Intervalne ocene

Interval poverenja za nepoznato matematičko očekivanje m obeležja sa normalnom $\mathcal{N}(m,\sigma)$ raspodelom

- Ako je standardna devijacija σ obeležja poznata

$$\mathbf{I} = \left(\bar{X}_n - a\frac{\sigma}{\sqrt{n}}, \bar{X}_n + a\frac{\sigma}{\sqrt{n}}\right),$$

  gde je $\bar{X}_n$ aritmetička sredina uzorka, n obim uzorka, β zadati nivo poverenja a vrednost a se pronalazi u tablici normalne $\mathcal{N}(0,1)$ raspodele po obrascu $a = \phi^{-1}\left(\frac{1+\beta}{2}\right)$.

- Ako je standardna devijacija σ nepoznata,

$$\mathbf{I} = \left(\bar{X}_n - a\frac{\bar{S}_n}{\sqrt{n-1}}, \bar{X}_n + a\frac{\bar{S}_n}{\sqrt{n-1}}\right),$$

  gde je $\bar{S}_n$ uzoračka standardna devijacija. Ako je obim uzorka $n \geq 30$, možemo primeniti centralnu graničnu teoremu i tada je $a = \phi^{-1}\left(\frac{1+\beta}{2}\right)$. Ukoliko je $n < 30$, tada je $a = t_{n-1,\frac{1+\beta}{2}}$ iz tablice Studentove raspodele.

Interval poverenja za nepoznatu disperziju σ^2 obeležja sa normalnom raspodelom $\mathcal{N}(m,\sigma)$

$$\textbf{dvostrani: } \mathbf{I} = \left(\frac{n\bar{S}_n^2}{a}, \frac{n\bar{S}_n^2}{b}\right), \quad \textbf{jednostrani: } \mathbf{I} = \left(0, \frac{n\bar{S}_n^2}{c}\right)$$

U prethodnim izrazima n je obim uzorka, $\bar{S}_n^2$ je uzoračka disperzija, a vrednosti a, b i c čitamo iz tablice Pirsonove χ^2 raspodele, sa $n-1$ stepenom slobode:

$$a = \chi^2_{n-1,\frac{1+\beta}{2}}, \quad b = \chi^2_{n-1,\frac{1-\beta}{2}}, \quad c = \chi^2_{n-1,\beta}.$$

Jednostrani interval poverenja za nepoznatu disperziju ima za donju granicu broj 0, zato što disperzija ne može imati negativnu vrednost. Koristi se kod testiranja parametarskih hipoteza.

Interval poverenja za nepoznatu proporciju p obeležja sa binomnom $\mathcal{B}(n,p)$ raspodelom

Proporcija p predstavlja verovatnoću uspešne realizacije obeležja sa binomnom raspodelom i ocenjuje se intervalom

$$\mathbf{I} = \left(p - a\sqrt{\frac{pq}{n-1}}, p + a\sqrt{\frac{pq}{n-1}}\right).$$

U prethodnom izrazu $p = \frac{K}{n}$, gde je K broj uspešnih realizacija u uzorku obima n, $q = 1-p$, a $a = \phi^{-1}\left(\frac{1+\beta}{2}\right)$.

Pomoću ovog intervala možemo oceniti i **nepoznat broj pozitivnih realizacija** obeležja sa binomnom $\mathcal{B}(n,p)$ raspodelom **u populaciji obima** N tako što ćemo obe granice intervala za proporciju pomnožiti obimom populacije N tj.

$$\mathbf{I} = \left(N(p - a\sqrt{\frac{pq}{n-1}}), N(p + a\sqrt{\frac{pq}{n-1}}) \right).$$

---

[163] *Dati su podaci o visini u centimetrima 36 slučajno odabranih sportista jednog kluba. Naći 90% interval poverenja za srednju vrednost visine sportista ako je poznata standardna devijacija* $\sigma = 10cm$.

183.7,	*185.9,*	*178.0,*	*185.5,*	*185.1,*	*184.4,*	*184.3,*	*183.3,*	*184.4,*
165.2,	*178.9,*	*173.9,*	*173.7,*	*194.3,*	*186.3,*	*183.8,*	*173.4,*	*176.2,*
176.5,	*180.4,*	*177.1,*	*174.3,*	*187.6,*	*198.8,*	*188.1,*	*180.1,*	*187.1,*
164.9,	*172.6,*	*185.5,*	*167.9,*	*165.1,*	*199.2,*	*168.4,*	*169.1,*	*175.3*

Rešenje: Pretpostavljamo da obeležje koje posmatramo, visina sportiste, ima normalnu raspodelu $\mathcal{N}(m,\sigma)$ (vidi napomenu). Srednja vrednost nekog obeležja je upravo matematičko očekivanje, pa zadatak rešavamo nalazeći interval poverenja za nepoznato očekivanje m sa poznatom devijacijom σ: $\mathbf{I} = \left(\bar{X}_n - a\frac{\sigma}{\sqrt{n}}, \bar{X}_n + a\frac{\sigma}{\sqrt{n}}\right)$.
Obim datog uzorka je $n = 36$, aritmetička sredina iznosi $\bar{x}_n = \frac{1}{36}(183.7 + 185.9 + \cdots + 175.3) = 179.9528$, a kako je zadati nivo poverenja $\beta = 0.90$, iz tablice Gausove raspodele čitamo da je $a = \phi^{-1}\left(\frac{1+0.90}{2}\right) = \phi^{-1}(0.95) = 1.645$.
Uvrštavanjem izračunatih vrednosti u formulu dobijamo da je

$$I = \left(179.9528 - 1.645 \cdot \frac{10}{\sqrt{36}}, 179.9528 + 1.645 \cdot \frac{10}{\sqrt{36}}\right).$$

Interval poverenja za srednju vrednost visine sportista posmatranog kluba, sa nivoom poverenja od 90%, je

$$(177.211cm, 182.694cm).$$

Napomena. Pretpostavka da posmatrano obeležje ima normalnu raspodelu ne važi uvek u praksi. Ali, centralna granična teorema nam garantuje da će za dovoljno veliko n rezultati dobijeni pomoću gornjih formula približno važiti za srednju vrednost posmatranog obeležja ako to obeležje ima raspodelu koja zadovoljava uslove centralne granične teoreme.
Napomena. U literaturi i kompjuterskim programima se često uvodi veličina $se(\bar{x}_n) = \frac{\sigma}{\sqrt{n}}$ koju nazivamo **standardna greška**, (engl. *standard error*). Takođe se u literaturi sa $z = \phi^{-1}\left(\frac{1+\beta}{2}\right)$ obeležava takozvana z-**vrednost** (engl. z-*value*). Sa tim oznakama formulu za interval poverenja pišemo $(\bar{x}_n - z \cdot se(\bar{x}_n), \bar{x}_n + z \cdot se(\bar{x}_n))$.

---

[164] *Aparat za merenje krvnog pritiska u mmHg je testiran na slučajno odabranim zdravim regrutima na regrutaciji. Izmerene vrednosti su: 118, 100, 119, 122, 113, 115, 113, 131, 119, 118, 116, 136, 128, 114, 123, 125, 136, 119, 115, 124, 125, 120, 121, 128, 124, 102. Naći 99% interval poverenja za prosečni krvni pritisak zdravih regruta.*

Rešenje: U ovom zadatku standardna devijacija obeležja nije poznata, pa koristimo interval poverenja za nepoznato očekivanje m sa nepoznatom devijacijom σ : $\mathbf{I} = (\bar{X}_n - a \cdot \frac{\bar{S}_n}{\sqrt{n-1}}, \bar{X}_n + a \cdot \frac{\bar{S}_n}{\sqrt{n-1}})$.
Uzorak sadrži $n = 26$ elemenata, $\bar{x}_n = (118 + 100 + \cdots + 102)/26 = 120.154$, $\bar{s}_n = \sqrt{(118^2 + 100^2 + \cdots + 102^2)/26 - 120.154^2} = 8.3006$. Pošto je obim uzorka manji od 30, vrednost a nalazimo iz tablice Studentove raspodele. Za zadati nivo poverenja $\beta = 0.99$, $a = t_{26-1,\frac{1+0.99}{2}} = t_{25,0.995} = 2.787$.

Traženi interval poverenja za prosečni krvni pritisak zdravih regruta je

$$I = (120.154 - \frac{2.787 \times 8.30057}{\sqrt{25}}, 120.154 + \frac{2.787 \times 8.30057}{\sqrt{25}}) = (115.527, 124.781).$$

---

[165] *U tabeli*

ocena	*1*	*2*	*3*	*4*	*5*
broj đaka	*5*	*9*	*4*	*5*	*5*

su prikazane ocene na pismenom ispitu u slučajno odabranom odeljenju petog razreda. Poznato je da se ocene ponašaju u skladu sa Normalnom raspodelom. Proceniti 95%-im intervalom poverenja srednju ocenu na pismenom u petim razredima, ako

(a) je poznato da standardno odstupanje ocena na pismenom iznosi $\sigma = 1.4$.

(b) standardno odstupanje ocena na pismenom nije poznato.

Rešenje: Deo zadatka pod (a) rešavamo pomoću intervala poverenja za nepoznato m sa poznatim σ. Podaci su dati u vidu grupisanog uzorka, tako da je obim uzorka $n = 5 + 9 + 4 + 5 + 5 = 28$ i $\bar{x}_n = \frac{1}{28}(1 \cdot 5 + 2 \cdot 9 + 3 \cdot 4 + 4 \cdot 5 + 5 \cdot 5) = 2.8571$.
Za $\beta = 0.95$, $a = \phi^{-1}\left(\frac{1+0.95}{2}\right) = \phi^{-1}(0.975) = 1.96$.
Interval poverenja za srednju ocenu je

$$(2.8571 - 1.96\frac{1.4}{\sqrt{28}}, 2.8571 + 1.96\frac{1.4}{\sqrt{28}}) = (2.3385, 3.3757).$$

Kako u zadatku pod (b) standardna devijacija (odstupanje) nije poznata, koristimo interval poverenja za nepoznato m sa nepoznatim σ. Nepoznatu standardnu devijaciju ocenjujemo uzoračkom standardnom devijacijom

$$\bar{s}_n = \sqrt{\frac{1}{28}(1^2 \cdot 5 + 2^2 \cdot 9 + 3^2 \cdot 4 + 4^2 \cdot 5 + 5^2 \cdot 5) - 2.8571^2} = \sqrt{1.9084} = 1.4073.$$

Obim uzorka je manji od 30, tako da je $a = t_{28-1,\frac{1+0.95}{2}} = t_{27,0.975} = 2.052$.
Interval poverenja za srednju ocenu u ovom slučaju je

$$(2.8571 - 2.052\frac{1.4073}{\sqrt{27}}, 2.8571 + 2.052\frac{1.4073}{\sqrt{27}}) = (2.3014, 3.4128).$$

Napomena. U većini slučajeva, pa tako i ovde, možemo uočiti da poznavanje standardne devijacije obeležja smanjuje interval poverenja, odnosno povećava preciznost ocene nepoznatog očekivanja. Međutim, ovaj podatak u praksi najčešće nije poznat.

---

[166] *Kontrola kvaliteta u fabrici za proizvodnju deterdženta je izmerila masu pojedinačnih pakovanja iz slučajno odabranog uzorka. Rezultati su dati u tabeli.*

masa [g]	*[4800, 4900)*	*[4900, 4950)*	*[4950, 5000)*	*[5000, 5050)*	*[5050, 5100)*	*[5100, 5200]*
f_i	*8*	*31*	*96*	*109*	*48*	*8*

(a) *Naći 95% interval poverenja za prosečnu masu deterdženta u jednom pakovanju.*

(b) *Naći 99% interval poverenja za prosečnu masu deterdženta u jednom pakovanju.*

Rešenje: Standardna devijacija posmatranog obeležja (masa deterdženta) nije poznata, i zato koristimo formulu za interval poverenja za matematičko očekivanje sa nepoznatim σ. Uzorak je intervalni, pa potrebne vrednosti računamo na sledeći način:
$n = \sum f_i = 8 + 31 + ... + 8 = 300$,

$$\bar{x}_n = \frac{1}{n}\sum(x_i f_i) = \frac{1}{300}(48508 \cdot 8 + 4925 \cdot 31 + ... + 5150 \cdot 8) = 5005.33,$$

$$\bar{s}_n = \sqrt{\frac{1}{300}(4850^2 \cdot 8 + 4925^2 \cdot 31 + ... + 5150^2 \cdot 8) - 5005.33^2} = 55.7965.$$

Obim uzorka je veći od 30, pa za određivanje vrednosti koristimo tablicu Gausove raspodele. U delu zadatka pod (a) je $a_1 = \phi^{-1}\left(\frac{1+0.95}{2}\right) = \phi^{-1}(0.975) = 1.96$, dok je u delu pod (b) $a_2 = \phi^{-1}\left(\frac{1+0.99}{2}\right) = \phi^{-1}(0.995) = 2.576$.

Uvrštavanjem izračunatih vrednosti dobijamo intervale poverenja za očekivanu masu deterdženta:

$$(a)\ I_{95\%} = (5005.33 - 1.96\frac{55.7965}{\sqrt{299}}, 5005.33 + 1.96\frac{55.7965}{\sqrt{299}}) = (4999.01, 5011.65),$$

$$(b)\ I_{99\%} = (5005.33 - 2.576\frac{55.7965}{\sqrt{299}}, 5005.33 + 2.576\frac{55.7965}{\sqrt{299}}) = (4997.02, 5013.64).$$

Napomena. Uočavamo da se povećanjem nivoa poverenja povećava i odgovarajući interval poverenja. Ovo oslikava zakonitost koja važi kod svih statističkih ispitivanja - veća pouzdanost vodi do manje preciznosti i obrnuto. Zbog ove zakonitosti ne možemo birati prevelike nivoe poverenja (npr. 0.999 i slično) jer bi odgovarajući intervali poverenja bili previše široki i stoga praktično neupotrebljivi.

---

[167] *Od 1000 novorođenčadi u jednom porodilištu 517 su bili dečaci. Naći 90% interval poverenja za procenat dečaka.*

Rešenje: Nepoznati procenat dečaka ćemo oceniti pomoću intervala poverenja za nepoznatu proporciju p: $\mathbf{I} = \left(p - a\sqrt{\frac{pq}{n-1}}, p + a\sqrt{\frac{pq}{n-1}}\right)$.
Posmatrani uzorak je obima $n = 1000$ i u njemu je bilo $k = 517$ uspešnih realizacija, tako da je realizovana proporcija $\bar{p} = \frac{517}{1000} = 0.517$ i $\bar{q} = 1 - 0.517 = 0.483$. Za zadati nivo poverenja $\beta = 0.9$, tablična vrednost $a = \phi^{-1}\left(\frac{1+0.9}{2}\right) = \phi^{-1}(0.95) = 1.645$.
Traženi interval poverenja je

$$\left(0.517 - 1.645\sqrt{\frac{0.517 \cdot 0.483}{999}}, 0.517 + 1.645\sqrt{\frac{0.517 \cdot 0.483}{999}}\right) = (0.491, 0.543),$$

odnosno sa nivoom poverenja 90% možemo tvrditi da će biti između 49.1% i 54.3% dečaka.

---

[168] *Kontrola kvaliteta iz zadatka* [166] *treba da utvrdi 95% interval poverenja za proporciju pakovanja čija je masa van dozvoljenog odstupanja tj. ispod* 4900 g *ili preko* 5100 g. *Ako se godišnje proizvede 350000 pakovanja deterdženta, odrediti 95% interval poverenja za broj pakovanja koja su nepropisne mase.*

Rešenje: Iz tablice date u zadatku [166] vidimo da je broj pakovanja iz kontrolnog uzorka koji imaju manje od 4900 g jednak 8, isto kao i broj onih koji imaju više od 5100 g. Stoga je broj uspešnih realizacija $k = 8 + 8 = 16$ (pod uspehom uvek podrazumevamo ono što je od interesa u zadatku, u ovom konkretnom slučaju uspeh je kada je pakovanje nepropisne mase). Obim uzorka $n = 300$, pa je $\bar{p} = \frac{16}{300} = 0.053$ i $\bar{q} = 1 - 0.053 = 0.947$. Tablična vrednost $a = \phi^{-1}\left(\frac{1+0.95}{2}\right) = 1.96$.
Uvrštavanjem dobijenih rezultata u obrazac za interval poverenja za nepoznatu proporciju, dobijamo

$$\left(0.053 - 1.96\sqrt{\frac{0.053 \cdot 0.947}{299}}, 0.053 + 1.96\sqrt{\frac{0.053 \cdot 0.947}{299}}\right) = (0.0276, 0.0783).$$

Kako je obim cele populacije jednak godišnjoj proizvodnji $N = 350000$, dobijamo da je 95% interval poverenja za broj pakovanja koja su nepropisne mase

$$(350000 \cdot 0.0276, 350000 \cdot 0.0783) = (9660, 27405).$$

---

[169] *Na jednoj kontrolnoj tački puta izmerena je brzina slučajno odabranih automobila i rezultati su prikazani u sledećoj tabeli:*

brzina [km/h]	*[60, 80]*	*(80, 90]*	*(90, 100]*	*(100, 110]*	*(110, 120]*	*(120, 140]*
f_i	*5*	*20*	*36*	*23*	*15*	*1*

(a) Naći 95% interval poverenja za prosečnu brzinu automobila na toj kontrolnoj tački.

(b) Naći 99% interval poverenja za prosečnu brzinu automobila na toj kontrolnoj tački, ako je poznata standardna devijacija $\sigma = 11.2$.

(c) Ako je najveća dozvoljena brzina na tom delu puta 90 km/h, odrediti 90% interval poverenja za procenat automobila koji se kreću nedozvoljenom brzinom.

(d) *Ako u toku jednog dana tim putem prođe 1200 automobila, i ako je kazna za prekoračenje brzine 1500 RSD, odrediti 90% interval poverenja za iznos novca koji bi sakupila saobraćajna policija ako bi naplatila kaznu svima koji su tog dana i na tom mestu prekoračili brzinu.*

Rešenje:

(a) Vrednosti potrebnih statistika su $n = 100$, $\bar{x}_n = 97.4$, $\bar{s}_n = 11.863$, tablična vrednost je $a = 1.96$, pa je odgovarajući interval poverenja

$$I = (95.063km/h,\ 99.737km/h).$$

(b) Vrednosti potrebnih statistika su $n = 100$, $\bar{x}_n = 97.4$, $\sigma = 11.2$, tablična vrednost je $a = 2.576$, pa je odgovarajući interval poverenja

$$I = (94.515km/h,\ 100.285km/h).$$

(c) Obim uzorka je $n = 100$, uzoračka proporcija je $\bar{p} = 0.75$, $\bar{q} = 0.25$, a tablična vrednost $a = 1.645$. Sledi da je

$$I = (67.8\%,\ 82.2\%).$$

(d) Interval poverenja dobijamo množenjem granica prethodnog intervala $I = (1200 \cdot 1500 \cdot 0.678, 1200 \cdot 1500 \cdot 0.822) = (1220400\ RSD,\ 1479600\ RSD)$.

---

[170] *Iz obeležja sa normalnom raspodelom uzet je uzorak obima $n = 27$. Izračunata je uzoračka disperzija $\bar{s}^2_{27} = 9.7344$. Naći dvostrani i jednostrani interval poverenja za disperziju sa nivoom poverenja $\beta = 0.95$.*

Rešenje: Intervale poverenja za nepoznatu disperziju obeležja sa normalnom raspodelom nalazimo pomoću sledećih formula: $\mathbf{I} = \left(\frac{n\bar{S}^2_n}{a}, \frac{n\bar{S}^2_n}{b}\right)$, i $\mathbf{I} = \left(0, \frac{n\bar{S}^2_n}{c}\right)$.
Obim uzorka i uzoračka disperzija su poznati: $n = 27$ i $\bar{s}^2_{27} = 9.7344$, tako da još treba odrediti tablične vrenosti iz tablice χ^2 rapodele. Za zadati nivo poverenja $\beta = 0.95$, imamo da je
$a = \chi^2_{n-1,\frac{1+\beta}{2}} = \chi^2_{27-1,\frac{1+0.95}{2}} = \chi^2_{26,0.975} = 41.9$, $\quad b = \chi^2_{n-1,\frac{1-\beta}{2}} = \chi^2_{27-1,\frac{1-0.95}{2}} = \chi^2_{26,0.025} = 13.8$, $\quad c = \chi^2_{26,0.95} = 38.9$.
Dakle, dvostrani interval je

$$I = \left(\frac{27 \times 9.7344}{41.9}, \frac{27 \times 9.7344}{13.8}\right) = (6.27, 19.05),$$

a jednostrani

$$I = \left(0, \frac{n\bar{S}^2_n}{c}\right) = \left(0, \frac{27 \times 9.7344}{38.9}\right) = (0, 6.76).$$

---

[171] *Uz pretpostavku da obeležje „masa pakovanja” iz zadatka* [166] *ima normalnu raspodelu, izračunati 98-procentni dvostrani i jednostrani interval poverenja za nepoznatu disperziju uzorka.*