

STANDARDIZIRANE VRIJEDNOSTI. NORMALNA DISTRIBUCIJA.

U statističkoj analizi često se uspoređuje *relativni položaja podatka u nizovima*. To se postiže s pomoću *standardizirane vrijednosti z*.

STANDARDIZIRANE VRIJEDNOSTI niza numeričke varijable X definira se izrazom:

$$(1) z_i = \frac{x_i - \bar{x}}{\sigma}, \quad \bar{z} = 0, \quad \sigma_z = 1.$$

Standardizirana vrijednost z dobije se tako da se od vrijednosti varijable X oduzme njezina aritmetička sredina, a razlika podijeli sa standardnom devijacijom. Aritmetička sredina standardiziranih vrijednosti jednaka je 0. Standardna devijacija standardiziranih vrijednosti jednaka je 1. Standardizirane vrijednosti ne ovise o mjernim jedinicama varijabli.

PRIMJER 1.

Prosječan broj bodova na testu iz statistike bio je 75, s prosječnim odstupanjem 10 bodova. Pristupnik je na tom testu postigao 80 bodova. Prosječan broj bodova na testu iz matematike bio je 65, s prosječnim odstupanjem 15 bodova. Na testu iz matematike isti je pristupnik imao 74 boda. Valja odgovoriti na pitanje na kojem je testu student postigao bolji uspjeh.

Na to pitanje možemo odgovoriti samo uz pomoć standardizirane vrijednosti zato što su dana dva podatka za jednog pristupnika (uspjeh iz statistike i matematike) koja pripadaju nizovima s različitim aritmetičkim sredinama i standardnim devijacijama. Standardizirana je vrijednost za uspjeh iz statistike ($x = 80$):

$$z_{\text{stat.}} = \frac{x - \bar{x}_{\text{stat.}}}{\sigma_{\text{stat.}}} = \frac{80 - 75}{10}, \quad z_{\text{stat.}} = 0.5.$$

Na testu iz statistike pristupnik je postigao natprosječan rezultat ($80 > 75$), a odstupanje je od aritmetičke sredine $+0.5$ standardnih devijacija. Standardizirana je vrijednost za uspjeh iz matematike ($x = 74$):

$$z_{\text{mat.}} = \frac{x - \bar{x}_{\text{mat.}}}{\sigma_{\text{mat.}}} = \frac{74 - 65}{15}, \quad z_{\text{mat.}} = 0.6.$$

I na testu iz matematike pristupnik je postigao natprosječan rezultat ($74 > 65$). Rezultat na testu odstupa od aritmetičke sredine $+0.6$ standardnih devijacija. Dakle, relativno gledajući, uspjeh pristupnika na testu iz matematike bolji je nego na testu iz statistike, zato što je $z_{\text{mat.}} > z_{\text{stat.}}$.

Koeficijenti su varijacije rezultata testova:

$$V_{\text{mat.}} = \frac{\sigma_{\text{mat.}}}{\bar{x}_{\text{mat.}}} 100 = \frac{15}{65} 100, \quad V_{\text{mat.}} = 23.08\%. \quad V_{\text{stat.}} = \frac{\sigma_{\text{stat.}}}{\bar{x}_{\text{stat.}}} 100 = \frac{10}{75} 100, \quad V_{\text{stat.}} = 13.33\%.$$

Koeficijenti varijacije pokazuju da je raspršenost rezultata testa iz matematike veća od raspršenosti rezultata iz statistike.

Dodatni sadržaj veže se uz sadržaj u udžbeniku
Šošić, I., *Statistika*, Školska knjiga, Zagreb, 2014.
na stranici 151.

Prosudite li se o intervalima u kojima se nalaze proporcije broja vrijednosti numeričkog niza, primjenjuje se više pravila koja se oslanjaju na aritmetičku sredinu i standardnu devijaciju. Takvo je i sljedeće pravilo. Za *bilo kako raspoređeni numerički niz* općenito vrijedi da se u intervalu od $(\bar{x} - 2\sigma)$ do $(\bar{x} + 2\sigma)$ nalazi najmanje 75 % njegovih članova, u intervalu od $(\bar{x} - 3\sigma)$ do $(\bar{x} + 3\sigma)$ najmanje 90 % članova, a u intervalu od $(\bar{x} - 4\sigma)$ do $(\bar{x} + 4\sigma)$ najmanje 94 % njegovih članova.

PRIMJER 2.

Proizvođač kruha svakog dana kupcima isporučuje u prosjeku 1 858 kg kruha, s prosječnim odstupanjem od 40 kg. Može li se narudžba kupaca od 2 050 kg smatrati neuobičajeno velikom? (b) Prosječna posljednja cijena vrijednosnice TR-AM u 100 dana trgovanja iznosi 56.54 eura, s prosječnim odstupanjem od 2.50 eura. Što se može reći za posljednju cijenu od 52.04 eura?

(a) Odstupanje narudžbe od $x = 2\,050$ prema prosjeku izrazit ćemo u jedinicama standardne devijacije:

$$z = \frac{x - \bar{x}}{\sigma} = \frac{2\,050 - 1\,858}{40}, \quad z = 4.8.$$

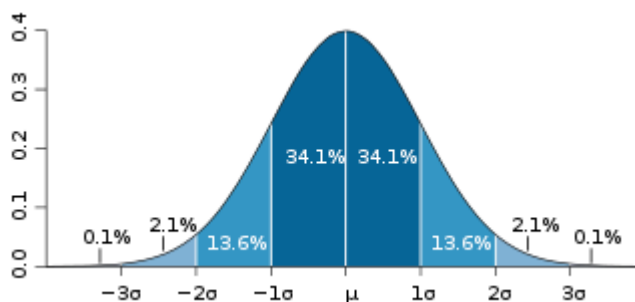
Prema navedenom pravilu, *izvan* pojasa $(\bar{x} - 4\sigma)$ do $(\bar{x} + 4\sigma)$ može se očekivati najviše 6 % vrijednosti (3 % manje od donje granice i 3 % više od gornje granice). Vrijednost 4.8 govori da narudžba od 2 050 odstupa od prosjeka naviše za 4.8 standardnih devijacija i predložuje neuobičajeno veliku vrijednost.

(b) Odstupanje posljednje cijene od 52.04 eura prema prosjeku izrazit ćemo u jedinicama standardne devijacije.

$$z = \frac{x - \bar{x}}{\sigma} = \frac{52.04 - 56.54}{2.5}, \quad z = -1.8.$$

Prema navedenom pravilu, odstupanje posljednje cijene 52.04 od prosjeka (odstupanje prema nižoj cijeni od 1.8 standardnih devijacija) ne može se uzeti kao neuobičajeno veliko.

Raspored podataka katkad se približno ravna prema *normalnoj distribuciji*. Normalna distribucija ubraja se među najvažnije statističke distribucije. Za tu je distribuciju svojstveno da joj je poznat analitički izraz (funkcija). Po obliku je simetrična i zvonolika. Sve tri srednje vrijednosti normalne distribucije (aritmetička sredina, medijan, mod) međusobno su jednake.



Slika 1. Normalna (Gaussova) distribucija.

Uz pomoć svojstava normalne distribucije mogu se donositi različiti zaključci u pogledu obuhvata broja vrijednosti u nekom rasponu varijacije i prosudbe o položaju pojedine vrijednosti prema prosjeku.

Ako se podatci raspoređuju po normalnoj distribuciji sa sredinom \bar{x} i standardnom devijacijom σ , tada je u intervalu od $(\bar{x} - \sigma)$ do $(\bar{x} + \sigma)$ obuhvaćeno (zaokruženo) 68 % podataka, u intervalu od $(\bar{x} - 2\sigma)$ do $(\bar{x} + 2\sigma)$ 95 % podataka, a u intervalu od $(\bar{x} - 3\sigma)$ do $(\bar{x} + 3\sigma)$ 99.7 % podataka, to jest gotovo svi podatci.

Iz svojstva normalne distribucije slijedi da vrlo malo podataka raspoređenih prema normalnoj distribuciji odstupa od aritmetičke sredine za više od tri standardne devijacije (ukupno 0.3 %, to jest 0.15 % na svakom kraku). Pri primjeni pravila normalne distribucije nužno je provjeriti jesu li podatci raspoređeni prema normalnoj distribuciji.

PRIMJER 3.

Dobavljač kupcu isporučuje paketiće kave prosječne neto-težine 100 g. Dopusšteno je prosječno odstupanje od prosječne težine 2 g. Pretpostavimo li da je distribucija paketića kave oblika normalne distribucije sa sredinom 100 i standardnom devijacijom 2, određuje se broj paketića težine od 96 do 104 g u pošiljci 25 000 isporučenih paketića.

Težina paketića od 96 g od prosječne težine odstupa -2σ , a težina od 104 g odstupa $+2\sigma$ od prosječne težine. Prema tome, interval od 96 do 104 odgovara intervalu od -2σ do $+2\sigma$ normalne distribucije. U tom je intervalu zaokruženo 95 % podataka, što znači da se u isporučenoj robi očekuje 23 750 paketića kave težih od 96 do 104 g.