Robustnost statističkih postupaka i robustni ocenitelji parametara lokacije i skale

1. Robustnost statističkih postupaka

Pojam robustnosti statističkih postupaka u ovom tekstu razmotrićemo imajući u vidu sledeća tri aspekta:[1]

1. robustnost statističkih testova za testiranje hipoteza;
2. robustnosti parametara i
3. robustnost ocenitelja parametara.

Robustnost statističkih testova za testiranje hipoteza

Matematički (probabilistički) modeli klasičnih statističkih postupaka koji se najčešće koriste u psihologiji (t-test, analiza varijanse i regresiona analiza) uključuju pretpostavke za koje se, samim korišćenjem određenog statističkog postupka, implicitno prihvata da *dopustivo dobro aproksimiraju* realnost ispitivanih fenomena. Naglasak u tome je na *aproksimativnosti*, tj. na istraživačima dobro poznatoj činjenici da realnost psiholoških fenomena manje ili više odstupa od matematičkih modela koji su u osnovi određenog statističkog postupka. Statistička sredstva predstavljaju jedno od osnovnih "oruđa" (srećom, ne jedino) istraživačima u psihologiji. Stoga, nepostojanje "dopustivo dobre aproksimacije" proučavane realnosti pretpostavkama matematičkog modela na kojima se zasniva određeni statistički postupak može imati dalekosežne posledice na razvoj psihološke nauke. Klasični statistički postupci koji se najčešće primenjuju u analizama podataka psiholoških istraživanja izvedeni su postuliranjem određenih idealizovanih uslova od kojih je "normalnost raspodele varijable X u populaciji" sastavni deo matematičkih modela većine klasičnih parametrijskih testova. (Normalna raspodela je, na taj način, psihološkim rečnikom rečeno, postala vrsta "statističkog arhetipa" za ogromnu većinu populacija na čijim uzorcima su primenjivane statističke metode).[2]

---

[1] Termin robustan potiče od latinske reči *robustus* što znači jak, snažan, krepak, čvrst.

[2] Ilustracije radi, nulta distribucija uzorkovanja t-statistika koji se uobičajeno koristio za testiranje pretpostavke da dve subpopulacije (npr. muškarci i žene) imaju "prosečno gledano" jednaku prostornu sposobnost (H_0: $\mu_1 = \mu_2$) izvedena je pod sledećim pretpostavkama:

a) varijabla X (prostorna sposobnost u našem slučaju) ima normalnu funkciju gustine u svakoj od subpopulacija sa parametrima μ i σ^2 (subpopulacija muškaraca: $X \sim N(\mu,\sigma^2)$; subpopulacija žena: $X \sim N(\mu,\sigma^2)$).

b) varijabla X ima jednake varijanse u dvema subpopulacijama: $\sigma^2{}_1 = \sigma^2{}_2 = \sigma^2$;

c) subpopulacije imaju jednake aritmetičke sredine na varijabli X: ($\mu_1 = \mu_2 = \mu$);

d) opservacije su nezavisne, a uzorci jedinica posmatranja jednostavni slučajni uzorci;

Statistik za testiranje nulte hipoteze u ovom slučaju, dobro poznati t-statistik, definiše se na sledeći način:

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{SE_{M_1\text{-}M_2}}$$

U obrascu za t statistik M_1-M_2 je razlika između aritmetičkih sredina uzoraka, a $SE_{M_1 - M_2} = \sqrt{S^2 (\frac{1}{n_1} + \frac{1}{n_2})}$ je ocena standardne greške za razliku između aritmetičkih sredina. Pri

U kojoj meri će disparatnost skupa pretpostavki koje su u osnovi određenog statističkog postupka i ispitivane realnosti imati (negativni) efekat zavisi, pre svega, od toga kako određena statistička metoda funkcioniše u različitim "nemodelskim", "neidealnim", tj. realnim uslovima. Osetljivost, tj. robustnost statističkih postupaka na aproksimativnost u procesu njihove primene, tj. na veliki broj malih ili mali broj velikih odstupanja od pretpostavki na kojima se ovi postupci zasnivaju predstavlja, po mišljenju pisca ovog teksta, jedno od najvažnijih svojstava statističkih postupaka. Poznavanje tog svojstva statističkih postupaka koji se primenjuju u analizama realnih podataka od kritičnog je značaja jer od toga zavisi valjanost naučnih zaključaka do kojih se dolazi njihovim korišćenjem.

Sam termin *robustnost* (i to pod navodnicima) prvi je upotrebio statističar Box (G. E. P. Box) 1953. godine u članku koji je upravo bio posvećen ispitivanju osetljivosti statističkih testova za poređenje varijansi dveju populacija na neispunjenost pretpostavke o normalnosti distribucija u populaciji. Ovim terminom Boks je označio relativnu neosetljivost statističkih testova za testiranje značajnosti razlika između aritmetičkih sredina na određeni tip ne-normalnosti distribucija u populaciji. Vremenom je upotreba termina robustnost u navedenom smislu proširena tako da obuhvati stepen osetljivosti statističkih testova za testiranje hipoteza na neispunjenost bilo koje pretpostavke koja čini sastavni deo matematičkog modela koji je u osnovi izvođenja samog testa. Dakle, statistički postupak ili test koji funkcioniše "razumno dobro" u uslovima blage ili umerene narušenosti pretpostavki probabilističkog modela na kojima se zasniva njegovo matematičko izvođenje smatra se robustnim statističkim postupkom. "Razumno dobro" funkcionisanje robustnog statističkog testa podrazumeva da distribucija uzorkovanja statistika za testiranje nulte hipoteze ne odstupa bitno u uslovima blage (i umerene) narušenosti pretpostavki probabilističkog modela, kao i to da snaga statističkog testa takođe nije bitno narušena. Drugim rečima, verovatnoća grešaka I i II tipa u statističkom zaključivanju korišćenjem <u>robustnog</u> statističkog testa u uslovima narušavanja pretpostavki probabilističkog modela na kojem se test zasniva veoma je blizu nominalnim vrednostima koje slede na osnovu probabilističkog modela. Međutim, u odsustvu

---

tom se ocena varijanse za kvantitativnu varijablu u populaciji, u oznaci S^2, dobija na osnovu varijansi uzoraka na sledeći način:

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Ukoliko su uslovi pod a, b, c i d iz probabilističkog modela ovog postupka ispunjeni tada, kao što je poznato, statistik t kao slučajna varijabla ima Studentovu distribuciju uzorkovanja koja je definisana parametrom ν, tj. stepenima slobode (pri čemu je $\nu = n_1 + n_2 - 2$, a n_1 i n_2 predstavljaju veličine uzoraka). *Ukoliko su*, dakle, *uslovi uključeni u probabilistički model koji stoji u osnovi* prikazanog postupka t-testa *dopustivo dobra aproksimacija* realnosti ovaj postupak se može primeniti kao valjan postupak za testiranje nulte hipoteze o jednakosti aritmetičkih sredina muškog i ženskog dela ljudske populacije u pogledu varijable X, tj. prostorne sposobnosti u našem slučaju. U tom će slučaju verovatnoća da t statistik uzme na slučajnim uzorcima neku od svojih vrednosti u određenom intervalu – a koja se dobija na osnovu njegove nulte distribucije uzorkovanja – biti adekvatna. Od tačnosti ocene ove verovatnoće zavisiće i da li će odluka o odbacivanju ili neodbacivanju nulte hipoteze, pa i krajnji supstantivni, tj. naučni zaključak o ispitivanom fenomenu biti utemeljeni na tačnim ili pogrešnim brojkama! Naravno, ovde se prirodno nameću barem dva pitanja: 1. Kakve su posledice neispunjenosti uslova probabilističkog modela, tj. da li će zaključci do kojih ćemo doći primenom statističkog postupka – ako pretpostavke na kojima se on zasniva ne aproksimiraju dopustivo dobro realnost – biti nužno pogrešni? 2. Da li su sve pretpostavke statističkog modela podjednako važne, tj. da li se neispunjenost nekih od uslova može zanemariti bez bojazni po ispravnost donetih zaključaka?

preciznog kvantitativnog kriterijuma za dozvoljeno odstupanje verovatnoća dveju vrsta grešaka u statističkom zaključivanju pri narušenosti pretpostavljenih uslova probabilističkog modela nije moguća ni objektivna ocena robustnosti statističkog testa, a zaključci o (ne)robustnosti statističkog testa mogu biti veoma subjektivni i varijabilni. Postoje pokušaji da se ovaj problem reši definisanjem granica u kojima se – u uslovima narušenosti pretpostavki na kojima test počiva – može kretati aktuelna verovatnoća greške I tipa za dati nominalni nivo značajnosti (uobičajeno u oznaci α) pri primeni statističkog testa. Bredli (J. V. Bradley) je 1978. godine predložio dva kriterijuma koji se razlikuju po stepenu strogosti: prema strožem kriterijumu statistički test je robustan ako je – pri narušenosti pretpostavki na kojima test počiva – stvarna verovatnoća greške pri odbacivanju tačne nulte hipoteze u rasponu $\alpha \pm 0.1\alpha$, pri čemu je α nominalni nivo značajnosti; prema blažem kriterijumu, koji Bredli ujedno smatra najliberalnijim kriterijumom koji se "može smatrati ozbiljnim", test se može smatrati robustnim ukoliko se stvarna verovatnoća greške I tipa u uslovima narušenih pretpostavki testa kreće u granicama od α - 0.5α do $\alpha + 0.5\alpha$.

Ispitivanja robustnosti statističkih testova izvode se primenom simulacionih eksperimenata u kojima se porede distribucije uzorkovanja statistika za testiranje hipoteza pri uzorkovanju iz "idealnih" populacija (pretpostavljenih statističkim modelom postupka) i distribucije uzorkovanja za isti statistik kada se uzorkovanje vrši iz populacija koje nisu sasvim u skladu sa pretpostavljenim probabilističkim modelom statističkog testa.

Robustnost parametara

Robustnost određenog parametra θ može se definisati na sledeći način: parametar $\theta(F)$ (pri čemu je F funkcija distribucije, tj. kumulativna distribucija verovatnoća) robustan je ukoliko brzina njegove promene, pri proizvoljno malim promenama funkcije distribucije F, ne može biti proizvoljno velika. Drugim rečima, parametar $\theta(F)$ – pri čemu je F funkcija distribucije – predstavlja robustan parametar ako male perturbacije u F ne dovode do velikih promena u $\theta(F)$. Dakle, male promene u F ne bi trebalo da odvedu $\theta(F)$ ka proizvoljno velikim vrednostima. Minimalni zahtev koji se postavlja nekom parametru da bi se on mogao smatrati robustnim jeste da *tačka otkaza* parametra bude *veća od nule* i da njegova *funkcija uticaja* bude *ograničena*. Tačka ili granica otkaza (eng. breakdown point; breakdown bound) nekog parametra $\theta(F)$ za distribuciju F predstavlja najmanju vrednost ε za koju parametar $\theta(F_\varepsilon)$ može postići proizvoljno veliku vrednost. Vrednost ε predstavlja meru distance ili razlike distribucije F i distribucije F_ε. Za aritmetičku sredinu, varijansu i koeficijent linearne korelacije tačka otkaza jednaka je nuli, tj. najmanjoj mogućoj vrednosti za tačku otkaza, pa se ovi parametri prema njihovoj tački otkaza ne mogu smatrati robustnima. Drugim rečima, za infinitezimalnu distancu ε između dveju distribucija, tj. njihovih funkcija F i F_ε, ovi parametri mogu uzeti vrednosti koje se razlikuju za proizvoljno veliki broj. Funkcija uticaja (eng. influence function), pak, predstavlja graničnu vrednost količnika:

$$\frac{\theta(F_\varepsilon) - \theta(F)}{\varepsilon}$$

sa približavanjem ε ka nuli zdesna. Tako je, na primer, funkcija uticaja $IF_{\mu,F}(x)$ za aritmetičku sredinu kao parametar ($\mu = \int x dF(x)$) jednaka x - μ ($x \in R$) pa – budući da je njena funkcija uticaja neograničena po x – aritmetička sredina prema ovom pokazatelju ne spada u robustne parametre. Isto tako, varijansa distribucije F ($\sigma^2 = \int (x - \mu)^2 dF$), ako varijansa postoji i za poznatu aritmetičku sredinu distribucije μ, ima funkciju uticaja oblika $(x - \mu)^2 - \sigma^2$ ($x \in R$) pa, očigledno, zbog neograničenosti ove funkcije uticaja ni varijansa nije robustan parametar. U vezi s tim, postavlja se, dakle, pitanje opravdanosti korišćenja ovih parametara za opis tipične osobe ili varijabilnosti u populaciji za varijable za koje pretpostavka o normalnosti ne izgleda održiva. Slični problemi postoje i u vezi sa korišćenjem koeficijenta linearne korelacije kao parametra u populacijama za koje pretpostavka o bivarijacionoj normalnoj distribuciji ne izgleda opravdana.

Moguće je postaviti pitanje da li je izbor parametara koji preovlađuje u psihološkim istraživanjima – sa stanovišta robustnosti samih parametara – "najsrećniji" izbor. Mogli bismo reći da je sam izbor "nerobustnih" parametara kojima se najčešće bave istraživači u psihologiji pod uticajem teorijskih pretpostavki o "sveprisutnoj" normalnosti raspodela psiholoških varijabli u populaciji, tj. "mita o normalnosti" koji opstaje uprkos tome što distribucije psiholoških varijabli izgleda skoro nikada nisu normalne! U prilog tome govore i rezultati istraživanja Mićerija (Micceri) iz 1989. godine u kojem su ispitivana svojstva 440 empirijskih distribucija sa 46 psiholoških testova (uglavnom testova sposobnosti, postignuća i ličnosti) koje su dobijene na velikim uzorcima: ne samo da su *sve* ispitivane distribucije bile nesaglasne sa pretpostavkom o normalnosti u populacijama iz kojih su uzorci dobijeni, već je samo 3% realnih distribucija podataka koje su analizirane u ovom istraživanju bilo simetrično! Dakle, rezultati Mićerijevog istraživanja *ne* sugerišu da, barem kada je o psihološkim varijablama reč, "Bog voli normalnu krivu".

Robustnost ocenitelja parametara

Pored testova statističkih hipoteza statistički postupci uključuju i mnoge statistike ocenitelje koji treba da posluže kao kvalitetna ocena populacijskih numeričkih svojstava, tj. parametara. Uzoračka stabilnost (varijansa) ovih ocenitelja može biti pod velikim uticajem distribucionih svojstava varijabli kojima se istraživači bave ali i idiosinkratičnih svojstava uzorka na kojem se primenjuju. Tako, aritmetička sredina uzorka predstavlja najprecizniju ocenu parametra, tj. aritmetičke sredine populacije (predstavlja tzv. nepristrasni ocenitelj sa minimalnom varijansom) ukoliko se uzorkovanje vrši iz populacije sa *normalnom* funkcijom gustine. Ukoliko se, pak, uzorkovanje vrši iz distribucije koja odstupa od normalne funkcije gustine tada aritmetička sredina uzorka kao ocenitelj parametra gubi ova svojstva. Na primer, ako se uzorkovanje vrši iz "mešane" normalne raspodele – raspodele koja predstavlja konvoluciju ('mešavinu') više normalnih raspodela i pokazuje, prema tome, malo odstupanje od normalnosti – medijana uzorka kao ocenitelj parametra ima znatno manju varijansu od uzoračke aritmetičke sredine. Pored toga, dobro je poznato da prisustvo malog broja nesaglasnih vrednosti – autlajera (eng. outlier) u uzorku opservacija koje služe za računanje aritmetičke sredine može da dovede do potpunog otkaza (eng. break down) ovog ocenitelja, tj. do njegove neupotrebljivosti.

Postupak ocenjivanja ili ocenitelj (eng. estimator) parametra smatra se robustnim ukoliko nije osetljiv na "mala odstupanja" od idealizovanih pretpostavki za koje je optimizovan. Pri tome se pod malim odstupanjima podrazumeva mali broj velikih ili veliki broj malih odstupanja. Robustnost ocenitelja obuhvata dva svojstva: *robustnost u efikasnosti* i *rezistentnost*.

Robustnost u efikasnosti datog ocenitelja podrazumeva da je varijansa nepristrasnog ocenitelja ili prosečna kvadratna greška za pristrasni ocenitelj blizu minimalne vrednosti za različite distribucije.[3] Sa stanovišta analize realnih podataka ovo svojstvo ocenitelja obezbeđuje njegovo dobro funkcionisanje u uslovima kada uzorci potiču iz populacija čije distribucije imaju "gušće" krajeve u odnosu na normalnu raspodelu ili, pak, iz populacija čija distribuciona svojstva nije moguće precizno definisati.

Svojstvo rezistentnosti ocenitelja definiše se preko tačke ili granice otkaza za konačni uzorak (eng. finite sample breakdown point ili finite sample breakdown bound), tj. najveće proporcije opservacija koje se mogu neograničeno menjati a da promena u vrednosti ocenitelja bude ograničena. Tačka otkaza za konačni uzorak ocenitelja T_n definisana je formalno na sledeći način:

$$\varepsilon_n^*(T_n, Z_n) = \min\left\{\frac{k}{n}; \sup_{Z_n'}\left\|T_n(Z_n') - T_n(Z_n)\right\| = \infty\right\},$$

pri čemu Z_n' obuhvata raspone podataka na svim skupovima koji se dobijaju zamenom k vrednosti u Z_n arbitrarnim vrednostima. Prema tome, tačka otkaza je najmanja proporcija neograničeno izmenjenih podataka koja može odvesti ocenitelj izvan svih granica, tj. dovesti do toga da se vrednost ocenitelja drastično promeni, da postane proizvoljno velika ili mala. Aritmetička sredina, varijansa i koeficijent linearne korelacije kao ocenitelji imaju najmanju moguću vrednost tačke otkaza za konačni uzorak, tj. $\frac{1}{n}$, pri čemu je n veličina uzorka. Očigledno, samo jedan neograničeno izmenjen podatak može da dovede do otkazivanja ovih ocenitelja, tj. može tako dobijene ocene učiniti neupotrebljivim. Granična vrednost tačke otkaza za konačni uzorak za neki ocenitelj tipično je jednaka tački otkaza parametra koji se ocenjuje. Budući da je $\lim_{n\to\infty}\frac{1}{n} = 0$, tačka otkaza parametara koji se ocenjuju navedenim oceniteljima jednaka je nuli, tj. najnižoj mogućoj vrednosti. Prema tome, ocenitelj je utoliko robustniji ukoliko je njegova tačka otkaza veća od $\frac{1}{n}$. Ako se posmatra iz perspektive analize realnih podataka robustnost ocenitelja definisana teorijski preko tačke otkaza govori o *rezistentnosti* ocenitelja *na autlajere*, tj. nesaglasne vrednosti.

Autlajeri – nesaglasne vrednosti (eng. outliers)

Autlajeri ili nesaglasne vrednosti su one vrednosti u skupu podataka koje su neobičajeno daleko ili koje su veoma različite od glavnine podataka. Prisustvo autlajera u uzorku podataka najčešće se tumači kao posledica grešaka merenja, grešaka u unosu podataka ili (što je sa statističkog aspekta najvažnije) kao odraz intrinzične varijabilnosti izvora podataka. Određenje nesaglasnih vrednosti kao odraza intrinzične varijabilnosti podrazumeva da postoji statistički model o izvoru koji generiše podatke, tj. o prirodi distribucije varijable u populaciji iz koje je generisan uzorak. Autlajer bi se u tom slučaju mogao posmatrati kao opservacija koja je uzorkovana iz druge populacije u odnosu na onu iz koje je glavnina podataka.

[3] Pristrasnost, varijansa i prosečna (srednja) kvadratna greška ocenitelja definisani su u L. Tenjović (2002). Statistika u psihologiji – priručnik, glava VI.1. Statističko ocenjivanje parametara.

Moguće je, isto tako, postojanje nesaglasnih vrednosti u podacima tumačiti i u smislu uzorkovanja iz populacije čija distribucija ima "gušće krajeve" od pretpostavljene raspodele. Autlajeri u bivarijacionom slučaju mogu predstavljati opservacije koje odstupaju od glavnine podataka po veličini ("autlajeri pomerenosti" – eng. shift outliers) ili po strukturi. Dok prva vrsta autlajera prati osnovnu strukturu povezanosti među varijablama i predstavlja prosto daleke opservacije koje prate trend glavnine podataka, dotle druga vrsta nesaglasnih vrednosti upravo narušava strukturu povezanosti između varijabli. Ukoliko su, na primer, dve osobine ličnosti u pozitivnoj linearnoj vezi, autlajer "po veličini" predstavljao bi ispitanika koji ima neuobičajeno visoke (ili niske) rezultate na obema crtama, dok bi autlajer "po strukturi" bila jedinica posmatranja sa neobično visokim rezultatom na jednoj, a neuobičajeno niskim rezultatom na drugoj osobini.

Otkrivanje autlajera u podacima neobično je važno. Pre svega, na taj način moguće je otkriti grube greške u merenju ili unosu podataka koje bi, ukoliko ostanu neuočene, mogle potpuno kompromitovati zaključke istraživanja. S druge strane, prave nesaglasne vrednosti, tj. rezultati izrazito netipičnih jedinica posmatranja mogu – budući da se u analizama podataka psiholoških istraživanja još uvek skoro isključivo koriste statistički postupci koji su nepostojani, tj. nerezistentni na autlajere – dovesti do nesrazmernog izobličenja rezultata i tako dovesti do neadekvatnih ocena ključnih parametara koji su od interesa za istraživača. Zavisno od toga koji matematički model distribucije, po pretpostavci, dobro opisuje raspodelu jedne ili više varijabli u populaciji postoje različiti statistički kriterijumi i raznovrsni grafički postupci za detekciju autlajera. Najčešće primenjivani klasični statistički kriterijumi za nesaglasne vrednosti u jednodimenzionalnom slučaju jesu postojanje podataka čije odstupanje od aritmetičke sredine je veće od 2.5 ili 3 standardne devijacije (u zavisnosti od veličine uzorka) ili čija je udaljenost od gornje (ili donje) četvrti (eng. uper and lower forth, vrednosti koje grubo odgovaraju percentilima 75 i 25) veća od izraza koji se dobija množenjem međučetvrtnog raspršenja (razlike gornje i donje četvrti) vrednošću 1.5. U novije vreme sve se više koristi "robustni kriterijum" za deklarisanje jednodimenzionalnih autlajera koji se definiše na sledeći način: rezultat x_i je nesaglasna vrednost ako je

$$\boxed{\frac{|x_i - \text{Mdn}|}{\text{MAD}/0.6745} > 2.24}$$

Pri tome, Mdn je medijana, MAD je medijana apsolutnih odstupanja svih rezultata od medijane (eng. Median Absolute Deviation), a konstanta 0.6745 služi za reskaliranje mere MAD tako da u slučaju normalne raspodele MAD može poslužiti kao ocena standardne devijacije populacije.

Uočavanje potencijalnih nesaglasnih vrednosti u univarijacionom i bivarijacionom slučaju relativno je jednostavno i za to je ponekad dovoljno pažljivo i znalačko posmatranje valjanog grafičkog prikaza podataka (npr. kutijasti dijagram – eng. boxplot, dijagram raspršenja i slično).

***

Precizne definicije robustnosti na osnovu funkcije uticaja i tačke otkaza mogu se koristiti u formalnom dokazivanju robustnosti ili nerobustnosti parametra i ocenitelja. Utvrđivanje robustnosti ili nerobustnosti na autlajere za ocenitelje korišćenjem empirijskih studija podrazumeva odabir ili definiciju određenih pokazatelja robustnosti na autlajere. Stepen u kojem je ocenitelj rezistentan na autlajere se potom procenjuje uzimanjem u obzir empirijski dobijenih vrednosti ovih pokazatelja.

2. Robustni ocenitelji lokacije i skale

U poslednjih 50 godina definisan je veliki broj robustnih ocenitelja lokacije (centralne tendencije) i skale (varijabilnosti) koji imaju višu tačku otkaza od aritmetičke sredine i standardne devijacije i, istovremeno, manju osetljivost na autlajere. U ovom tekstu, samo kao ilustraciju, definisaćemo dva jednostavna robustna ocenitelja lokacije: *postriženu aritmetičku sredinu* i *modifikovani M-ocenitelj iz jednog koraka* i jedan ocenitelj skale: *medijansko apsolutno odstupanje.*

Postrižena aritmetička sredina (eng. Trimmed mean)

Postrižena aritmetička sredina, u oznaci $M_{t(\gamma)}$, definisana je na sledeći način:

$$\boxed{M_{t(\gamma)} = \frac{\sum_{i=g+1}^{n-g} x_{(i)}}{n-2g}}$$

Pri tome, $x_{(i)}$ je redosledni statistik (rezultat koji je na *i*-tom mestu po veličini kada se rezultati poređaju od najmanjeg ka najvećem), n je ukupan broj rezultata, γ je proporcija najnižih i najviših rezultata koji se odbacuju, a g je jednako $[\gamma*n]$. Oznaka $[\gamma*n]$ znači da, ukoliko je proizvod $\gamma*n$ decimalan broj, g predstavlja samo njegov celobrojni deo. Na primer, ako je $\gamma*n$ jednako 9.8 tada je g = 9. Dakle, postrižena aritmetička sredina predstavlja aritmetičku sredinu rezultata koji preostanu kada se g najnižih i g najviših rezultata odbace. Najčešće se, pri računanju postrižene aritmetičke sredine uzima da je γ jednako 0.20, tj. odbacuje se 20% najnižih i 20% najviših rezultata.

Modifikovani M-ocenitelj lokacije iz jednog koraka (eng. Modified One-step M-estimator)

Ovaj ocenitelj, u oznaci MOM (akronim engleskog naziva ocenitelja), definisan je na sledeći način:

$$\boxed{MOM = \frac{\sum_{i=i_1+1}^{n-i_2} x_{(i)}}{n-i_1-i_2}}$$

Pri tome, $x_{(i)}$ je redosledni statistik, i_1 je broj rezultata za koje važi: $\frac{(x_i - Mdn)}{MAD/0.6745} \langle 2.24$, i_2 je broj rezultata za koje važi: $\frac{(x_i - Mdn)}{MAD/0.6745} \rangle 2.24$. Dakle, MOM je, u stvari, aritmetička sredina rezultata, koji prema ranije definisanom robustnom kriterijumu <u>ne</u> predstavljaju autlajere.

Medijansko apsolutno odstupanje (eng. Median Apsolute Deviation)

Medijansko apsolutno odstupanje, u oznaci MAD (akronim engleskog naziva ocenitelja), definisano je na sledeći način:

$$\boxed{MAD = Mdn \{ |x_1 - Mdn| \ldots |x_n - Mdn| \}}$$

Dakle, MAD predstavlja Medijanu apsolutnih odstupanja svih rezultata od Medijane.