránok* (**Iacoviello et al, 2014**) esetén a személyen belüli variancia (az un. *attention bias variability*) magasabb releváns ingerek bemutatásakor (**Price at al, 2015**). Az index jó stabilitással bír, így e tesztek esetén reliabilitási mutatóként is használható.

# Módszer

A jelen munkában az Érzelmi Arcok Figyelmi Próba segítségével nyert reakcióidőkön végzett robusztus adatelemzés eredményeit mutatom be, 333 serdülőtől (átlagéletkor: 16,09; szórás: 1,20) származó adatokkal. Az adatfelvétel során 144 próba esett egy vizsgálati személyre, ebből személyenként 57 inger kongruens volt, 57 inkongruens. A 144 próba emellett harmadolva lett az érzelmi ingerek tekintetében is, 38-38 próba esett, szomorú, boldog és mérges ingerekre.

Az eredeti adatok leíró statisztikája

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA’s |
| 201 | 345 | 390 | 422.4672 | 453 | 8935 | 488 |

##### Az adattisztítás módszertana; *outlierek*

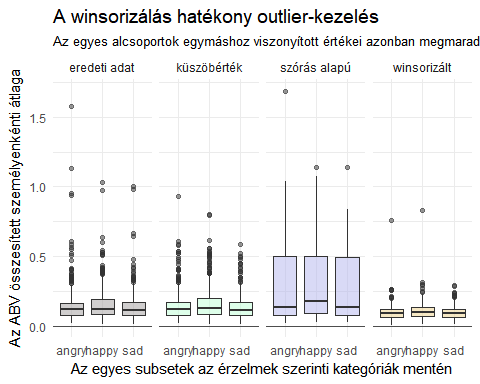
Az outlierek kezelésének három főbb módszerét alkalmaztuk, mely a szórás alapján történő kiszűrést, az előre megadott reakcióidőre vonatkozó küszöbértékek szerinti elhagyást, és az újraskálázást (azon belül is a winsorizing módszerét) foglalja magában. Az adatelemzésre az R szoftvert használtuk. A winsorizálás módszere követte **Price** módszertanát (**2015**), azaz személyenként az adatok 25. és 75. percentilisén kívülre eső végső tartományaiban az interquartilis terjedelem (*interquartile range*, IQR) másfélszeres értékét is alul, illetve felülmúló outliereket skáláztuk újra. Például ha a 25. percentilis 600 ms, az IQR pedig 200 ms volt, ez esetben az eloszlásnál a 300 ms alatti értékeket winsorizáltuk. Az újraskálázás során az adott outlier a még az elfogadható tartományba eső legkisebb, illetve legnagyobb értéket kapta meg. A példánál maradva, 300 ms alatt minden érték 300 ms-re lett átkódolva. Az ABV-index számítása szintén követte a fenti munkában tárgyalt módszertant, azaz a 114 próbát un. *bin*-ekre osztottuk, 29 bin alapján 4 tartománya volt képezhető a próbáknak. Minden személyre binenként számoltunk un. ICvC-indexet, mely az inkongurens ingerekre adott reackióidő és a kongruensekre adott válaszok különbségét jelentette. Következő lépésben a binenként ily módon nyert indexek standard szórését vettük, majd elosztottuk binenként a reakcióidő átlagával. Így egy személyen belüli, különböző binenként számított variabilitás indexet kaptunk, mely az elvárás szerint magas volt, ha a személy figyelme letapadt vagy adott inger elvonta azt. A fenti munkától eltérően azonban, az ABV-indexeket személyenként és binenként az érzelmi ingerek tekintetében is vettük, arra keresve a választ, hogy adott érzelmi tartalmú bemutatott inger miként hat a figyelmi teljesítményben megmutatkozó teljesítmény tekintetében.

Az egyes outlier-kezelő módszerekkel nyert reakcióidők leíró statisztikája

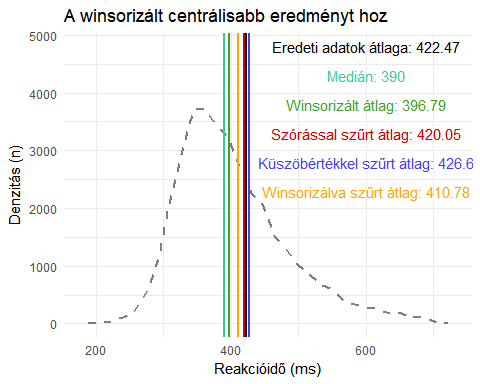
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| method | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA’s |
| winsorizing | 203 | 345 | 390 | 410.7822 | 451 | 3552 | 488 |
| SD based filtering | 202 | 343 | 389 | 420.0452 | 461 | 4216 | 6888 |
| RT threshold applied | 301 | 353 | 395 | 426.5969 | 457 | 2492 | 2576 |

# Eredmények

Az eredeti adatokon képzett ABV-index értékeit összevetve az egyes outlier-kezelési technika révén megtisztított adatok képzettekkel, az alábbi trendek figyelhetők meg.



Az előzetes eredmények arra utalnak, a 25., 75. percentilisekhez tartozó reakcióidőket és 1.5-szörös IQR-t felhasználva (**Price et al, 2015**) az adatok winsorizálásához, a módszer, összehasonlítva az eredményeket az eredeti, a szórás alapján szűrt és az előre meghatározott határértékek mentén történő tisztítással nyert adatokon végzett számításokkal, jelentősen képes javítani egy adatmegoszlást az outlierek tekintetében. Fontos ugyanakkor, hogy látszólag az outlierek kiszűrésében hatékony módszer a rögzítetten adott szórásértékek mentén történő adattisztítás, noha a csoportközi különbségeket arányaiban megőrzi, jelentősen módosítja is az adatmegoszlást, adatveszteséget okoz. Mint később látni fogjuk, a centrális pont (*“measure of location”*) kérdésében, a winsorizált átlag az adateloszlás szélső tartományai által kevésbé befolyásolt, stabilabb, a mediánhoz közelítő értéket hoz, a szórás révén tisztított adatokon képzett aritmetikai átlaghoz képest. Az eredmény összefügg azzal, amit **Wilcox (2017)** felvázol: nem szimmetrikus megoszlások esetén a winsorizált átlag az adatok centrális lokációját emeli ki, míg a hagyományos átlag torzít. (Szimmetrikus megoszlás esetén értelemszerűen a kettő megegyezik). Ha egy adateloszlás tipikus tagját keressük, a hagyományos aritmetikai átlag eszerint azért nem lesz pontos, mert egy az eloszlás száraira eső kisebbség nagyban képes megváltoztatni annak értékét. A *winsorizálás* ebben az értelemben azáltal fektet nagyobb figyelmet egy megoszlás központi értékeire, hogy határértékek mentén transzformálja, esetünkben 0-ra állítja az adott kvartiliseken kívülre eső értékeket.



Jól látható, hogy a (25%-os trimmeléssel készült) winsorizált átlag az adateloszlás centrális értékeire fókuszálva határoz meg lokációt (zöld), míg a hagyományos átlagot nagyban vezérlik az eloszlás szélső quartiliseinek értékei (az egyes átlagok jobb kivehetősége érdekében az ábrán 700 ms alá szűrtük le a reakcióidőket). Mind a szórással szűrt adatokon képzett, mind pedig a rögzített küszöbértékkel szűrt adatokon képzett aritmetikai átlag a teljes, eredeti adatokon kapott átlaghoz közelít, ennek értelmében hasonlóképp érzékeny maradhatott a szélsőbb, extrémebb adatok vonatkozásában. A winsorizálással szűrt adatok átlaga 410,78 ms, melyet az adatok újrakódolása is okoz, küszöbértékeken túli adatokat átkódoltuk **Price (2015)** módszertana szerint.

# Megbeszélés

Noha a három standard szórásnyi sugárban elvégzett adattisztítás módszere látványos eredményt hozott az **első ábra** szerint, az átlagok tekintetében hasonlóképpen teljesít ez a módszer mintha az eredeti adateloszlást használnánk fel végül az elemzésekhez. A két átlag közti különbség alig eredményezett nagyobb különbséget mint 2 ms. A szórás menti szűrés ugyanakkor az adateloszlást is megváltoztatta, így használata nem javasolt. A rögzített küszöbértékek mentén történő szűrés sem az outlierek, sem az átlagok tekintetében nem bizonyult hatékony módszernek - megfontolandó a küszöbök vizsgálathoz igazítása a teljes adateloszlás fényében. A winsorizálás robusztusabb eredményeket hozott, tisztított az adatokon: mind az eredeti adatokon képzett 25%-os trimmeléssel képzett winsorizált átlag, mind pedig a winsorizálással tisztított és újraskálázott adatok aritmetikai átlaga is közelebb áll a mediánhoz, mely egy robusztusabb, a normalitást sértő adatmegoszlások esetén az eddigi kutatások szerint pontosabb elemzéseket lehetővé tevő módszert valószínűsít, a konfidencia intervallumok tekintetében úgy, mint ahogy a statisztikai erő vonatkozásában is. Az egyes adatkezelési módok összevetésének további módja az ABV indexek lehetséges kapcsolatainak feltárása például egy rumináció skálán elért pontszámokkal; illetve az egyes érzelmek által képzett csoportok közti statisztikai különbségek feltárása módszerenként.

# Felhasznált irodalom

**Erceg-Hurn, D.A. & Miresovich V. M.** (2008). “Modern Robust Statistical Methods - An easy way to maximize the accuracy and power of your research.” *American Psychologist* 63.(7.): 591 - 601.

**Epstein, J. N., Langberg, J. M., P. J. Rosen, A. Graham, M. E. Narad, T. N. Antonini, W. B. Brinkham, T. Froehlich, J. I. Simon, M. Altaye** (2011). “Evidence for higher reaction time variability for children with ADHD on a range of cognitive tasks including reward and event rate manipulations.” *Neuropsychology* 25.(4.): 427 - 441.

**Iacoviello, B. M., G. Wu, R. Abend, J. W. Murrough, A. Feder, E. Fruchter, Y. Levinstein, I. Wald, C. R. Bailey, D. S. P., A.Neumeister, Y. Bar - Haim, D. S. Charney** (2014). “Attention Bias Variability and Symptions of Posttraumatic Stress Disorder.” *Journal of Trauma and Stress* 27.(2.): 232-239.

**Price, R. B., J. M. Kuckertz, G. J. Siegle, C. D. Ladouccer, J. S. Silk, N. D. Ryan, R. E. Dahl, N. Amir** (2015). “Empirical recommendations for improving the stability of the dot-probe task in clinical research.” *Psychological Assessment* 27.(2.): 365-376.

**Wilcox, R. R., H. J. Keselman, L. M. Lix, J. Algine, K. N. Deering** (2008). “A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes.” *Psychological Methods* 13.(2.): 110-129.

**Wilcox, R. R., H. J. Keselman** (2003). “Modern robust data analysis methods: Measures of central tendency.” *Psychological Methods* 8.(3.): 254 - 274.

**Wilcox, R. R.** (2017). Introduction to Robust Estimation and Hypothesis Testing, *Elsevier*.