



1. Uvodna statistika

1.1 Uloga i važnost statistike u analizi podataka u opskrbnim lancima

Statistika igra ključnu ulogu u modernim opskrbnim lancima, gdje su učinkovito upravljanje, planiranje i kontrola ključni. Statističke metode koriste se za prikupljanje, analizu i tumačenje podataka, omogućujući tvrtkama da bolje razumiju i optimiziraju svoje opskrbe lance.

Istaknimo neke od važnih uloga statistike u analizi opskrbnog lanca.

Deskriptivna statistika ključna je za opisivanje osnovnih svojstava podataka o opskrbnom lancu, kao što su srednja vrijednost, standardna devijacija, medijan, kvartili i druge mjere. Ovi alati nam pomažu razumjeti distribuciju i karakteristike podataka kao što su prosječna vremena isporuke, količine na zalihama i prosječni troškovi, što pridonosi boljem razumijevanju i upravljanju opskrbnim lancem.

Osim toga, statističke tehnike kao što su regresija, analiza vremenskih serija i analiza uzoraka koriste se za predviđanje budućih događaja i trendova u opskrbnim lancima. To uključuje predviđanje, tj. prognoziranje potražnje, zaliha i vremena isporuke, što omogućuje bolje planiranje i prilagodbu ponude.

Statistika igra ključnu ulogu u prepoznavanju obrazaca u podacima, omogućujući bolje razumijevanje ponašanja opskrbnog lanca, uključujući sezonske obrasce, trendove i cikluse potražnje.

Optimizacija zaliha još je jedno ključno područje u kojem statistika pomaže u određivanju optimalnih količina narudžbe koje minimiziraju troškove skladištenja i naručivanja, koristeći metode kao što je EOQ (engl. *Economic Order Quantity* - ekonomična količina narudžbe).

Osim toga, statistika se također koristi za procjenu rizika opskrbnog lanca, kao što je vjerojatnost kašnjenja u isporukama, oštećenja tijekom transporta i drugih potencijalnih problema.

Statističkim praćenjem i kontrolom procesa identificiramo odstupanja od standarda, što nam omogućuje poboljšanje kvalitete i učinkovitosti procesa opskrbnog lanca.



Osim toga, statistika se koristi za praćenje i poboljšanje kvalitete proizvoda i usluga u opskrbnom lancu, uključujući kontrolu kvalitete kod dobavljača.

Konačno, statistika je ključni alat za donošenje utemeljenih odluka o nabavi, zalihama, odabiru dobavljača i drugim aspektima upravljanja opskrbom, pridonoseći učinkovitom i djelotvornom radu cijelog opskrbnog lanca.

U analizi opskrbnog lanca statistika se koristi za optimizaciju procesa, smanjenje troškova, povećanje učinkovitosti i poboljšanje zadovoljstva kupaca. Omogućuje bolje razumijevanje dinamike opskrbnog lanca i bolje upravljanje rizicima, što je ključno za uspješno poslovanje tvrtki i organizacija u današnjem globalnom okruženju.

1.2 Osnovni pojmovi statistike

Variable

Variable su osnovni građevni blokovi u statistici jer predstavljaju svojstva ili karakteristike koje se mjere ili promatraju u anketi, eksperimentu ili uzorku podataka. Variable su ključne za razumijevanje i analizu podataka jer omogućuju istraživačima, analitičarima i statističarima da opisuju, analiziraju i razumiju fenomene.



Važno je razumjeti različite vrste varijabli i njihovu važnost u statistici.

Kvalitativne (deskriptivne, kategoričke) variable su variable koje predstavljaju kvalitativne karakteristike ili kategorije koje se ne mogu prebrojati ili klasificirati prema matematičkom redu. Primjeri uključuju spol (muški, ženski), boju očiju (plave, smeđe, zelene) ili vrstu automobila (limuzina, karavan, SUV). Kvalitativne variable često su korisne za opisivanje demografskih karakteristika ili osobina.

Kvantitativne (numeričke) variable su variable koje predstavljaju numeričke vrijednosti koje se mogu prebrojati ili izmjeriti i mogu se sortirati po nekom matematičkom redu. Primjeri uključuju dob, visinu, temperaturu, prihod ili rezultate istraživanja. Kvantitativne variable često se koriste za analizu i kvantitativno istraživanje fenomena.

Zavisne i nezavisne variable. Zavisna varijabla je ona koju želimo istražiti, mjeriti ili predvidjeti, dok je nezavisna varijabla ona koja treba utjecati na zavisnu varijablu. Na primjer,



ako želimo istražiti utječe li razina obrazovanja na dohodak, dohodak je zavisna varijabla, a razina obrazovanja nezavisna varijabla.

Diskretne i kontinuirane varijable. Varijable se također mogu podijeliti na diskretne i kontinuirane. Diskretne varijable imaju ograničen skup mogućih vrijednosti i obično su predstavljene cijelim brojevima. Primjer je broj djece u obitelji, gdje su moguće vrijednosti 0, 1, 2 itd. Kontinuirane varijable, s druge strane, imaju beskonačan broj mogućih vrijednosti i obično se mjere pomoću decimalnih brojeva. Primjer je visina osoba, gdje je moguć beskonačan broj vrijednosti unutar zadanog raspona.

Varijable su osnovni alati za istraživanje i analizu podataka. Razumijevanje i pravilno definiranje varijabli ključno je za provođenje statističkih analiza i proučavanje fenomena u istraživanju. Varijable omogućuju istraživačima izražavanje i kvantificiranje različitih aspekata stvarnosti, omogućujući bolje razumijevanje fenomena, donošenje odluka i predviđanje budućih događaja. Također omogućuju korištenje različitih statističkih tehnika za testiranje hipoteza, predviđanja i boljeg razumijevanja uzročno-posljedičnih veza između varijabli.

1.3 Osnovni statistički koncepti s primjerima

Prosjek (srednja vrijednost)

Srednja **vrijednost** (engl. *mean*), također poznata kao **prosjek**, jedna je od osnovnih statističkih mjera. Srednja vrijednost je aritmetički prosjek svih vrijednosti u skupu podataka. Izračunava se zbrajanjem svih podataka, a zatim dijeljenjem s brojem podataka.



Izračunavanje prosjeka:

- Zbrojite sve vrijednosti u skupu podataka.
- Podijelite zbroj s brojem vrijednosti u skupu.
- Jednadžba za izračunavanje prosjeka ($\bar{x}$) je: $\bar{x} = (x_1 + x_2 + x_3 + \dots + x_n) / n$

Gdje je $\bar{x}$ prosjek. $x_1, x_2, x_3, \dots, x_n$ su vrijednosti u skupu podataka. n je broj vrijednosti u skupu podataka.



Primjer:

Zamislite skup podataka koji predstavlja ocjene učenika na ispitu iz matematike: 80, 85, 90, 75, 95. Da biste izračunali prosjek, zbrojite sve ove vrijednosti i podijelite s brojem ocjena, koji je u ovom slučaju 5:

$$\text{Prosjeck} = (80 + 85 + 90 + 75 + 95) / 5 = 425 / 5 = 85$$

Dakle, prosječni studentski rezultat je 85. Prosjek je koristan za mjerenje centralne tendencije podataka i daje nam grubu ideju o tome što možemo očekivati kao "tipičnu" vrijednost u skupu podataka. Međutim, srednja vrijednost može se značajno promijeniti ako su u podacima prisutni ekstremi. Stoga je važno poznavati druge statističke mjere kao što su medijan i mod kako bi se bolje razumjela distribucija podataka.

Medijan

Medijan je statistički koncept koji se koristi za mjerenje pozicijske srednje vrijednosti skupa podataka. To je vrijednost koja dijeli uređene podatke na dvije jednake polovice. To znači da polovica podataka ima vrijednosti manje ili jednake medijanu, a druga polovica ima vrijednosti veće ili jednake medijanu. Medijan je jedna od osnovnih mjera centralne tendencije u statistici i koristi se za opisivanje distribucije podataka, posebno kada su podaci iskrivljeni ili sadrže izvanredne vrijednosti.



Kako izračunati medijan?

- Prvo morate sortirati skup podataka od najmanje do najveće vrijednosti.
- Ako je broj podataka paran (n), tada je medijan prosjek dviju srednjih vrijednosti. To znači da je medijan jednak prosjeku vrijednosti na poziciji $n/2$ i $(n/2 + 1)$ kada su podaci poredani uzlaznim redoslijedom.
- Ako je broj podataka neparan, tada je vrijednost medijana na sredini.

Primjer:

Zamislite sljedeći skup podataka koji predstavlja broj sati sna koje su ljudi imali u određenom razdoblju: 7, 6, 5, 8, 6, 9, 7

Prvo posložite podatke uzlaznim redoslijedom: 5, 6, 6, 7, 7, 8, 9



Budući da je broj podataka neparan (7), medijan će biti vrijednost na srednjoj poziciji, što je četvrta vrijednost u poredanom skupu podataka. Dakle, medijan je u ovom slučaju jednak 7 sati. To znači da polovica ljudi u ovom skupu podataka spava 7 ili manje sati, dok druga polovica spava 7 ili više sati.

Mod

Mod je jedna od osnovnih statističkih metrika koja se koristi za mjerenje centralne tendencije skupa podataka. Mod predstavlja vrijednost koja se najčešće pojavljuje u skupu podataka. To je vrijednost koja ima najveću učestalost pojavljivanja među svim vrijednostima u skupu podataka.

Modus je koristan za identificiranje najčešće vrijednosti u skupu podataka i posebno je koristan pri analizi kvalitativnih (kategorijskih) varijabli gdje su vrijednosti nenumeričke.

Ako postoji više modova u skupu podataka (više vrijednosti koje se javljaju sa sličnom maksimalnom učestalošću), govorimo o višemodalnoj distribuciji. Ako svi podaci imaju istu učestalost pojavljivanja, tada skup podataka nema mod.

Primjer: zamislite skup podataka koji predstavlja boje automobila na parkiralištu:

Crvena, Plava, Crvena, Zelena, Plava, Plava, Plava, Crvena

U ovom slučaju mod je "Crvena", jer se ova vrijednost najčešće pojavljuje (tri puta), dok se "Plava" i "Zelena" pojavljuju rjeđe.

Mod je jednostavan za izračunavanje jer jednostavno identificira vrijednost s najvećom učestalošću pojavljivanja u skupu podataka. Mod se koristi za opisivanje karakterističnih vrijednosti u podacima i može biti koristan u razumijevanju koja je vrijednost najkarakterističnija za određenu situaciju ili skupinu.

Raspon varijacije

Razlika između maksimalne i minimalne vrijednosti u skupu podataka je statistički koncept koji se naziva raspon. Time se mjeri kolika je razlika između maksimalne i minimalne vrijednosti u skupu podataka. Raspon je jednostavan način za procjenu raspona vrijednosti u skupu podataka i mjerenje varijabilnosti između minimalnih i maksimalnih vrijednosti.



Izračun raspona varijacije je jednostavan:

- Najprije pronađite minimalnu vrijednost (min) i maksimalnu vrijednost (max) u skupu podataka.
- Zatim izračunajte razliku između maksimalne i minimalne vrijednosti ($\max - \min$).

Primjer: zamislite skup podataka koji predstavlja dob sudionika događaja: 20, 25, 30, 35, 40. Da biste izračunali raspon varijacije, prvo pronađite minimalnu vrijednost (20) i maksimalnu vrijednost (40) u skupu podataka. Zatim izračunajte razliku između maksimalne i minimalne vrijednosti: $VR = 40 - 20 = 20$

Dakle, raspon varijacije u ovom slučaju je 20 godina. To znači da je razlika između najstarijeg i najmlađeg sudionika 20 godina.

Dekompozicija varijacije je korisna za procjenu raspona vrijednosti u skupu podataka, ali je prilično jednostavna i ne uzima u obzir sve vrijednosti u skupu podataka. Za detaljniju analizu varijabilnosti i disperzije podataka obično se koriste druge statističke mjere kao što su varijanca ili kvartili.

Varijanca i standardna devijacija

Varijanca je prosječni zbroj kvadrata odstupanja od srednje vrijednosti. To je kvadrat standardne devijacije. **Standardna devijacija** je statistička mjera koja se koristi za mjerenje disperzije ili varijabilnosti u skupu podataka. Govori koliko su vrijednosti udaljene od srednje vrijednosti (prosjeaka) u skupu. Standardna devijacija jedna je od najčešće korištenih mjera disperzije u statistici i izračunava se izračunavanjem kvadratnog korijena varijance.

Izračunavanje standardne devijacije:

- Prvo izračunajte varijancu. Varijanca se izračunava uzimanjem prosjeka svih vrijednosti u skupu za svaku vrijednost u skupu, zatim kvadriranjem i zbrajanjem tih razlika.
- Kada dobijete vrijednost varijance (σ^2), izračunajte standardnu devijaciju izračunavanjem kvadratnog korijena varijance. To se radi vađenjem kvadratnog korijena iz σ^2 :



Standardna devijacija $\sigma = \sqrt{\sigma^2}$

Standardna devijacija mjeri koliko su vrijednosti disperzirane oko srednje vrijednosti u skupu podataka. Veća vrijednost standardne devijacije znači da su vrijednosti raširenije i više se razlikuju od srednje vrijednosti, dok niža vrijednost standardne devijacije označava manju raspršenost.



Primjer: zamislite skup podataka koji predstavlja ocjene učenika na ispitu iz matematike: 80, 85, 90, 75, 95. Formula koja će biti prikazana u nastavku vrijedi samo ako pet vrijednosti s kojima smo započeli čine cjelokupnu populaciju. Prvo izračunavate prosjek (srednju vrijednost), koji iznosi 85. Zatim izračunavate varijaciju, koja iznosi 50.

Prvo izračunajte odstupanja svake podatka od srednje vrijednosti i kvadrirajte rezultat svake:

$$(80 - 85)^2 = (-5)^2 = 25, \quad (85 - 85)^2 = (0)^2 = 0, \quad (90 - 85)^2 = (5)^2 = 25, \quad (75 - 85)^2 = (-10)^2 = 100, \quad (95 - 85)^2 = (10)^2 = 100$$

Varijanca je srednja vrijednost ovih vrijednosti:

$$\sigma^2 = \frac{25 + 0 + 25 + 100 + 100}{5} = \frac{250}{5} = 50$$

Na kraju, izračunavate standardnu devijaciju uzimajući kvadratni korijen varijance:

$$\text{Standardna devijacija} = \sqrt{50} \approx 7.07$$

Dakle, standardna devijacija u ovom slučaju je oko 7,07. To znači da su u prosjeku rezultati učenika udaljeni oko 7,07 jedinica od prosjeka. Standardna devijacija se često koristi u analizi distribucije podataka i u procjeni varijabilnosti vrijednosti u skupu.

Kvantili

Kvantili su vrijednosti koje dijele uređene podatke u određene dijelove. Na primjer, kvantili dijele podatke na četiri jednaka dijela. Prvi kvartil (Q1) dijeli donjih 25% podataka, drugi kvartil (Q2) jednak je medijanu, a treći kvartil (Q3) dijeli gornjih 25% podataka.



Primjer: u skupu podataka 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, prvi kvartil (Q1) jednak je 6, drugi kvartil (Q2) jednak je 11, a treći kvartil (Q3) je jednak 16.



1.4 Prikaz statistike

Prikaz statistike uključuje korištenje različitih metoda i alata, s ciljem da se podaci prezentiraju na jasan, transparentan i informativan način.

Evo nekoliko uobičajenih načina za prikaz statistike:

Tablice

Tablice su osnovna metoda za prikaz podataka. Primjeri uključuju tablice frekvencija, koje pokazuju broj pojavljivanja za različite vrijednosti, i podatkovne tablice, koje prikazuju više informacija o podacima.

Studenti – ostvarene ocjene	Zbrojne oznake	Frekvencija
41 - 49		3
50 - 58		6
59 - 67		5
68 - 76		6
77 - 85		2
		Total =22

Slika 1.1 Primjer tablice.

Grafički prikazi

Grafički prikazi učinkovit su alat za vizualizaciju podataka. Uključuju različite vrste grafikona kao što su stupčasti grafikoni, linijski grafikoni, tortni grafikoni, histogrami, dijagram pravokutnika itd.



Slika 1.2 Primjeri grafičkih prikaza podataka.

Linijski grafikoni koriste se za vizualizaciju trendova i promjena tijekom vremena, što ih čini idealnim za praćenje podataka koji se kontinuirano razvijaju. Posebno su učinkoviti za prikazivanje odnosa između varijabli i isticanje uzoraka, kao što su povećanja, smanjenja ili



fluktuacije. Linijski grafikoni obično se koriste u područjima kao što su financije, znanost i poslovanje za analizu vremenskih serija podataka, usporedbu trendova u kategorijama ili predviđanje budućeg razvoja na temelju povijesnih podataka.

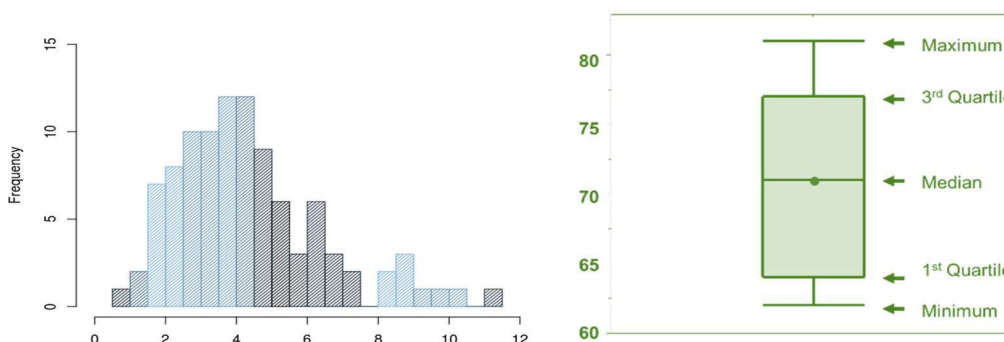
Stupčasti grafikoni koriste se za usporedbu količina u različitim kategorijama, što ih čini idealnim za predstavljanje diskretnih podataka. Posebno su učinkoviti za isticanje razlika, sličnosti i trendova među skupinama. Stupčasti grafikoni obično se koriste kada trebate prikazati frekvencije, postotke ili druge numeričke mjere na jasan i vizualno jednostavan način. Široko se primjenjuju u poslovanju, obrazovanju i istraživanju za analizu i priopćavanje kategoričkih podataka.

Polarni grafikoni, također poznati kao radarski ili paukovi grafikoni, koriste se za prikaz višestrukih podataka u više dimenzija u kružnom formatu. Idealni su za usporedbu nekoliko varijabli ili entiteta prema istim kriterijima, ističući prednosti i slabosti na jasan, vizualni način. Često se koriste u analizi učinka, donošenju odluka i konkurentskim usporedbama, kao što je procjena značajki proizvoda, timskih vještina ili rezultata ankete u različitim kategorijama.

Tortni grafikoni koriste se za predstavljanje proporcija ili postotaka cjeline, što ih čini idealnim za vizualizaciju relativnih veličina različitih kategorija. Posebno su učinkoviti kada želite pokazati kako dijelovi doprinose ukupnom iznosu ili na prvi pogled usporediti proporcije. Tortni grafikoni obično se koriste u izvješćima, prezentacijama i istraživanjima za prikaz podataka poput tržišnog udjela, raspodjele proračuna ili demografske distribucije.

Histogrami

Histogrami su grafički prikazi distribucije podataka. Koriste se za prikaz frekvencije vrijednosti varijable u različitim intervalima.



Slika 1.3 Histogram i dijagram pravokutnika (Box-Plot).



Dijagram pravokutnika (Box plot)

Dijagram pravokutnika, ili okvir s brkovima, vrsta je grafikona koji se koristi u deskriptivnoj statistici kao prikladan način grafičkog predstavljanja grupa numeričkih podataka njihovim sažimanjem s pet brojeva: minimum, prvi kvartil, medijan, treći kvartil i maksimum.

Izbor metode za prikaz statistike ovisi o prirodi podataka, ciljevima analize i ciljanoj publici. Važno je odabrati metodu koja najbolje odgovara vašoj poruci i čini podatke lakšim za razumijevanje.

1.5 Distribucija frekvencija

Distribucija frekvencija, također poznata kao tablica frekvencija ili histogram, način je prikazivanja broja pojavljivanja različitih vrijednosti varijable u skupu podataka. Pomoću distribucije frekvencija možete identificirati obrasce, distribucije i učestalosti vrijednosti u podacima. Obično se koristi za analizu kvalitativnih (kategoričkih) varijabli, ali se također može koristiti za prikaz diskretnih vrijednosti kvantitativnih (numeričkih) varijabli.



Proces stvaranja distribucije frekvencija uključuje sljedeće korake:

- Prikupljanje podataka: prvo prikupite podatke za koje želite izraditi distribuciju frekvencija.
- Identificirajte različite vrijednosti: identificirajte različite vrijednosti koje se pojavljuju u vašim podacima. Ovo su kategorije ili diskretne vrijednosti koje želite analizirati.
- Brojanje pojavljivanja: brojite koliko se puta svaka vrijednost pojavljuje u skupu podataka.
- Napravite tablicu frekvencija: izradite tablicu koja prikazuje sve različite vrijednosti varijable i broj pojavljivanja za svaku vrijednost.
- Crtanje histograma: ako imate veliki broj različitih vrijednosti, možete izraditi histogram koji prikazuje distribuciju frekvencija. Ovo je grafički prikaz koji prikazuje broj pojavljivanja svake vrijednosti u obliku stupaca.

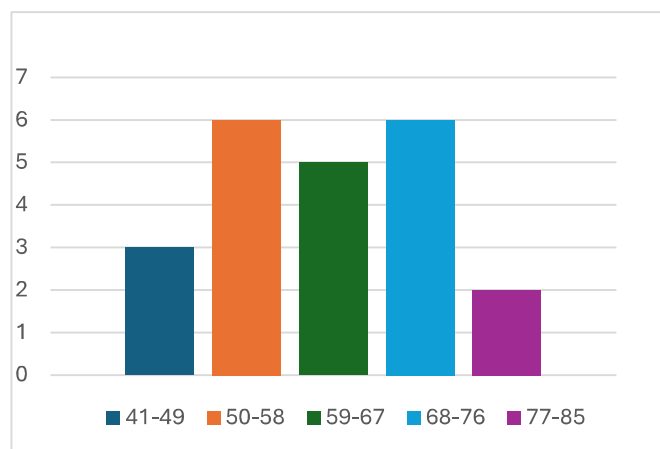


Primjer distribucije frekvencija: Zamislite da analiziramo distribuciju učestalosti ocjena koje su postigli učenici. Prikupili smo podatke od 22 učenika i želimo vidjeti koliko je učenika osvojilo određeni broj bodova.

Studenti – ostvarene ocjene	Zbrojne oznake	Frekvencija
41 - 49		3
50 - 58		6
59 - 67		5
68 - 76		6
77 - 85		2
		Total =22

Slika 1.22 Tablica distribucije frekvencija.

Grafikon distribucije frekvencija (histogram) bi prikazao stupce za svaki raspon ocjena s visinom koja predstavlja broj učenika u svakom razredu frekvencija. Na taj način možemo jasno vidjeti koja je klasa frekvencija najčešća i kako su ostale oznake u skupu podataka raspoređene. Distribucije frekvencija su koristan alat za vizualizaciju i analizu kvalitativnih podataka i za brzo prepoznavanje obrazaca.



Slika 1.23 Grafikon distribucije frekvencija.

1.6 Deskriptivna i inferencijalna statistika

Deskriptivna statistika: deskriptivna statistika bavi se opisom i sažimanjem podataka iz uzorka ili populacije koja se proučava. Koristi se za analizu i razumijevanje podataka, ali ne i za donošenje zaključaka o populaciji kao cjelini. Glavni cilj deskriptivne statistike je opisati karakteristike podataka, na





primjer izračunati srednju vrijednost, medijan, raspon, standardnu devijaciju i stvoriti grafičke prikaze kao što su histogrami ili grafikoni. Koristi se za izradu sažetaka i grafikona koji pomažu u vizualizaciji podataka.

Inferencijalna statistika : inferencijalna statistika bavi se donošenjem zaključaka o populaciji iz uzorka. To znači da inferencijalna statistika omogućuje izvođenje zaključaka o populaciji kao cjelini iz analize uzorka. Koristi različite statističke metode kao što su testiranje hipoteza, intervali pouzdanosti i regresijska analiza kako bi se razumjelo mogu li se promatrani rezultati uzorka generalizirati na populaciju. Na primjer, ako želimo otkriti je li srednja dob u uzorku reprezentativna za populaciju u cjelini, koristit ćemo se inferencijalnom statistikom.

Inferencijalna statistika

Inferencijalna statistika grana je statistike koja se fokusira na zaključke koje možemo izvući iz podataka koje prikupljamo. Glavni zadatak je izvući opće zaključke o populaciji ili uzorku iz analize uzorka podataka.

Glavni ciljevi inferencijalne statistike su:

Procjena parametara populacije: inferencijalna statistika omogućuje nam procjenu parametara populacije kao što su srednja vrijednost, varijanca, proporcije i druge karakteristike iz uzorka.

Testiranje hipoteza: inferencijalna statistika može se koristiti za testiranje hipoteza o populaciji na temelju uzorkovanih podataka. To uključuje statističko testiranje, gdje uspoređujemo uzorak s pretpostavkama o populaciji.

Stvaranje intervala pouzdanosti: inferencijalna statistika omogućuje nam izračunavanje intervala koji sadrže procijenjene vrijednosti parametara populacije s određenom razinom pouzdanosti.

Primjer inferencijalne statistike: pretpostavimo da želimo procijeniti prosječnu visinu svih studenata na sveučilištu. Budući da je nemoguće provjeriti sve učenike, uzimamo uzorak od 100 učenika i mjerimo njihovu visinu.

Zatim koristimo inferencijalnu statistiku za izračunavanje intervala pouzdanosti za prosječnu visinu svih učenika. Naš uzorak ima srednju visinu od 170 cm i standardnu devijaciju od 5 cm.



Uz pretpostavku da su visine učenika u populaciji **približno normalno distribuirane**, možemo koristiti standardnu pogrešku srednje vrijednosti za izračun intervala pouzdanosti. Na primjer, ako želimo interval pouzdanosti od 95%, koristimo standardnu pogrešku i kvantile normalne distribucije.

Približan interval pouzdanosti od 95% za prosječnu visinu svih studenata na sveučilištu bio bi:

$$170 \text{ cm} \pm 1.96 \times \left(\frac{5 \text{ cm}}{\sqrt{100}} \right) = 170 \text{ cm} \pm 0.98 \text{ cm}$$

To znači da s 95%-tnom sigurnošću možemo reći da je prosječna visina svih učenika između približno 169,02 cm i 170,98 cm. Ovaj interval pouzdanosti omogućuje nam da zaključimo o prosječnoj visini svih studenata na sveučilištu iz ukupnog uzorka.

Zajedno ove statističke metode omogućuju logističkim tvrtkama bolje razumijevanje njihovih procesa, predviđanje budućih događaja i donošenje informiranih odluka za poboljšanje učinkovitosti i konkurentnosti.

1.7 Korelacija i regresija

To su statističke metode koje se koriste za proučavanje odnosa između varijabli i predviđanje vrijednosti. Obje metode pomažu razumjeti kako jedna varijabla utječe na drugu i koliko se dobro jedna varijabla može koristiti za predviđanje druge. Evo objašnjenja svake od ove dvije metode:



Korelacija

Korelacija se koristi za mjerenje stupnja povezanosti između dviju kvantitativnih (numeričkih) varijabli. Ona govori postoji li linearna veza između dvije varijable i koliko je jaka ta veza. Korelacija se mjeri koeficijentom korelacije, koji ima oblik **vrijednosti između -1 i 1**.

Koeficijent korelacije 1 znači savršenu pozitivnu korelaciju, što znači da su varijable savršeno korelirane i da se kreću u istom smjeru.

Koeficijent korelacije -1 znači savršenu negativnu korelaciju, što znači da su dvije varijable potpuno obrnuto korelirane i da se kreću u suprotnim smjerovima.

Koeficijent korelacije 0 znači da ne postoji linearna povezanost između varijabli.



Primjer: korelacija između broja sati učenja i ocjena koje studenti postižu bit će pozitivna ako povećanje broja sati učenja obično odgovara višim ocjenama.

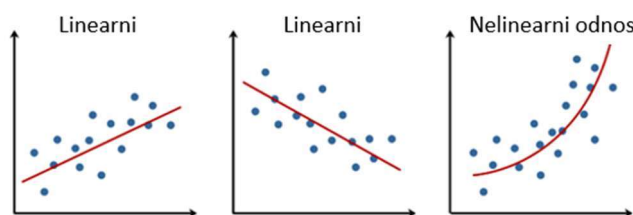
Regresija

Regresija se koristi za modeliranje i predviđanje vrijednosti jedne kvantitativne varijable (zavisne varijable) iz vrijednosti druge kvantitativne varijable (nezavisne varijable). Postoje različite vrste regresije, uključujući **jednostavnu linearnu regresiju, višestruku linearnu regresiju**, logističku regresiju itd.



Jednostavna linearna regresija: koristi se za modeliranje odnosa između jedne nezavisne varijable i jedne zavisne varijable. Model je linearan i obično se prikazuje jednadžbom ravne crte ($y = a + bx$), gdje je a sjecište s y –osi, a b nagib krivulje.

Višestruka linearna regresija: koristi se kada želite modelirati odnos između nekoliko nezavisnih varijabli i jedne zavisne varijable.



Slika 1.24 Grafikon jednostavne linearne regresije.

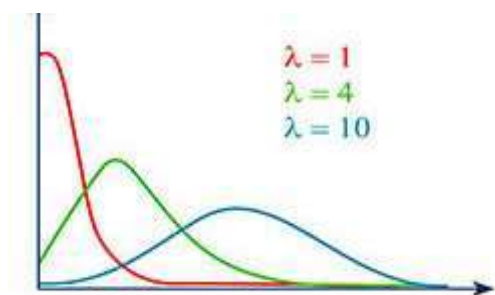
Primjer: jednostavna linearna regresija može se koristiti za modeliranje odnosa između broja obavljenih zadataka učenja (nezavisna varijabla) i završne ocjene ispita (zavisna varijabla).

1.8 Distribucija vjerojatnosti

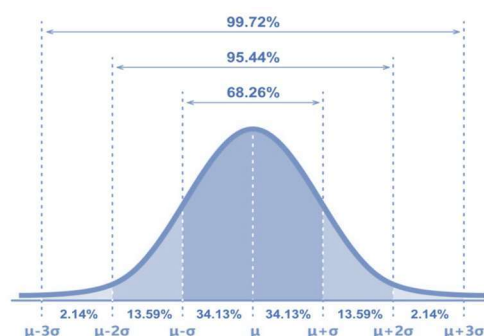
U statistici, distribucija vjerojatnosti opisuje vjerojatnosti različitih vrijednosti koje varijabla može poprimiti. To je matematički model koji nam pomaže razumjeti i analizirati slučajne pojave i predvidjeti kako će se vrijednosti raspodijeliti pod određenim okolnostima. Postoji nekoliko različitih distribucija vjerojatnosti, svaka sa svojim karakteristikama i primjenama u različitim situacijama. Evo nekih od najpoznatijih distribucija vjerojatnosti u statistici:



Normalna (Gaussova) distribucija: normalna distribucija jedna je od najvažnijih i najčešće korištenih distribucija. Opisuje simetričnu i zvonoliku distribuciju s poznatim parametrima: središnjom vrijednošću (μ) i standardnom devijacijom (σ). Mnogi prirodni fenomeni približni su normalnoj distribuciji.



Slika 1.26 Grafikon Poissonove distribucije.



Slika 1.25 Grafikon normalne distribucije.

Binomna distribucija: binomna distribucija koristi se za modeliranje broja uspjeha (npr. broja "glava") u određenom broju neovisnih Bernoullijevih pokusa. Ima dva parametra: broj pokušaja (n) i vjerojatnost uspjeha (p).

Poissonova distribucija: Poissonova distribucija koristi se za modeliranje broja događaja koji se događaju u određenom vremenskom ili prostornom razdoblju. Obično se koristi za modeliranje rijetkih događaja kao što su nesreće, pozivi hitnim službama itd. Parametar distribucije je prosječna stopa (λ).

Eksponencijalna distribucija: eksponencijalna distribucija poseban je slučaj gama distribucije i koristi se za modeliranje vremena do prvog događaja u Poissonovom procesu. Parametar distribucije je prosječna stopa događaja (λ).

Studentova t-distribucija: Studentova t-distribucija koristi se za procjenu intervala pouzdanosti i testiranje hipoteza kada imate mali uzorak i ne znate standardnu devijaciju populacije. To je važno pri analizi uzoraka gdje pretpostavka o normalnoj distribuciji može biti krhka.

Hi-kvadrat distribucija: Hi-kvadrat distribucija koristi se za analizu distribucije frekvencija u tablicama, za testiranje neovisnosti i za testiranje hipoteza. Često se koristi u statističkim testovima kao što je hi-kvadrat test.



F-distribucija: F-distribucija se koristi kada se uspoređuje varijabilnost između dva uzorka. Koristi se u analizi varijance (ANOVA) i drugim statističkim testovima.

Ove distribucije vjerojatnosti temeljni su građevni blokovi u statistici i koriste se za modeliranje i analizu različitih vrsta podataka u različitim kontekstima. Odabir ispravne distribucije vjerojatnosti ključan je pri provođenju statističkih analiza i predviđanju rezultata.

LITERATURA 1. POGLAVLJA

- *Introductory Statistics*. Bentham Science Publishers, Kahl, A. (Publish 2023). DOI:10.2174/97898151231351230101
- *Introductory Statistics 2e*, Openstax, Rice University, Houston, Texas 77005, Jun 23, senior contributing authors: Barbara Illowsky and Susan dean, De anza college, Publish Date: Dec 13, 2023, (<https://openstax.org/details/books/introductory-statistics-2e>);
- *Introductory Statistics 4th Edition*, Susan Dean and Barbara Illowsky, Adapted by Riyanti Boyd & Natalia Casper (Published 2013 by OpenStax College) July 2021, (<http://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf>);
- *Introductory Statistics 7th Edition*, Prem S. Mann, eastern Connecticut state university with the help of Christopher Jay Lacke, Rowan university, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, 2011
- *Journal of the Royal Statistical Society 2024*, A reputable journal publishing cutting-edge research and articles on various aspects of statistics, including theoretical advancements and practical applications. Recent issues have featured studies on sampling and hypothesis testing.
- *Introduction to statistics, made easy second edition*, Prof. Dr. Hamid Al-Oklah Dr. Said Titi Mr. Tareq Alodat, March 2014
- *Statistics for Business and Economics, Thirteenth Edition*, David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran, 2017, 2015 Cengage Learning®



- Statistics for Business, First edition, Derek L Waller, 2008 Copyright © 2008, Derek L Waller, Published by Elsevier Inc. All rights reserved

DODATNE POVEZNICE NA LITERATURU I YOUTUBE VIDEOZAPISE 1. POGLAVLJA

- <https://open.umn.edu/opentextbooks/textbooks/196>
- <https://www.scribbr.com/category/statistics/>
- https://stats.libretexts.org/Bookshelves/Introductory_Statistics
- https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
- https://saylordotorg.github.io/text_introductory-statistics/
- [https://drive.uqu.edu.sa/_/mskhayat/files/MySubjects/20178FS%20Elementary%20Statistics/Introductory%20Statistics%20\(7th%20Ed\).pdf](https://drive.uqu.edu.sa/_/mskhayat/files/MySubjects/20178FS%20Elementary%20Statistics/Introductory%20Statistics%20(7th%20Ed).pdf)
- <https://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf>
- <https://www.geeksforgeeks.org/introduction-of-statistics-and-its-types/>
- https://onlinestatbook.com/Online_Statistics_Education.pdf
- https://www.researchgate.net/profile/Tareq-Alodat-2/publication/340511098_INTRODUCTION_TO_STATISTICS_MADE_EASY/links/5e8de3dc4585150839c7b58a/INTRODUCTION-TO-STATISTICS-MADE-EASY.pdf
- <https://byjus.com/maths/statistics/>
- <https://www.khanacademy.org/math/statistics-probability>
- <https://www.youtube.com/watch?v=XZo4xyJXCak>
- <https://www.youtube.com/watch?v=LMSyiAJm99g>
- https://www.youtube.com/watch?v=VPZD_aij8H0
- <https://www.youtube.com/watch?v=TLwp5DwcqD4>
- <https://www.youtube.com/watch?v=fpFj1Re1l84>
- https://youtube.com/playlist?list=PLqzoL9-eJTNAB5st3mtP_bmXafGSH1Dtz&si=z-IXQ1iKbw2-ieJW



BAS4SC - Business Analytics Skills for the Future-proof Supply Chains

- <https://www.youtube.com/watch?v=44MJyNTxaP8>