Dr Dragoljub M. Brkić,
dipl. inž.
Tehnički opitni centar KoV,
Beograd

METODE ODBACIVANJA EKSTREMNIH VREDNOSTI SLUČAJNE PROMENLJIVE

UDC: 519.246

Rezime:

U ovom radu izloženo je nekoliko metoda odbacivanja ekstremnih vrednosti posmatrane slučajne promenljive u toku nekog eksperimenta ili ispitivanja. Radi ilustracije izloženih metoda predočen je određeni broj primera koji su urađeni primenom posebno razvijenog računarskog programa.

Ključne reči: ekstremna vrednost, Studentov ili t-test, Fišerov ili F-test, parametar raspodele, kvantil raspodele, Vejbulova raspodela, poverenje, granica poverenja.

METHODS OF REJECTING THE RANDOM VARIABLE EXTREME VALUES

Summary:

This study deals with several methods of rejecting the extreme values of a particular random variable during an experiment or testing. The described methods are illustrated by a number of examples obtained by a specially developed computer program.

Key words: extreme values, Student or t-test, Fisher or F-test, distribution parameter, distribution quantilles, Weibull distribution, confidence, confidence limit.

Uvod

U toku izvođenja nekog eksperimenta ili ispitivanja, posmatrana slučajna promenljiva može poprimiti vrednosti koje se veoma razlikuju od ostalih vrednosti. To su tzv. ekstremne vrednosti čija je pojava malo verovatna. Ako je ukupan broj vrednosti mali, ekstremna vrednost nepovoljno utiče na tačnost ocena parametara raspodele, pa je treba odbaciti što je opravdano i u skladu sa statističkim testovima.

U ovom radu, izloženo je nekoliko metoda odbacivanja ekstremnih vrednosti koje je u toku nekog eksperimenta ili ispitivanja poprimila posmatrana slučajna promenljiva. Primenom izloženih metoda odbacuju se one ekstremne vrednosti čije realizacije imaju veoma malu verovatnoću, koja je manja ili jednaka, na primer, jednom promilu.

Prva izložena metoda koristi ceo skup vrednosti koje je poprimila slučajna promenljiva, i zasnovana je na Studentovom ili *t*-testu i Fišerovom ili *F*-testu pomoću kojih se proveravaju jednakosti srednjih vrednosti i standardnih devijacija za dva slučaja: prvi, kada je u skupu vrednosti uključena i drugi, kada je isključena posmatrana ekstremna vrednost. Postavljeni kriterijumi u oba navedena testa moraju da budu ispunjeni da bi se posmatrana ekstremna vrednost odbacila ili u protivnom zadržala u skupu vrednosti.

Druga metoda je slična prvoj, a od nje se razlikuje po tome što se primenjuje na veći broj podskupova koji se izdvajaju na slučajan način iz osnovnog skupa vrednosti slučajne promenljive. Podskupovi vrednosti sadrže oko polovine broja vrednosti osnovnog skupa. Svi podskupovi se izdvajaju iz osnovnog skupa podataka po principu pseudoslučajnih brojeva koji predstavljaju redni broj vrednosti u posmatranom osnovnom skupu vrednosti. Svaki od ovih podskupova u prvom slučaju sadrži posmatranu ekstremnu vrednost, a u drugom ta vrednost je isključena. Na osnovu vrednosti u podskupovima određuju se srednje vrednosti i standardne devijacije na koje se dalje primenjuje t-test i F-test radi donošenja odluke o tome da li posmatranu ekstremnu vrednost treba odbaciti ili zadržati u skupu osnovnih vrednosti. Koristeći srednje vrednosti i standardne devijacije podskupova vrednosti slučajne promenljive i primenom t-testa i F-testa, prati se koliko puta od ukupnog broja podskupova posmatranu ekstremnu vrednost nije trebalo odbaciti.

Na osnovu ukupnog broja podskupova vrednosti i broja podskupova vrednosti u kojima je došlo do uspešnog ishoda postupka, određuje se uspešnost, kao i granice poverenja. Ako se unapred zada minimalno prihvatljiva vrednost za ovu uspešnost i ako je donja granica poverenja veća od te minimalno prihvatljive vrednosti za uspešnost, posmatranu ekstremnu vrednost treba zadržati, a u protivnom je odbaciti, što je izloženo u okviru druge metode. Treća izložena metoda odbacivanja ekstremne vrednosti zasnovana je na upoređenju te vrednosti sa utvrđenim kvantilima poznate raspodele slučajne promenljive. U ovom radu razmatran je slučaj troparametarske Vejbulove raspodele. Kriterijum za zadržavanje posmatrane ekstremne vrednosti je ako je ona veća od donjih kvantila i manja od gornjih kvantila. Donji i gornji kvantili određuju se na osnovu tačkastih ocena parametara raspodele za oba slučaja, kada je ekstremna vrednost prisutna u skupu vrednosti i kada je ona namerno isključena iz posmatranog skupa vrednosti. Rešavanje predmetnog problema je složeno i zahteva primenu elektronskog računara. Radi ilustracije izloženih metoda, dato je nekoliko primera koji su urađeni primenom računarskog programa koji je razvijen specijalno za rešavanje ove problematike.

Metode odbacivanja ekstremnih vrednosti

METODA STATISTIČKIH TESTOVA NA CELOM SKUPU VREDNOSTI

Odbacivanje donje ekstremne vrednosti

Neka su $t_1, t_2, ..., t_n$ vrednosti koje je slučajna promenljiva t poprimila u toku jednog eksperimenta. Tačkaste ocene srednje vrednosti i standardne devijacije date su sledećim izrazima:

$$\hat{m}_1 = \bar{t}_1 - \frac{1}{n}\sum_{i=1}^{n} t_i \qquad (1)$$

$$\hat{\sigma}_1 = s_1 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(t_i - \hat{m}_1)^2} \qquad (2)$$

Ako se iz skupa vrednosti koje je uzela slučajna promenljiva t: $\{t_1, t_2, ..., t_n\}$ odbaci najmanja vrednost t_{min}, dobiće se „krnji" skup od n-1 vrednost. Na osnovu ovih vrednosti „krnjeg" skupa treba ponovo odrediti tačkaste ocene za srednju vrednost i standardnu devijaciju koristeći sledeće izraze:

$$\hat{m}_2 = \bar{t}_2 = \frac{1}{n-1}\sum_{i=1}^{n-1} t_i \qquad (3)$$

$$\hat{\sigma}_2 = s_2 = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n-1}(t_i - \hat{m}_2)^2} \qquad (4)$$

Koristeći tačkaste ocene $\hat{m}_1, s_1, \hat{m}_2$ i s_2, kao slučajne promenljive, može se formirati nova slučajna promenljiva:

$$\hat{t} = \frac{\hat{m}_1 - \hat{m}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad (5)$$

gde je $n_1 = n$ i $n_2 = n - 1$.

Slučajna promenljiva $\hat{t}$, data izrazom (5), ima Studentovu ili t-raspodelu sa $n_1 + n_2 - 2$ stepeni slobode. Kada se usvoji vrednost donjeg kvanta (rizika), p, za ovaj broj stepeni slobode $n_1 + n_2 - 2$ može se odrediti kvantil Studentove ili t-raspodele $t_p(n_1 + n_2 - 2)$.

Za ovako određenu vrednost kvantila t-raspodele, ako je ispunjen uslov:

$$-t_p(n_1+n_2-2) < \hat{t} < \left|t_p(n_1+n_2-2)\right| \qquad (6)$$

ekstremnu vrednost t_{min} ne bi trebalo odbaciti. U protivnom, ako uslov dat izrazom (6) nije ispunjen, donju ekstremnu ili minimalnu vrednost t_{min} trebalo bi odbaciti.

Opravdanost odbacivanja t_{min} trebalo bi pojačati ispunjavanjem još jednog uslova. Radi toga treba formirati novu slučajnu promenljivu:

$$\hat{F} = \frac{s_1^2}{s_2^2} \qquad (7)$$

koja ima Fišerovu ili F-raspodelu sa n_1-1 i n_2-1 stepeni slobode.

Za usvojene vrednosti donjeg i gornjeg kvanta (rizika), p i q, respektivno, a za ove brojeve stepeni slobode $n_1 - 1$ i $n_2 - 1$, mogu se odrediti donji i gornji kvantili F-raspodele: $F_1(p; n_1 - 1; n_2 - 1)$ i $F_2(q; n_1 - 1; n_2 - 1)$. Za ovako određene vrednosti kvantila, ako je ispunjen uslov:

$$F_1(p;\ n_1-1;\ n_2-1) < \hat{F} < F_2(q;\ n_1-1;\ n_2-1) \qquad (8)$$

donju ekstremnu vrednost t_{min} ne bi trebalo odbaciti. U protivnom, tj. ako uslov dat izrazom (8) nije ispunjen, t_{min} bi trebalo odbaciti.

Ispunjavanje uslova datog izrazom (6) ukazuje na to da nije došlo do značajnijeg pomeranja, a ispunjavanje uslova datog izrazom (8) da nije došlo do značajnijeg rasipanja posle odbacivanja donje ekstremne ili minimalne vrednosti t_{min}. Dakle, ako su ispunjena oba uslova, data izrazima (6) i (8), tada vrednost t_{min} ne treba odbaciti. U protivnom, ako oba uslova istovremeno nisu ispunjena, tada vrednost t_{min} treba odbaciti. Na taj način, dobija se veća sigurnost u opravdanost odbacivanja vrednosti t_{min}.

Odbacivanje gornje ekstremne vrednosti

Neka su t_1, t_2, ..., t_n vrednosti koje je slučajna promenljiva t poprimila u toku jednog eksperimenta, i neka je t_{max} najveća vrednost u tom skupu vrednosti. Ne odbacujući ovu maksimalnu vrednost t_{max}, pomoću izraza (1) i (2) odrede se tačkaste ocene za srednju vrednost i standardnu devijaciju slučajne promenljive t.

Ako se iz skupa vrednosti koje je uzela slučajna promenljiva t: $\{t_1, t_2, ..., t_n\}$ odbaci najveća vrednost t_{max}, dobiće se „krnji" skup od n–1 vrednost. Na osnovu ovih vrednosti „krnjeg" skupa, treba ponovo odrediti tačkaste ocene za srednju vrednost i standardnu devijaciju koristeći izraze (3) i (4). Na isti način, kao i u slučaju prethodnog postupka odbacivanja minimalne vrednosti t_{min}, koristeći tačkaste ocene za srednje vrednosti i standardne devijacije, formira se slučajna promenljiva koja je data izrazom (5) i slučajna promenljiva koja je data izrazom (7).

Ako vrednosti tačkastih ocena, datih izrazima (5) i (7), ispunjavaju uslove date izrazima (6) i (8), tada gornju ekstremnu vrednost t_{max} ne treba odbaciti. U protivnom, ova vrednost se može odbaciti. Sigurnost, S, u opravdanost odbacivanja gornje ekstremne ili maksimalne vrednosti t_{max} može se izraziti pomoću usvojenih rizika p i q, tj. $S = 1 - (p + q)$.

METODA STATISTIČKIH TESTOVA NA PODSKUPOVIMA VREDNOSTI

Odbacivanje donje ekstremne vrednosti

a) postupak usrednjavanja

Neka su t_1, t_2, ..., t_n vrednosti koje je slučajna promenljiva t poprimila u toku jednog eksperimenta. Iz skupa ovih vrednosti izdvoji se na slučajan način podskup od N vrednosti. Broj N može se odrediti pomoću sledećeg izraza:

$$N = 1 + \left\lfloor \frac{1}{2} n \right\rfloor \tag{9}$$

gde je n ukupan broj vrednosti za t celog skupa, a $\lfloor Q \rfloor$ je celobrojna vrednost broja $Q=n/2$. Pri izdvajanju podskupa vodi se računa da donja ekstremna vrednost t_{min} bude prisutna.

Na osnovu tako izdvojenih vrednosti podskupa, pomoću izraza (1) i (2), određuju se tačkaste ocene za srednju vrednost i standardnu devijaciju, vodeći računa da se u tim izrazima n zameni sa N. Posle toga ponovo se iz celog skupa vrednosti na slučajan način izdvoji podskup od N vrednosti, ali tako da donja ekstremna vrednost t_{min} ne bude prisutna. Ponovo se odrede tačkaste ocene za srednju vrednost i standardnu devijaciju, vodeći računa da se i u ovom slučaju u izrazima (1) i (2) n zameni sa N.

Tako se dobijaju po dve ocene za srednju vrednost i standardnu devijaciju: $\hat{m}_1, s_1, \hat{m}_2$ i s_2.

Ponavljajući ceo ovaj postupak N_p puta (preporučuje se da N_p bude veće od n), dobiće se skupovi vrednosti za $\hat{m}_1, s_1, \hat{m}_2$ i s_2.

Usrednjene vrednosti za srednje vrednosti i standardne devijacije date su sledećim izrazima:

$$\hat{m}_1^{*} = \frac{1}{N_p}\sum_{i=1}^{N_p}\hat{m}_{1_i} \tag{10}$$

$$\hat{m}_2^{*} = \frac{1}{N_p}\sum_{i=1}^{N_p}\hat{m}_{2_i} \tag{11}$$

$$s_1^{*} = \frac{1}{N_p}\sum_{i=1}^{N_p}s_{1_i} \tag{12}$$

$$s_2^{*} = \frac{1}{N_p}\sum_{i=1}^{N_p}s_{2_i} \tag{13}$$

Ako se u izrazima (5) i (7) tačkaste ocene za srednje vrednosti i standardne devijacije zamene sa usrednjenim vrednostima datim odgovarajućim izrazima od (10) do (13), a n_1 i n_2 sa N_p, dobiće se statistike: $\hat{t}^{*}$ i $\hat{F}^{*}$. Ako vrednosti ovih statistika ispunjavaju uslove date izrazima (6) i (8), onda donju ekstremnu vrednost t_{min} ne treba odbaciti; u protivnom ova vrednost t_{min} može se odbaciti.

b) postupak ocene uspešnosti

Kada se na osnovu vrednosti prvog podskupa odrede tačkaste ocene za srednje vrednosti i standardne devijacije: $\hat{m}_1, s_1, \hat{m}_2$ i s_2, kao i statistike: $\hat{t}^{*}$ i $\hat{F}^{*}$ tada se pomoću izraza (6) i (8) izvrši provera da li se donja ekstremna vrednost t_{min} može zadržati. Ako se dobije potvrdan odgovor, onda se to smatra pozitivnim ishodom ili uspehom misije. Proveravanje se nastavlja sve do poslednje probe (izdvajanja podskupa iz osnovnog polaznog skupa vrednosti). Neka je N_p ukupan broj proba (izdvajanja podskupova) i neka je M broj povoljnih ishoda, tj. broj podskupova u kojima je dobijen potvrdan odgovor da se donja ekstremna vrednost t_{min} može zadržati. Na osnovu izloženog može se oceniti verovatnoća zadržavanja ekstremne vrednosti t_{min} pomoću sledećeg izraza:

$$\hat{P}_z = \frac{M}{N_p + 1} \tag{14}$$

Za ovu verovatnoću uspeha ili uspešnost mogu se odrediti i granice poverenja pomoću sledećih izraza:

$$p_1 = 1 - x_{N_p+1-M;\ M+1;\ \alpha_1} \tag{15}$$

$$p_2 = x_{M+1;\ N_p+1-M;\ \alpha_2} \tag{16}$$

gde je u prethodnim izrazima $x_{a;b;\gamma}$ gornji kvantil beta raspodele; α_1 i α_2 su donji i gornji rizik, respektivno.

Za usvojene vrednosti donjeg i gornjeg rizika α_1 i α_2, kao i minimalno prihvatljivu vrednost verovatnoće uspeha P_{min}, da bi se zadržala donja ekstremna vrednost t_{min}, potrebno je da bude ispunjen sledeći uslov:

$$p_1 \geq P_{min} \tag{17}$$

Ako uslov dat izrazom (17) nije ispunjen, tada se donja ekstremna vrednost t_{min} može odbaciti sa poverenjem $\mathcal{P} = 1-(\alpha_1+\alpha_2)$.

Odbacivanje gornje ekstremne vrednosti

a) postupak usrednjavanja

Neka su $t_1, t_2, ..., t_n$ vrednosti koje je slučajna promenljiva t poprimila u toku jednog eksperimenta i neka je t_{max} najveća vrednost u tom skupu vrednosti. Iz skupa tih vrednosti izdvoji se, na slučajan način, podskup od N vrednosti. Broj vrednosti u podskupu N određuje se pomoću izraza (9). Dalji postupak je isti kao i u slučaju odbacivanja donje ekstremne vrednosti, samo što se umesto donje ekstremne vrednosti t_{min}, posmatra gornja ekstremna vrednost t_{max}.

b) postupak ocene uspešnosti

Postupak utvrđivanja opravdanosti zadržavanja ili odbacivanja gornje ekstremne vrednosti isti je kao i postupak izložen u slučaju odbacivanja donje ekstremne vrednosti, samo što se umesto donje ekstremne vrednosti t_{min} posmatra gornja ekstremna vrednost t_{max}.

METODA POREĐENJA EKSTREMNE VREDNOSTI SA KVANTILIMA

Odbacivanje donje ekstremne vrednosti

Neka su $t_1, t_2, ..., t_n$ vrednosti koje je slučajna promenljiva t poprimila u toku jednog eksperimenta i neka je t_{min} najmanja vrednost u skupu tih vrednosti. Pretpostavlja se da slučajna promenljiva t ima troparametarsku Vejbulovu raspodelu čija je funkcija gustine raspodele $f(t, a, b, c)$, gde je a – parametar položaja (početka) b – parametar razmere (skale) i c – parametar oblika čije vrednosti treba odrediti. Tačkaste ocene $\hat{a}, b$ i $\hat{c}$, za parametre a, b i c, respektivno, mogu se odrediti primenom jedne od poznatih metoda.

Donji kvantil troparametarske Vejbulove raspodele dat je sledećim izrazom:

$$t_p = a + b\left[-ln(1-p)\right]^{\frac{1}{c}} \tag{18}$$

gde je $p = F(t_p)$ donji kvant (rizik) ili vrednost funkcije raspodele $F(t)$ za $t = t_p$.

Na osnovu skupa vrednosti slučajne promenljive t: $\{t_1, t_2, ..., t_n\}$, u koji je uključena i donja ekstremna vrednost t_{min}, odrede se tačkaste ocene $\hat{a}_1$, $\hat{b}_1$ i $\hat{c}_1$. Zamenom parametara a, b i c ovim odgovarajućim ocenama u izraz (18) dobija se tačkasta ocena donjeg kvantila troparametarske Vejbulove raspodele:

$$\hat{t}_{p1} = \hat{a}_1 + \hat{b}_1\left[-ln(1-p)\right]^{\frac{1}{\hat{c}_1}} \tag{19}$$

Zatim se iz skupa vrednosti $\{t_1, t_2, ..., t_n\}$ izbaci najmanja ili donja ekstremna vrednost t_{min}, tako da skup ostaje sa $n-1$ vrednosti. Koristeći ove vrednosti „krnjeg“ skupa, ponovo se odrede tačkaste ocene parametara raspodele: $\hat{a}_2, b_2$ i $\hat{c}_2$. U izrazu (18) teorijski parametri a, b i c zamene se ovim tačkastim ocenama i ponovo se odredi vrednost donjeg kvantila:

$$\hat{t}_{p2} = \hat{a}_2 + \hat{b}_2\left[-ln(1-p)\right]^{\frac{1}{\hat{c}_2}} \tag{20}$$

U izrazima (19) i (20) preporučuje se da vrednost za p bude jedan promil ($p = 0{,}001$). Pošto se odrede ove dve tačkaste ocene donjih kvantila, pomoću izraza (19) i (20), tada se najmanja ili donja ekstremna vrednost t_{min} uporedi sa vrednostima ovih kvantila, pa ako je:

$$t_{\min} < \hat{t}_{p1} \wedge t_{\min} < \hat{t}_{p2} \qquad (21)$$

tada se t_{min} može odbaciti, kao malo verovatna vrednost, autlajer (outlayer), jer verovatnoća njene pojave je najverovatnije manja od jednog promila (0,001). Ova tvrdnja je utoliko tačnija, ukoliko je veći broj vrednosti n, slučajne promenljive t. U protivnom, ako uslov dat izrazom (21) nije ispunjen, tada t_{min} ne treba odbaciti.

Odbacivanje gornje ekstremne vrednosti

Neka su $t_1, t_2, ..., t_n$ vrednosti koje je slučajna promenljiva t poprimila u toku jednog eksperimenta i neka je t_{max} najveća vrednost u skupu tih vrednosti. Pretpostavlja se da slučajna promenljiva t ima troparametarsku Vejbulovu raspodelu čija je funkcija gustine raspodele $f(t, a, b, c)$, gde je a – parametar položaja (početka), b – parametar razmere (skale) i c – parametar oblika čije vrednosti treba odrediti. Tačkaste ocene $\hat{a}, \hat{b}$ i $\hat{c}$ za parametre a, b i c, respektivno, mogu se odrediti primenom jedne od poznatih metoda.

Gornji kvantil troparametarske Vejbulove raspodele dat je sledećim izrazom:

$$t_q = a + b\left(-lnq\right)^{\frac{1}{c}} \qquad (22)$$

gde je $q = 1 - F(t_q) = R(t_q)$ gornji kvant (rizik) ili vrednost funkcije pouzdanosti $R(t)$ za $t = t_q$.

Na osnovu skupa vrednosti slučajne promenljive t: $\{t_1, t_2, ..., t_n\}$ u koji je uključena i gornja ekstremna vrednost t_{max}, odrede se tačkaste ocene $\hat{a}_1, \hat{b}_1$ i $\hat{c}_1$. Zamenom parametara a, b i c ovim odgovarajućim ocenama u izraz (22), dobija se tačkasta ocena gornjeg kvantila troparametarske Vejbulove raspodele:

$$\hat{t}_{q1} = \hat{a}_1 + \hat{b}_1\left(-lnq\right)^{\frac{1}{\hat{c}_1}} \qquad (23)$$

Zatim se iz skupa vrednosti $\{t_1, t_2, ..., t_n\}$ izbaci najveća ili gornja ekstremna vrednost t_{max}, tako da skup ostaje sa $n-1$ vrednosti. Koristeći ove vrednosti „krnjeg" skupa, ponovo se odrede tačkaste ocene parametara raspodele: $\hat{a}_2, \hat{b}_2$ i $\hat{c}_2$.

U izrazu (22) teorijski parametri a, b i c zamene se ovim tačkastim ocenama i ponovo se odredi vrednost gornjeg kvantila:

$$\hat{t}_{q2} = \hat{a}_2 + \hat{b}_2\left(-lnq\right)^{\frac{1}{\hat{c}_2}} \qquad (24)$$

U izrazima (23) i (24) preporučuje se da vrednost za q bude jedan promil ($q = 0,001$).

Pošto se odrede ove dve tačkaste ocene gornjih kvantila, pomoću izraza (23) i (24), tada se najveća ili gornja ekstremna vrednost t_{max} uporedi sa vrednostima ovih kvantila, pa ako je:

$$t_{\max} > \hat{t}_{q1} \wedge t_{\max} > \hat{t}_{q2} \qquad (25)$$

tada se t_{max} može odbaciti, kao malo verovatna vrednost, autlajer (outlayer), jer je verovatnoća njene pojave najverovatnije manja od jednog promila (0,001). Ova tvrdnja je utoliko tačnija, ukoliko je veći broj vrednosti n, slučajne promenljive t. U protivnom, ako uslov dat izrazom (25) nije ispunjen, tada t_{max} ne treba odbaciti.

Ilustrativni primeri

PRIMER 1

Pri usvojenim vrednostima parametara Vejbulove raspodele: $a = 250$, $b = 100$ i $c = 2,5$ pomoću specijalno razvijenog računarskog programa, generisati $n = 50$ pseudoslučajnih brojeva koji predstavljaju vrednosti pseudoslučajne promenljive t koja ima Vejbulovu raspodelu sa datim vrednostima parametara ove raspodele. Koristeći tako dobijeni ceo skup vrednosti za t, proveriti da li se može odbaciti donja ekstremna vrednost t_{min}. Usvojiti da su rizici opravdanosti odbacivanja vrednosti t_{min} međusobno jednaki i da iznose 5% ($p = q = 0,05$). Pri rešavanju ovog problema koristiti empirijske vrednosti za srednju vrednost m i standardnu devijaciju σ.

Rešenje:

Za date vrednosti parametara Vejbulove raspodele, pomoću računara, dobijen je sledeći skup od $n = 50$ pseudoslučajnih brojeva.

308,60	383,94	337,82	352,81	334,92
328,63	337,19	315,94	431,45	303,98
321,88	273,17	365,51	426,29	306,37
310,01	347,14	335,99	379,92	392,61
341,88	373,64	271,37	265,04	347,68
311,50	340,32	371,16	340,00	338,33
294,32	340,45	366,59	331,48	274,00
330,34	363,20	376,33	440,85	277,47
301,20	423,33	294,69	319,37	285,65
290,30	352,36	308,88	348,98	352,72

Najmanja i najveća vrednost u datom skupu vrednosti su: $t_{min} = 265,04$ i $t_{max} = 440,85$. Koristeći računarski program, dobijeni su sledeći rezultati:

Za $n = 50$ (kada je uključena vrednost t_{min}):

$\hat{a} = 246,0517$ $\hat{b} = 103,1371$ i $\hat{c} = 2,3342$

$\hat{m} = 337,3525$ i $\hat{\sigma} = s = 42,0800$.

Za $n = 49$ (kada je isključena vrednost t_{min}):

$\hat{a} = 260,2428$ $\hat{b} = 88,6120$ i $\hat{c} = 2,0088$

$\hat{m} = 338,8283$ i $\hat{\sigma} = s = 41,1879$.

Studentov ili t-test:

izračunata vrednost t-statistike, $t = -0,1763$
kritična vrednost t-statistike, $t_p(N) = -1,6607$
broj stepeni slobode $N = 97$
donji rizik $p = 0,05$.

Fišerov ili F-test:

izračunata vrednost F-statistike, $F = 1,0438$
donja kritična vrednost F-statistike, $F_p(N1, N2) = 0,6241$
gornja kritična vrednost F-statistike, $F_q(N1, N2) = 1,6044$
prvi broj stepeni slobode $N1 = 50$
drugi broj stepeni slobode $N2 = 49$
donji rizik $p = 0,05$
gornji rizik $q = 0,05$.
Pošto su istovremeno ispunjena oba uslova:
$-1,6607 < t = -0,1763 < 1,6607$ i
$0,6241 < F = 1,0438 < 1,6044$,
to se sa poverenjem $\mathcal{P} = 1 - (0,05 + + 0,05) = 0,90$ može doneti odluka da ne treba odbaciti donju ekstremnu vrednost $t_{min} = 265,04$.

PRIMER 2

Korišćenjem podskupova vrednosti izdvajanih na slučajan način (sa vraćanjem) iz celog skupa vrednosti pseudo-

slučajnih brojeva datih u Primeru 1, proveriti da li se može odbaciti donja ekstremna vrednost t_{min}.
a) na osnovu usrednjavanja m i σ po podgrupama;
b) na osnovu uspešnosti neodbacivanja t_{min} u podgrupama.
Usvojiti da su rizici opravdanosti odbacivanja vrednosti t_{min} međusobno jednaki i da iznose 5% ($p = q = 0,05$), kao i vrednosti donjeg i gornjeg rizika $\alpha_1=\alpha_2= = 0,05$. Takođe, usvojiti da je minimalna verovatnoća uspešnosti neodbacivanja t_{min}, $P_{min} = 0,90$.
Za rešavanje problema pod a) usvojiti da je broj podgrupa $N_p = 300$, a pod b) $N_p = 50$.

Rešenje:

a) Prosečne vrednosti za srednje vrednosti i standardne devijacije, kada je donja ekstremna vrednost t_{min} uključena i kada je ona isključena, respektivno:

$m_1 = 336,59$ $s_1 = 38,96$
$m_2 = 334,05$ $s_2 = 40,78$.

Studentov ili t-test:

izračunata vrednost t-statistike, $t = 0,7813$
kritična vrednost t-statistike, $t_p(N) = -1,6474$
broj stepeni slobode $N = 598$
donji rizik $p = 0,05$.

Fišerov ili F-test:

izračunata vrednost F-statistike, $F = 0,9126$
donja kritična vrednost F-statistike, $F_p(N1, N2) = 0,8265$
gornja kritična vrednost F-statistike, $F_q(N1, N2) = 1,2099$
prvi broj stepeni slobode $N1 = 299$
drugi broj stepeni slobode $N2 = 299$
donji rizik $p = 0,05$
gornji rizik $q = 0,05$.

Pošto su istovremeno ispunjena oba uslova:

$-1,6474 < t = 0,7813 < 1,6474$ i
$0,8265 < F = 0,9126 < 1,2099$,

to se sa poverenjem $\mathcal{P} = 1 - (0,05 + 0,05) = 0,90$ može doneti odluka da ne treba odbaciti donju ekstremnu vrednost $t_{min} = 265,04$.

b) Od $N = n = 50$ proba (podgrupa) bilo je $M = m = 50$ uspešnih ishoda (slučajeva kada t_{min} nije trebalo odbaciti). Na osnovu toga dobijena je ocenjena frekvencija uspešnosti $f = m/(n + 1) = 0,9804$ i granice poverenja $p_1 = 0,942952$ i $p_2 = 0,998995$.
Pošto je donja granica poverenja $p_1 = 0,942952$ veća od $P_{min} = 0,90$ sa velikim poverenjem se može zadržati donja ekstremna vrednost $t_{min} = 265,04$.

PRIMER 3

Znajući da skup vrednosti pseudoslučajne promenljive t, dat u Primeru 1, ima Vejbulovu raspodelu sa parametrima: $a = 250$, $b = 100$ i $c = 2,5$ na osnovu poređenja donje ekstremne vrednosti t_{min} sa donjim kvantilima t_{1p} i t_{2p}, određenim kada je t_{min} prisutno i kada je ono odsutno, primenom odgovarajućeg izloženog postupka proveriti da li se donja ekstremna vrednost t_{min} može odbaciti. Usvojiti da je red kvantila (kvanti) $p = q = 0,001$.

Rešenje:

Na osnovu izloženog teorijskog postupka i primenom računarskog programa dobijene su sledeće vrednosti donjih kvantila: $t_{1p} = 251{,}40$ i $t_{2p} = 263{,}09$. Oni su određeni sa sledećim tačkastim ocenama parametara Vejbulove raspodele:

$\hat{a} = 246{,}0517$ $\hat{b} = 103{,}1371$ i $\hat{c} = 2{,}3342$;
za $n = 50$ (kada je uključena vrednost t_{min}) i
$\hat{a} = 260{,}2428$ $\hat{b} = 88{,}6120$ i $\hat{c} = 2{,}0088$;
za $n=49$ (kada je isključena vrednost t_{min}).
Pošto su istovremeno ispunjena oba uslova:

$t_{min} = 265{,}04 > t_{1p} = 251{,}40$ i
$t_{min} = 265{,}04 > t_{2p} = 263{,}40$,

sa velikim poverenjem se može doneti odluka da ne treba odbaciti donju ekstremnu vrednost $t_{min} = 265{,}04$.

PRIMER 4

Koristeći opisane postupke i odgovarajući računarski program za odbacivanje ekstremnih vrednosti, kao i skup vrednosti datih u Primeru 1, odrediti neku vrednost za t_{min}, koja je u datom skupu vrednosti jednaka 265,04, koja se po kriterijumima datim u tim postupcima može odbaciti. Za rešavanje ovog problema koristiti sve podatke date u primerima 1, 2 i 3.

Rešenje:

Ako se u skupu vrednosti pseudoslučajne promenljive t, koje su date u Primeru 1, donja ekstremna vrednost t_{min} smanji sa 265,04 na 240,5 i primene sva tri opisana postupka za odbacivanje ekstremnih vrednosti, kao u primerima 1, 2 i 3, onda prema kriterijumima prva dva postupka, vrednost $t_{min} = 240{,}5$ ne treba odbaciti. Međutim, primenom trećeg postupka, koji se odnosi na upoređivanje t_{min} sa donjim kvantilima t_{1p} i t_{2p}, ovu vrednost treba odbaciti. U ovom slučaju dobijene su sledeće tačkaste ocene parametara raspodele:

$\hat{a} = 222{,}569$ $\hat{b} = 128{,}083$ i $\hat{c} = 2{,}8684$;
za $n = 50$ (kada je uključena vrednost $t_{min} = 240{,}5$) i
$\hat{a} = 259{,}719$ $\hat{b} = 89{,}232$ i $\hat{c} = 2{,}0260$;
za $n = 49$ (kada je isključena vrednost $t_{min} = 240{,}5$).

Na osnovu ovih tačkastih ocena dobijene su sledeće vrednosti donjih kvantila: $t_{1p} = 234{,}096$ i $t_{2p} = 262{,}669$.

Pošto nova donja ekstremna vrednost $t_{min} = 240{,}5$ nije veća od oba donja kvantila $t_{1p} = 234{,}096$ i $t_{2p} = 262{,}669$ prema postavljenom kriterijumu ovu vrednost treba odbaciti. Ovaj rezultat je prihvatljiv, kada se zna da su vrednosti date u Primeru 1 generisane sa parametrima raspodele: $a = 250$, $b = 100$ i $c = 2{,}5$, jer vrednost pseudoslučajne promenljive t ne može biti manja od parametra položaja $a = 250$.

Prva dva navedena postupka odnose se na slučajeve kada zakon raspodele slučajne promenljive t nije poznat i koriste se samo srednje vrednosti i standardne devijacije, a treći postupak se zasniva na poznatom zakonu raspodele (troparametarska Vejbulova raspodela). Zbog toga je i dobijen prihvatljiviji rezultat nego u prva dva slučaja, kod kojih, kada je veliki broj podataka n (vrednosti za t), odbacivanje donje ekstremne vrednosti ne utiče bitno na promenu srednjih vrednosti i standardnih devijacija. Dakle, prva dva postupka treba primenjivati u slučajevima kada nije poznat zakon raspodele posmatrane slučajne promenljive.

Zaključak

Izložene metode odbacivanja ekstremnih vrednosti, koje je u toku eksperimenta ili ispitivanja poprimila posmatrana slučajna promenljiva, uglavnom su poznate. Međutim, kriterijumi za ove metode su modifikovani. Tako, na primer, uveden je kriterijum istovremenog ispunjenja zahteva Studentovog ili t-testa i Fišerovog ili F-testa, da bi se posmatrana ekstremna vrednost odbacila ili zadržala u dobijenom skupu vrednosti slučajne promenljive. Drugim rečima istovremeno se proverava značajnost promene srednje vrednosti (mere istinitosti) i standardne devijacije (mere preciznosti) pri odsutnosti ili prisutnosti ekstremne vrednosti u skupu vrednosti. Uvođenjem ovakvog kriterijuma postiže se pouzdanija opravdanost odbacivanja ili zadržavanja ekstremne vrednosti.

Predložena metoda primene uspešnosti i metoda poređenja ekstremne vrednosti sa kvantilima su, uglavnom, nove metode koje, takođe, daju dobre rezultate.

Metoda primene uspešnosti zasnovana je na prvoj izloženoj metodi i primeni granica poverenja za uspešnost, a metoda poređenja ekstremnih vrednosti sa kvantilima zasnovana je na pretpostavci da je poznat zakon raspodele posmatrane slučajne promenljive.

Navedeni primeri, koji su urađeni korišćenjem računarskog programa, koji je razvijen specijalno za rešavanje ove problematike, pokazuju valjanost i osetljivost izloženih metoda odbacivanja ekstremnih vrednosti.

Literatura:

[1] Fisz, M.: Wahrscheinlichkeitsrechnung und Mathematische Statistik, VEB Deutscher Verlag der Wissenschaften, Berlin, 1962.

[2] P. Chapouille et R. De Pazzis, Fiabilité des Systèmes, Masson et C^ie^, Pariz, 1968.

[3] Gnédenko B.; Béliaev, Y.; Soloviev A.: Méthodes mathématiques en théorie de la fiabilité, Éditions, Mir, Moscou, 1972.

[4] B. L. Van Der Waerden, Mathematische Statistik, Springer-Verlag, Berlin, 1965.

[5] H. D. Brunk, An Introduction to Mathematical Statistics, Blaisdell Publishing Company, New York, 1965.

[6] Ivanović, B.: Teorijska statistika, Naučna knjiga, Beograd, 1979.

[7] Stojanović S.: Matematička statistika, Naučna knjiga, Beograd, 1980.

[8] Ivanović, Z.: Matematička statistika, Naučna knjiga, Beograd, 1976.