Originalni naučni rad

UDK 339.138:[330.3:311

DOI 10.7251/SVR1919132M

OSNOVA ZA IZRAČUNAVANJE POLOŽAJA REZULTATA –NORMALNA DISTRIBUCIJA

Prof. dr Branka Marković[1]

Nezavisni univerzitet Banja Luka

Apstrakt: Statistiku, možemo definisati kao nauku koja se bavi posmatranjem, prikupljanjem, grupisanjem (sređivanjem), analizom i tumačenjem prikupljenih podataka. Statistika može da se tumači u širem pojmu kao organizovana djelatnost, metodologija koja se koristi u istraživanjima ekonomskih, društvenih i prirodnih pojava.

Ako posmatramo statističke podatke kao dio velikog skupa informacija(podataka), pomoću statističkih metoda procjenjujemo nepoznate parametre osnovnog skupa ili možemo da vršimo testiranje hipoteza.

Osnovni skup (statistički skup, populacija ili masovna pojava) je skup statističkih jedinica (elemenata) i čije su karakteristike predmet statističkog istraživanja.

Uzorak je dio osnovnog skupa koji možemo statistički analizirati. Uzorak treba da ima sve karakteristike osnovnog skupa, da bih mogao predstavljati (reprezentovati) analiziranu pojavu.

Da bismo izvršili analizu i ocjenu osnovnog skupa, potreban je dobar uzorak. Uzorak je dobar jedino ukoliko je dovoljno reprezentativan, a u tom značenju mora biti slučajan i dovoljno velik.

Ključne riječi: *uzorak, osnovni skup, normalna raspodjela, populacija*

UVOD

Pri statističkom istraživanju masovnih pojava, koristi se teorijske raspodjele ili distribucije koje su predstavljene određenim parametrima. Takve rasporede zovemo rasporedima vjerovatnoća slučajne promjenljive. Ovakav raspored je ujedno i jednodimenzionalan teorijski raspored. Takvi rasporedi se u statistici dijele na prekidne i neprekidne što zavisi da li se radi o prekidnoj ili neprekidnoj slučajno promjenljivoj.

Za Normalnu distribuciju kažemo da je unimodalna, simetrična, neprekidna i jednodimenzionalna teorijska raspodjela (distribucija). Pošto svaku normalu distribuciju možemo tumačiti preko standardiziranog obilježja, pri čemu transformacijom ćemo dobiti standardiziranu normalnu distribuciju. Vjerovatnoća da ćemo vrijednost kontinuirane promjenljive imati u određenom intervalu mora biti jednaka vrijednosti funkcije distribucije.

[1] NUBL Banja Luka

Kada upoznamo osnovni skup, možemo da zaključimo da uzorak iako reprezentativan se razlikuje od osnovnog skupa, i zbog te razlike moramo vršiti procjenu karakteristika osnovnog skupa na osnovu karakteristika uzorka izabranog iz tog skupa. Pri tom slučajna promjenljiva poprima vrijednosti zavisno od slučaja tj. prema distribuciji vjerovatnoće.

Mjere srednje vrijednosti, mjere disperzije, mjere asimetrije i mjere zaobljenosti mogu se da dobiju iz podataka u cijelom osnovnom skupu. Međutim, ove mjere možemo izračunati iz konačnog osnovnog skupa. Ukoliko imamo beskonačan osnovni skup, podatke moramo računati samo iz dijela osnovnog skupa. Znači, moramo uzeti dio skupa koji zovemo uzorak. Taj uzorak koji uzmemo mora imati sve karakteristike osnovnog skupa i takav uzorak zovemo repezantativnim. Kod beskonačnih skupova, reprezentativni uzorak će nam poslužiti za procjenu istih karakteristika osnovnog skupa. Osnovni parametri uzorka se izračunavaju na isti način kao i parametri osnovnog skupa.

Za svaki uzorak koji izaberemo, možemo izračunati određeni parametar uzorka koji označavamo sa $\widehat{\theta}$(grčki: theta), a isti će nam poslužiti za procjenu parametara osnovnog skupa θ. Ukoliko izaberemo n uzoraka iz osnovnog skupa, toliko možemo izračunati parametara $\hat{\theta}$.Parametre osnovnog skupa θ ćemo procjeniti pomoću parametara uzorka $\hat{\theta}$, ti paremetri se razlikuju zato što je svaki paramerar $\hat{\theta}$ izračunat samo za n jedinica osnovnog skupa., koji je izabran od ukupnog broja jedinica N osnovnog skupa. Parametre $\hat{\theta}$ koje dobijemo za svaki uzorak se među sobom razlikuju, pošto je svaki parametar izračunat iz različitih uzoraka.

Zbog velikog broja uzoraka, za procjenu, moramo izabrati samo jedan. Na koji način ćemo izabrati taj uzorak, a njegovi parametri nam mogu poslužiti za procjenu parametara osnovnog skupa. Izračunati parametri $\hat{\theta}$ međusobom se razlikuju, zbog razlike u uzorku. Parametri će varirati od uzorka do uzorka. Varijacija koja postoji između parametara uzorka zove se sampling varijacija (sample=uzorak; tj samnpling=izbor uzorka). Pošto su uzorci izabrani slučajno, iz toga možemo zaključiti da je $\hat{\theta}$ slučajna promjenljiva. Slučajna promjenljiva poprima vrijednosti prema nekoj distribuciji vjerovatnoće.Ukoliko odredimo teorijsku distribuciju, da odredimo vjerovatnoću slučajne promjenljive $\hat{\theta}$. Ako se radi o diskretnim promjenljivim tada će $\hat{\theta}$ imati vrijednosti nekog realnog broja x, a ako se radi o kontinuiranim promjenljivim tada će vrijednost $\hat{\theta}$ uzimati neku vrijednost iz intervala između dva realna broja x_1 i x_2. Slučajno promjenljiva $\hat{\theta}$ poprima svoje vrijednosti prema distribuciji vjerovatnoće, a ona ima iste parametre kao teorijska distribucija: aritmetičku sredinu i varijansu.

Iz gore navedenog, možemo zaključiti da je procjena parametara osnovnog skupa θ pomoću parametara uzorka $\hat{\theta}$ intervalna procjena.

TEORIJSKE DISTRIBUCIJE

Ukoliko distribucija nastaje grupisanjem podataka, pri čemu nastaju grupe, klase ili razredi takvu distribuciju zovemo empirijska, originalna ili opažena distribucija. Postoje i distribucije koje možemo da očekujemo na osnovu našeg iskusta. Te distribucije zovemo teorijskom distribucijom. Za razliku od empirijskih distribucija, kod kojih nam nisu poznati parametri kao što je aritmetička sredina, mod, medijana, $\alpha_{3,}\alpha_4$, kod teorijske distibucije unaprijed znamo te parametre. Pošto se teorijska distribucija, u procjeni, javlja u ulozi distibucije vjerovatnoće,imamo dvije vrste vjerovatnoće:vjerovatnoća a priori (empirijska ili klasična) i aposretiori (statstička).

Vrste teorijskih distribucija

Često moramo računati vjerovatnoću za niz događaja, pri tome računamo apsolutnu frekvenciju svakog ishoda, zatim relativnu frekvenciju ili vjerovatnoću a priori svakog ishoda. pri čemu zbir svih vjerovatnoća kod relativnih frekvencija je jednaka 1, a cijeli niz tih vjerovatnoća čini distribuciju vjerovatnoće.

Ako imamo diskretne vrijednosti kod X promjenljive: $x_1; x_2; x_3; \ldots. x_k$ a njihova vjerovatnoća $p(x_1), p(x_2); p(x_3) \ldots\ldots. p(x_k)$pri čemu uređeni parovi izgledaju: $x_1 p(x_1) \ldots\ldots x_k p(x_k)$ $i = 1,2, \ldots k$ oni predstavljaju funkciju vjerovatnoće slučajne promjenljive X.

Uslovi funkcije vjerovatnoće diskretne promjenljive mora da zadovolji sljedeće:

- $P(x_i) \geq 0$, $i = 1,2,3 \ldots k$
- $\sum_{i=1}^{k} p(x_i) = 1$ (1.1)

Kumulativne distribucije ćemo dobiti ako se vjerovatnoće kumuliraju, a funkcija koju čine dijelovi niza zove se funkcija distribucije. Funkcija distribucije koju označavamo kao *F(X),* je funkcija koja će dati vjerovatnoću da će vrijednost *X* imati vrijednosti nekog realnog broja *x:*

$$F(x_i)=P(X\leq x_i)= \sum_{x\leq xi} p(x_i) \qquad (1.2)$$

Ako je promjenljiva kontinuirana , za nju ne možemo imati pojedinačne vrijednosti nekog realnog broja, nego je vjerovatnoća da će slučajna promjenljiva imati vrijednosti u određenom intervalu, pri čemu njena funkcija vjerovatnoće ima ove karakteristike:

- $F(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Funkcija distribucije vjerovatnoće kontinuiranih promjenljivi ima sljedeći izraz:

$$F(X)=a=\int_{-\infty}^{\infty} f(x)dx = 1 \qquad (1.3)$$

Distribuciju vjerovatnoće karakterišu njeni parametri, a u diskretinih i kontinuiranih promjenljivih to je očekivana vrijednost(aritmetička sredina) i varijansa.

U slučaju kada uzmemo aposteri (statističku) vjerovatnoću kao zamjenu apriori vjerovatnoće, kod koje imamo matematičke vrijednosti, kao granične vrijednosti tada možemo govoriti da se distribucija vjerovatnoće može smatrati teorijskim graničnim oblikom distribucije relativnih frekvencija, kada se broj opita kreće do beskonačnosti.

Teorijske distribucije se dijele na one za diskretne i kontinuirane promjenljive. U distribucije za diskretne promjenljive ubrajamo: Binomnu i Poissonova, a za kontinuirane promjenljive: Normalna (Gaussova), Studentova, F-distribucija, χ^2-distribucija.

Normalna (Gaussova ili gauss-Laplaceova) distribucija

Ova distribucija predstavlja najznačajniju teorijsku distribuciju. Prvi put je objavljena 1753. godine od strane Abrahama de Moivere i to kao granična vrijednost binomne raspodjele. Njemački matematičar Carl Fricdrich Gauss, (1777-1855) na bazi Laplasovih i Moiveovih istraživanja sa teorijom vjerovatnoće razvio osnove teorije greške i stvorio raspored vjerovatnoće poznat pod imenom *Normalna raspodjela* ili *Gaussova kriva*. On je raspored primjenio pri ocjeni slučajnih grešaka kod mjerenja vrijednosti onih obilježja koja se javljaju u velikom broju.

Funkcija gustine vjerovatnoće normalne distribucije se prikazuje pomoću formule:

$$F(x)=\frac{1}{\delta\sqrt{2\pi}}\cdot e^{\frac{8-(X-\mu)^2}{2\delta^2}}$$ matemetički model normalne distribucije

(1.4)

Normalna distribucija je zvonolika, simetrična, neprekidna jednodimenzionalna i unimodalna. Ovaj raspored je u potpunosti određen sa dva osnovna parametra a to je aritmetička sredina i standardna devijacija neprekidne slučajne promjenjive.

Normalna distribucija je predstavljena krivom koja ima zvonolik i idealno simetričan oblik u odnosu na vertikalnu pravu koju smo povukli iz tačke $x=\mu$, ona je unimodalna pošto se prostire od $-\infty\ do + \infty$, pošto je simetrična, mjera asimetrije je $\alpha_3=0$, a njema mjera zaobljenosti je $\alpha_4=3$,mod, medijana su jednaki njenoj aritmeričkoj sredini.

Iz matematičko modela funkcije znamo koliko iznos π=3,14159, dok je e=2,71828(baza prirodnih logaritama), ako nam je poznata standardna devijacija i aritmerička sredina možemo odrediti oblik normalne krive. Poznato je da aritmetička sredina slučajne promjenljive predstavlja udaljenost od koordinatnog početka, a standarna devijacija predstavlja razvučenost krive. Ukoliko nam se mijenja aritmerička sredina, kriva se pomjera duž apscise. Ukoliko su vrijednosti standardne devijacije manje, kriva će biti uža, ukoliko su vrijednosti standardne devijacije veće ona je šira. Za svaki dati slučaj možemo utvrditi vjerovatnoću gdje će slučajna promjenljiva javi u određenom intervalu koristeći pri tome formulu:

$$P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)$$
(1.5)

U ovom slučaju bismo za svaki konkretan slučaj (različite intervale)trebali računati površinu ispod normalne krive. Da bismo ovaj problem riješila, statistička teorija je uvela mogućnost da se svaka normalna distribucija može preformulisati na standardiziranu distribuciju.

Svaka normalna slučajna promjenljiva *X* se može prikazati izrazom $X=\mu \pm z\delta$, tako da ćemo umjesto slučajne promjenljive *X* dobiti standardizirano obilježje *z:*

$$Z=\frac{x-\mu}{\delta} \qquad (1.6)$$

Ovo nam odstupanje, a što moramo da znamo, predstavlja mjeru i smjer odstupanja, bilo koje veličine normalne slučajne promjenljive *X* u odnosu na njenu aritmetičku sredinu, a to odstupanje prikazujemo u broju standardnih devijacija. Ovo pretvaranje vrijednosti *X* u za to odgovarajuće vrijednosti *Z* formira neprekidnu standardiziranu slučajno promjenljivu, čije vrijednosti se kreću -∞ $do + \infty$. Ovim postupkom smo dobili funkciju standardizirane normalne distribucije:

$$f(Z)=\frac{1}{\sqrt{2\pi}} \cdot e^{\frac{z^2}{2}} \qquad (1.7)$$

Kod standardizirane normalne distribusije, aritmetička sredina je jednaka nuli, dok je standardna devijacija jednaka jedan. Koristeći formulu za funkciju standardizirane normalne distribucije, možemo doći do sljedećeeg zaključka: funkcija *f(z)* je eksponencijalna parna funkcija pošto je *f(-z)=f(+z)* i ta je kriva, kada se posmatra, u odnosu na vertikalnu pravu simetrična i dignuta iz tače *z=0* =aritmetička sredina .

Vrijednost krive je kada je *z=0* pri čemu je modijana, mod i aritmetička sredina jednake.Ukoliko uzmemo drugi izvod , pri čemu imamo prevojnu tačku kada je $z=\pm 1$ i funkciju

$f(\pm 1) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}} = 0{,}24197$, u ovoj tački kriva iz konveksnog prelazi u konkavan oblik. Na ovaj način se vrši obračun teorijskih ordinata standardizovane normalne krive. Ordinate su već izračunate i možemo ih pronaći u tablici standardiziranog normalonog rasporeda.

Funkciju rasporeda standardizirane promjenljive prikazujemo na sljedeći način:

$$F(z)=p(Z \leq z) = \int_{-\infty}^{z} f(z)dz = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz$$
(1.8)

Vjerovatnoća da će se vrijednost standardizirane(kontinuirane) normalne(slučajne)promjenljive Z naći se u intervalu između $(z_1; z_2)$

jednaka je razlici između vjerovatnoća vrijednosti funkcije raspodjele za $(z_{1;}z_2)$ i može se napisati:

$$P(z_1 < Z \leq z_2)=\int_{z_1}^{z_2} f(z)dz = F(z_2)\text{-}F(z_1)$$
(1.9)

Površina ispod krive je jednaka jedinici ili 100%, pošto je normalna standardizirana kriva simetrična iz toga će proisteći zaključak da je 50% površine ispod krive nalazi do tačke $z=0$ sa lijeve strane a ostalih 50% površine nalazi se sa desne strane od tačke $z=0$. Upravo zbog ovih osobina kod određivanja funkcije rasporeda standardizirane normalne, potrebno je samo da računski dobijemo površinu ispod krive koja poklapa interval od tačke na apscisi $z=0$ do određene vrijednosti z:

$$\Phi(z)= \int_0^z f(z)dz = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{z^2}{2}}dz$$
(2.0)

Veličine (površine) ispod normalne standardizirane krive za interval $(0{,}00\leq z \leq 3{,}99)$ je data u tablicama, a znači da je to vjerovatnoća da će slučajna promjenljiva z naći u intervalu između određene vrijednosti z i aritmetičke sredine. Na osnovu ove konstatacije uvijek možemo izračunati vjerovatnoću za standardiziranu normalnu promjenljivu ako se nalazi u intervalu između z_1 *i* z_2, tu razliku možemo naći u tabeli. Iz tabele ćemo izračunati vjerovatnoću, ako su z_1 *i* z_2 istog predznaka njihova vjerovatnoća će biti razlika vrijednosti iz tablica, ako su razlučitog predznaka ona je jednaka zbiru vrijednosti iz tabele $\Phi(-z)=\Phi(z)$ a vjerovatnoću dobijemo $P(z_1 < Z < z_2) = \Phi(z_1) - \Phi(z_2)$.

Ako imamo vrijednost z_1=1 pomoću tablica utvrdimo da je $\Phi(1)=0{,}34134$ ili procentualno 34,134%, što bi značilo da se 34,134% površine standardizirane normalne krive može nalaziti između tačaka $z=0$ *i* $z=1$ .Vjerovatnoću predstavlja površina ispod standardizirane normalne krive. Ovo nam govori da će ovaj procenat iznositi vjerovatnoću da će vrijednost Z nalaziti između vertikale podignute iz tačke z_1 *i* z_2. Pošto smo već rekli da je standardizirana normalna kriva simetrična , što bi značilo da će vjerovatnoća u istom opitu biti tolika i u intervalu između $z=0$ *i* $z=-1$. Kada saberemo površine ispod standardizirane normalne krive u intervalu $z=1$ *i* $z=-1$, one će iznositi 0,68268 (0,34134+0,34134) ili to iznosi 68,268%od ukupne površine ispod standardizirane normalne krive koja iznosi 1 ili 100%.

ZAKLJUČAK

Sve pojave koje analiziramo a mijenjaju. U tom svom mijenjanju nastoje ostati blizu tačke presjeka, tj. raspoređuju se po jednoj krivoj liniji zvonolikog oblika koju zovemo Gaussova ili normalna raspodjela(distribucija). Nultu tačku nema sa osom apscice, zato što je nigdje ne presijeca ali joj se približava. Osnovna karakteristika krive jeste da su simetrično sa lijeve i desne strane raspoređeno po 50% podataka. Kod diskrene varijable funkcija rasporeda vjerovatnoće nije neprekidna, ona ima onoliko tačaka prekida koliko ima vrijednosti. Za razliku od diskretnih varijabli, neprekidne slučajne varijable imaju unutar intervala iz koga se javljaju, jako mnogo vrijednosti.

Procjena nepoznatog parametara zasniva se na korištenju i izboru slučajnog uzorka iz osnovnog skupa. Nepoznati parametri se mogu procjenjivati pomoću intervala i brojčano. Pri intervalnoj procjeni moramo znati da li interval sadrži stvarnu vrijednost parametra koji nam je nepoznat i sa kojom vjerovatnoćom možemo tvrditi da se taj parametar nalazi unutar tog intervala. Kod intervalne procjene postoje granice unutar kojih se nalazi nepoznati parametar.

Procjena intervala aritmetičke sredine populacije (osnovnog skupa) se bazira na konstataciji da su sve prosječne vrijednosti uzoraka normalno raspoređene oko aritmetičke sredine populacije. Procjenu ćemo vršiti po principima vjerovatnoće Gaussove raspodjele. Sve aritmetičke sredine uzoraka nemaju istu aritmetičku sredinu kao populacija, nego one variraju oko te sredine. Upravo ta odstupanja i variranja aritmetičke sredine uzoraka od atitmetičke sredine populacije čini standardu grešku procjene aritmetičke sredine populacije. Pošto nam nikada nije poznata standardna devijacija populacije, ona se mora procjeniti iz standardne devijacije uzorka.

U društvu se može jako mnogo riješiti različitih situacija korištenjem Gaussove krive (distribucije), koja uzima samo vrijednosti kontinuirane promjenljive.

OSNOVA ZA IZRAČUNAVANJE POLOŽAJA REZULTATA – NORMALNA DISTRIBUCIJA

Prof. dr Branka Marković

Abstract: Statistics can be defined as a science that deals with observing, collecting, grouping, analyzing and interpreting data collected. Statistics can be interpreted more broadly as organized activity, a methodology used in studies of economic, social and natural phenomena.

If we view statistics as part of a large set of information (data), we use statistical methods to estimate unknown parameters of the basic set or to test hypotheses.

A basic set (statistical set, population or mass phenomenon) is a set of statistical units (elements) and whose characteristics are the subject of statistical research.

The sample is part of a basic set that we can statistically analyze. The sample should have all the characteristics of the basic set, in order to be able to represent the analyzed phenomenon.

A good sample is needed to perform the analysis and analyze the basic set. And a sample is only good if it is representative enough, and in that sense it must be random and large enough.

Key words: *sample, basic set, normal distribution, population*

LITERATURA

1. Goldfarb, B., Pardoux, C. *Introduction to the statistical methods*. Dunod, Paris(1993).
2. John, J., Whitaker, D.; Johnson, D. *Statistical Thinking for Managers*. Chapman & Hall/CRC, Boca Raton (2001).
3. Šošić, I. *Primijenjena statistika*. Školska knjiga, Zagreb. (2004).