



1. Uvodna statistika

1.1 Vloga in pomen statistike pri analizi podatkov v oskrbovalnih verigah

Statistika ima ključno vlogo v sodobnih oskrbovalnih verigah, kjer so učinkovito upravljanje, načrtovanje in nadzor ključnega pomena. Statistične metode se uporabljajo za zbiranje, analizo in razlago podatkov, kar podjetjem omogoča boljše razumevanje in optimizacijo oskrbovalnih verig.

Opišimo nekaj pomembnih vlog statistike pri analizi oskrbovalne verige.

Opisna statistika je ključna za opis osnovnih lastnosti podatkov oskrbovalne verige, kot so povprečje, standardni odklon, mediana, kvartili in druge mere. Ta orodja nam pomagajo razumeti porazdelitev in značilnosti podatkov, kot so povprečni dobavni roki, količine na zalogi in povprečni stroški, kar prispeva k boljšemu razumevanju in upravljanju oskrbovalne verige.

Poleg tega se za napovedovanje prihodnjih dogodkov in trendov v oskrbovalnih verigah uporabljajo statistične tehnike, kot so regresija, analiza časovnih vrst in analiza vzorcev. To vključuje napovedovanje povpraševanja, zalog in dobavnih rokov, kar omogoča boljše načrtovanje in prilagajanje oskrbe.

Statistika ima ključno vlogo pri prepoznavanju vzorcev v podatkih, kar omogoča boljše razumevanje obnašanja oskrbovalne verige, vključno s sezonskimi vzorci, trendi in cikli v povpraševanju.

Optimizacija zalog je še eno ključno področje, na katerem statistični podatki pomagajo določiti optimalne količine naročil, ki zmanjšujejo stroške skladiščenja in naročanja, z uporabo metod, kot je EOQ (Economic Order Quantity).

Poleg tega se statistični podatki uporabljajo tudi za ocenjevanje tveganj v oskrbovalni verigi, kot so verjetnost zamud pri dobavi, poškodbe med prevozom in druge morebitne težave.

S statističnim spremljanjem in nadzorom procesov ugotavljamo odstopanja od standardov, kar nam omogoča izboljšanje kakovosti in učinkovitosti procesov oskrbovalne verige.

Poleg tega se statistični podatki uporabljajo za spremljanje in izboljšanje kakovosti izdelkov in storitev v oskrbovalni verigi, vključno z nadzorom kakovosti pri dobaviteljih.



Statistični podatki so ključno orodje za sprejemanje bolj informiranih odločitev o nabavi, zalogah, izbiri dobaviteljev in drugih vidikih upravljanja oskrbe, kar prispeva k uspešnemu in učinkovitemu delovanju celotne oskrbovalne verige.

Pri analizi oskrbovalne verige se statistični podatki uporabljajo za optimizacijo procesov, zmanjšanje stroškov, povečanje učinkovitosti in izboljšanje zadovoljstva strank. Omogoča boljše razumevanje dinamike oskrbovalne verige in boljše obvladovanje tveganj, kar je ključnega pomena za uspešno delovanje podjetij in organizacij v današnjem globalnem okolju.

1.2 Osnovni pojmi statistike

Spremenljivke

Spremenljivke so osnovni gradniki statistike, saj predstavljajo lastnosti ali značilnosti, ki se merijo, ali opazujejo v raziskavi, poskusu ali vzorcu podatkov. Spremenljivke so bistvene za razumevanje in analizo podatkov, saj raziskovalcem, analitikom in statistikom omogočajo opisovanje, analizo in razumevanje pojavov.



Pomembno je razumeti različne vrste spremenljivk in njihov pomen v statistiki.

Kvalitativne (opisne, kategorične) spremenljivke so spremenljivke, ki predstavljajo kvalitativne značilnosti ali kategorije, ki jih ni mogoče prešteti ali razvrstiti po matematičnem redu. Primeri vključujejo spol (moški, ženska), barvo oči (modra, rjava, zelena) ali tip avtomobila (limuzina, kombi, SUV). Kvalitativne spremenljivke so pogosto uporabne za opisovanje demografskih značilnosti ali lastnosti.

Kvantitativne (numerične) spremenljivke so spremenljivke, ki predstavljajo numerične vrednosti, ki jih je mogoče prešteti ali izmeriti in jih je mogoče razvrstiti v nekakšen matematični red. Primeri vključujejo starost, višino, temperaturo, dohodek ali rezultate raziskav. Kvantitativne spremenljivke se pogosto uporabljajo za analizo in kvantitativno raziskovanje pojavov.

Odvisne in neodvisne spremenljivke. Odvisna spremenljivka je tista, ki jo želimo raziskati, izmeriti ali napovedati, medtem ko je neodvisna spremenljivka tista, ki naj bi vplivala na odvisno spremenljivko. Če želimo na primer raziskati, ali izobrazba vpliva na dohodek, je dohodek odvisna spremenljivka, izobrazba pa neodvisna spremenljivka.

Diskretne in zvezne spremenljivke. Spremenljivke lahko delimo tudi na diskretne in zvezne. Diskretne spremenljivke imajo omejen nabor možnih vrednosti in so običajno predstavljene s celimi



števíli. Primer je število otrok v družini, kjer so možne vrednosti 0, 1, 2 itd. Po drugi strani pa imajo zvezne spremenljivke neskončno število možnih vrednosti in se običajno merijo z decimalnimi števíli. Primer je višina oseb, kjer je v danem območju možnih neskončno število vrednosti.

Spremenljivke so osnovna orodja za raziskave in analizo podatkov. Razumevanje in pravilna opredelitev spremenljivk sta ključna za izvajanje statističnih analiz in preučevanje pojavov v raziskavah. Spremenljivke raziskovalcem omogočajo izražanje in količinsko opredelitev različnih vidikov resničnosti, kar omogoča boljše razumevanje pojavov, sprejemanje odločitev in napovedovanje prihodnjih dogodkov. Omogočajo tudi uporabo različnih statističnih tehnik za preverjanje hipotez, napovedovanje in boljše razumevanje vzročnih povezav med spremenljivkami.

1.3 Osnovni statistični koncepti s primeri

Povprečje (srednja vrednost)

Srednja vrednost, znana tudi kot **povprečje**, je ena od osnovnih statističnih mer. Povprečje je aritmetično povprečje vseh vrednosti v podatkovni zbirki. Izračunamo jo tako, da seštejemo vse podatke in jih nato delimo s številom podatkov.



Izračun povprečja:

- Seštejte vse vrednosti v naboru podatkov.
- Vsoto delite s številom vrednosti v nizu.
- Enačba za izračun povprečja (\bar{x}) je: $\bar{x} = (x_1 + x_2 + x_3 + \dots + x_n) / n$

Kje: \bar{x} je povprečje. $x_1, x_2, x_3, \dots, x_n$ so vrednosti v naboru podatkov. n je število vrednosti v naboru podatkov.

Primer:

Predstavljajte si nabor podatkov, ki predstavlja ocene študentov pri izpitu iz matematike: 80, 85, 90, 75, 95. Če želite izračunati povprečje, seštejte vse te vrednosti in jih delite s številom ocen, ki je v tem primeru 5:

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

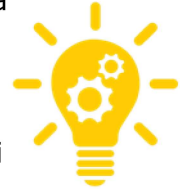
Tako je povprečna ocena učenca 85 točk. Povprečje je uporabno za merjenje osrednje tendence podatkov in nam daje grobo predstav o tem, kaj lahko pričakujemo kot "tipično" vrednost v nizu podatkov. Vendar se lahko povprečje precej spremeni, če so v podatkih prisotna odstopanja ali



izstopajoče vrednosti. Zato je pomembno poznati tudi druge statistične mere, kot sta mediana in modus, da bi boljše razumeli porazdelitev podatkov.

Mediana

Mediana je statistični pojem, ki se uporablja za merjenje srednje vrednosti niza podatkov. To je vrednost, ki razdeli urejene podatke na dve enaki polovici. To pomeni, da ima polovica podatkov vrednosti, ki so manjši ali enake mediani, druga polovica pa vrednosti, ki so večje ali enake mediani. Mediana je ena od osnovnih mer centralne tendence v statistiki in se uporablja za opis porazdelitve podatkov, zlasti kadar so podatki poševni ali vsebujejo izstopajoče vrednosti.



Kako izračunati mediano:

- Najprej morate niz podatkov razvrstiti od najmanjše do največje vrednosti.
- Če je število podatkov sodo (n), potem je mediana povprečje dveh srednjih vrednosti. To pomeni, da je mediana enaka povprečju vrednosti na položaju $n/2$ in $(n/2 + 1)$, ko so podatki razvrščeni v naraščajočem vrstnem redu.
- Če je število podatkov liho, je vrednost mediane na sredini.

Primer:

Predstavljajte si naslednji niz podatkov o številu ur spanja v določenem obdobju: 7, 6, 5, 8, 6, 9, 7

Najprej podatke razvrstite v naraščajočem vrstnem redu: 5, 6, 6, 7, 7, 8, 9

Ker je število podatkov liho (7), bo mediana vrednost na srednjem mestu, ki je četrta vrednost v urejenem nizu podatkov. Torej je mediana v tem primeru enaka 7 ur. To pomeni, da polovica ljudi v tem podatkovnem nizu spi 7 ur ali manj, druga polovica pa 7 ur ali več.

Modus

Modus je ena od osnovnih statističnih metrik, ki se uporablja za merjenje osrednje tendence niza podatkov. Modus predstavlja vrednost, ki se najpogosteje pojavlja v naboru podatkov. To je vrednost, ki ima največjo pogostost pojavljanja med vsemi vrednostmi v podatkovnem nizu.

Modus je uporaben za ugotavljanje najpogostejše vrednosti v nizu podatkov in je še posebej uporaben pri analizi kvalitativnih (kategoričnih) spremenljivk, kjer vrednosti niso številčne.

Če je v nizu podatkov več načinov (več vrednosti s podobno največjo frekvenco), govorimo o več-modalni porazdelitvi. Če imajo vsi podatki enako pogostost pojavljanja, potem niz podatkov nima načina.



Primer: predstavljajte si nabor podatkov, ki predstavlja barve avtomobilov na parkirišču:

Rdeča, modra, rdeča, zelena, modra, modra, modra, rdeča

V tem primeru je modus "rdeča", saj se ta vrednost pojavlja najpogosteje (trikrat), medtem ko se "modra" in "zelena" pojavljata manj pogosto.

Modus je preprosto izračunati, saj preprosto določi vrednost z največjo pogostostjo pojavljanja v naboru podatkov. Modus se uporablja za opis značilnih vrednosti v podatkih in je lahko koristen pri razumevanju, katera vrednost je najbolj značilna za določeno situacijo ali skupino.

Razpon variance (VR, Razpon, razpon)

Razlika med največjo in najmanjšo vrednostjo v nizu podatkov je statistični pojem, ki se imenuje razpon. Ta meri, kako velika je razlika med največjo (maksimalno) in najmanjšo (minimalno) vrednostjo v podatkovnem nizu. Razpon je preprost način za oceno razpona vrednosti v podatkovnem nizu in za merjenje variabilnosti med najmanjšo in največjo vrednostjo.

Izračun meje variacije je preprost:

- Najprej poiščite najmanjšo vrednost (min) in največjo vrednost (max) v naboru podatkov.
- Nato izračunajte razliko med največjo in najmanjšo vrednostjo (max - min).

Primer: predstavljajte si nabor podatkov, ki predstavlja starost udeležencev nekega dogodka: 20, 25, 30, 35, 40. Za izračun meje variacije najprej poiščite najmanjšo vrednost (20) in največjo vrednost (40) v nizu podatkov. Nato izračunajte razliko med največjo in najmanjšo vrednostjo: $VR = 40 - 20 = 20$

Tako je v tem primeru razpon odstopanja 20 let. To pomeni, da je razlika med najstarejšim in najmlajšim udeležencem 20 let.

Dekompozicija variance je uporabna za ocenjevanje razpona vrednosti v naboru podatkov, vendar je precej preprosta in ne upošteva vseh vrednosti v naboru podatkov. Za podrobnejšo analizo variabilnosti in razpršenosti podatkov se običajno uporabljajo druge statistične mere, kot so variance ali kvartili.

Varianca in standardni odklon

Varianca je povprečje kvadratnih odstopanj od povprečja. Je kvadrat standardnega odklona. **Standardni odklon** je statistična mera, ki se uporablja za merjenje razpršenosti ali variabilnosti v nizu podatkov. Pove, kako daleč so vrednosti od povprečja v nizu. Standardni odklon je ena



najpogosteje uporabljenih mer razpršenosti v statistiki in se izračuna z izračunom kvadratnega korena variacije (variance).

Izračun standardnega odklona:

- Najprej izračunajte variacijo (raztros). Variacijo (raztros) izračunamo tako, da za vsako vrednost v nizu vzamemo povprečje vseh vrednosti v nizu, nato pa te razlike kvadriramo in seštejemo.
- Ko dobite vrednost meje variacije (σ^2), izračunajte standardni odklon z izračunom kvadratnega korena meje variacije. To naredimo tako, da vzamemo kvadratni koren iz σ^2 :

Standardni odklon $\sigma = \sqrt{\sigma^2}$

Standardni odklon meri, kako razpršene so vrednosti okoli povprečja v nizu podatkov. Večja vrednost standardnega odklona pomeni, da so vrednosti bolj razpršene in se bolj razlikujejo od povprečja, medtem ko nižja vrednost standardnega odklona pomeni manjšo razpršenost.



Primer: Predstavljajte si nabor podatkov, ki predstavlja ocene študentov pri izpitu iz matematike: 80, 85, 90, 75, 95. Formula, ki jo bomo predstavili v nadaljevanju, velja le, če pet vrednosti, s katerimi smo začeli, tvori celotno populacijo. Najprej izračunajte povprečje, ki je 85. Nato izračunajte variantno razliko, ki je 50.

Najprej izračunajte odstopanja vsake podatkovne točke od povprečja in rezultat vsakega od njih izravnajte s kvadratom:

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

Varianca je povprečje teh vrednosti:

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

Na koncu izračunate standardni odklon tako, da vzamete kvadratni koren meje variacije:

$$\text{Standardni odklon} = \sqrt{50} \approx 7.07$$

Standardni odklon v tem primeru znaša približno 7,07. To pomeni, da so rezultati učencev v povprečju za približno 7,07 enote oddaljeni od povprečja. Standardni odklon se pogosto uporablja pri analizi porazdelitve podatkov in ocenjevanju variabilnosti vrednosti v nizu.



Kvantili

Kvantili so vrednosti, ki razdelijo urejene podatke na določene dele. Kvartili na primer delijo podatke na štiri enake dele. Prvi kvartil (Q1) deli spodnjih 25 % podatkov, drugi kvartil (Q2) je enak mediani, tretji kvartil (Q3) pa deli zgornjih 25 % podatkov.



Primer: v naboru podatkov 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 je prvi kvartil (Q1) enak 6, drugi kvartil (Q2) je enak 11, tretji kvartil (Q3) pa je enak 16.

1.4 Prikaz statističnih podatkov

Predstavitev statističnih podatkov vključuje uporabo različnih metod in orodij, da bi podatke predstavili na jasn, pregleden in informativen način.

Tukaj je nekaj pogostih načinov za prikaz statističnih podatkov:

Tabele

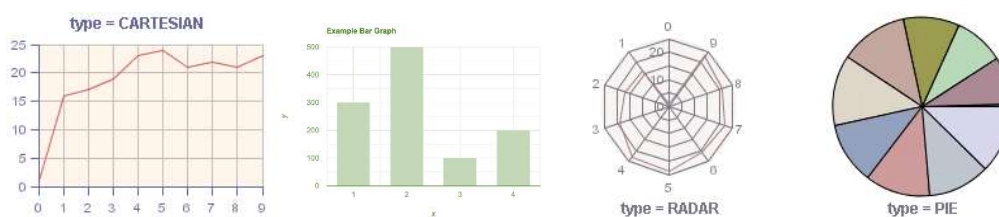
Tabele so osnovna metoda za prikazovanje podatkov. Primeri so tabele pogostosti, ki prikazujejo število pojavitev različnih vrednosti, in podatkovne tabele, ki prikazujejo več informacij o podatkih.

Marks Scored by Students	Tally Marks	Frequency
41 - 49		3
50 - 58		6
59 - 67		5
68 - 76		6
77 - 85		2
		Total =22

Slika 1.1 Primer tabele.

Grafični prikazi

Grafični prikazi so učinkovito orodje za vizualizacijo podatkov. Vključujejo različne vrste grafov, kot so črtni, linijski, krožni, histogramski, škatlasti itd.



Slika 1.2 Primeri grafičnih predstavitev podatkov.



Črtni diagrami se uporabljajo za vizualizacijo trendov in sprememb čez čas, zato so idealni za spremljanje podatkov, ki se nenehno spreminjajo. Posebej učinkoviti so za prikaz odnosov med spremenljivkami in poudarjanje vzorcev, kot so povečanja, zmanjšanja ali nihanja. Črtni diagrami se pogosto uporabljajo na področjih, kot so finance, znanost in poslovanje, za analizo podatkov časovnih vrst, primerjavo trendov med kategorijami ali napovedovanje prihodnjega razvoja na podlagi preteklih podatkov.

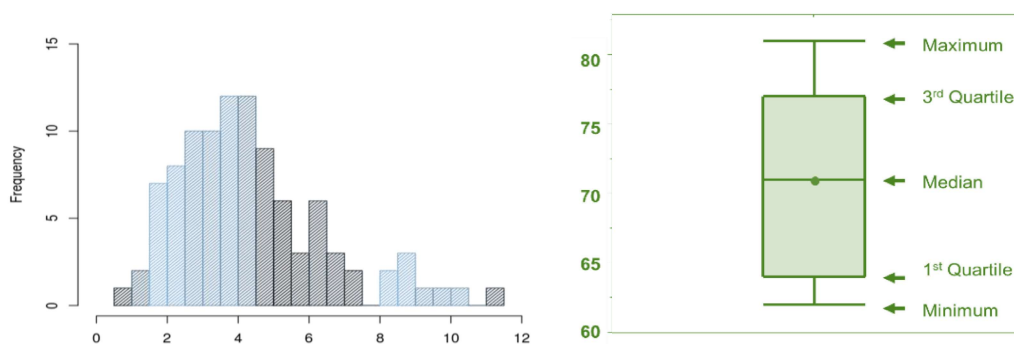
Stolpčni diagrami se uporabljajo za primerjavo količin v različnih kategorijah, zato so idealni za predstavitev diskretnih podatkov. Posebej učinkoviti so za poudarjanje razlik, podobnosti in trendov med skupinami. Stolpčni diagrami se pogosto uporabljajo, kadar je treba jasno in vizualno preprosto prikazati frekvence, odstotke ali druge številčne mere. Pogosto se uporabljajo v podjetjih, izobraževanju in raziskavah za analizo in sporočanje kategoričnih podatkov.

Radarski diagrami, znani tudi kot pajkovi diagrami, se uporabljajo za prikaz večrazsežnostnih podatkov v več dimenzijah v krožni obliki. Idealni so za primerjavo več spremenljivk ali entitet glede na ista merila, pri čemer jasno in vizualno poudarijo prednosti in slabosti. Radarski diagrami se pogosto uporabljajo pri analizi uspešnosti, sprejemanju odločitev in primerjavah s konkurenco, na primer pri ocenjevanju lastnosti izdelkov, sposobnosti ekipe ali rezultatov raziskav v različnih kategorijah.

Krožni diagrami se uporabljajo za prikaz deležev ali odstotkov celote, zato so idealni za vizualizacijo relativnih velikosti različnih kategorij. Še posebej učinkoviti so, kadar želite prikazati, kako deli prispevajo k celoti, ali kadar želite na prvi pogled primerjati razmerja. Krožni diagrami se pogosto uporabljajo v poročilih, predstavitvah in raziskavah za prikaz podatkov, kot so tržni delež, razdelitev proračuna ali demografska porazdelitev.

Histogrami

Histogrami so grafični prikazi porazdelitve podatkov. Uporabljajo se za prikaz pogostosti vrednosti spremenljivke v različnih časovnih intervalih.



Slika 1.3 Histogram in kvantilni graf (škatlasti graf).

Kvantilni grafikon (okvirni grafikon)

Kvantilni graf ali škatla z usti je vrsta grafa, ki se uporablja v opisni statistiki kot priročen način grafične predstavitve skupin številčnih podatkov, tako da jih povzamemo s petimi številkami: minimum, prvi kvartil, mediana, tretji kvartil in maksimum.

Izbira metode za prikazovanje statističnih podatkov je odvisna od narave podatkov, ciljev analize in ciljnega občinstva. Pomembno je, da izberete metodo, ki najbolj ustreza vašemu sporočilu in omogoča lažje razumevanje podatkov.

1.5 Porazdelitev pogostosti

Frekvenčna porazdelitev, znana tudi kot frekvenčna tabela ali histogram, je način prikaza števila pojavitev različnih vrednosti spremenljivke v nizu podatkov. S frekvenčno porazdelitvijo lahko ugotovite vzorce, porazdelitve in frekvence vrednosti v podatkih. Pogosto se uporablja za analizo kvalitativnih (kategoričnih) spremenljivk, lahko pa se uporablja tudi za prikaz diskretnih vrednosti kvantitativnih (numeričnih) spremenljivk.



Postopek ustvarjanja frekvenčne porazdelitve vključuje naslednje korake:

- Zbiranje podatkov: najprej zberite podatke, za katere želite ustvariti frekvenčno porazdelitev.
- Opredelitev različnih vrednosti: opredelite različne vrednosti, ki se pojavljajo v vaših podatkih. To so kategorije ali diskretne vrednosti, ki jih želite analizirati.
- Štetje pojavitev: preštejte, kolikokrat se posamezna vrednost pojavi v naboru podatkov.



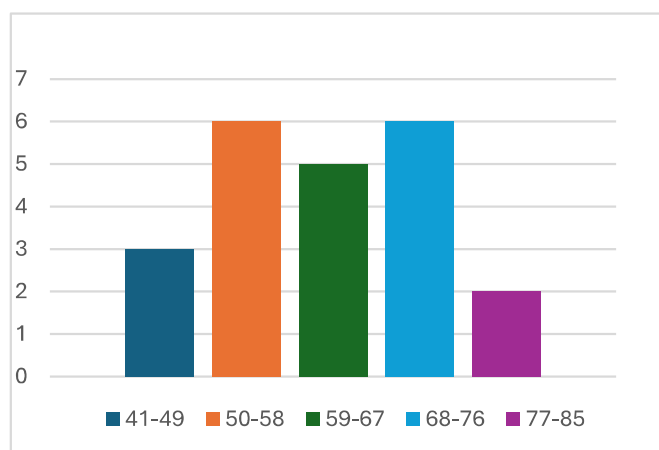
- Ustvarite frekvenčno tabelo: ustvarite tabelo, ki prikazuje vse različne vrednosti spremenljivke in število pojavitev za vsako vrednost.
- Risanje histograma: če imate veliko število različnih vrednosti, lahko ustvarite histogram, ki prikazuje frekvenčno porazdelitev. To je grafični prikaz, ki prikazuje število pojavitev vsake vrednosti v obliki stolpcev.

Primer frekvenčne porazdelitve: Predstavljajte si, da analiziramo frekvenčno porazdelitev ocen, ki so jih dosegli učenci. Zbrali smo podatke o 22 študentih in želimo ugotoviti, koliko študentov je doseglo določeno število točk.

Marks Scored by Students	Tally Marks	Frequency
41 - 49		3
50 - 58		6
59 - 67		5
68 - 76		6
77 - 85		2
		Total =22

Slika 1.4 Tabela porazdelitve pogostosti.

Graf frekvenčne porazdelitve (histogram) bi prikazal stolpce za vsako območje ocen, višina pa bi predstavljala število učencev v vsakem frekvenčnem razredu. Tako lahko jasno vidimo, kateri frekvenčni razred je najpogostejši in kako so porazdeljene druge ocene v podatkovnem nizu. Frekvenčne porazdelitve so uporabno orodje za vizualizacijo in analizo kvalitativnih podatkov ter za hitro prepoznavanje vzorcev.

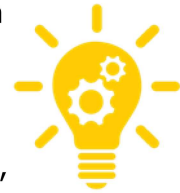


Slika 1.5 Graf porazdelitve pogostosti.



1.6 Opisna in sklepalna statistika

Opisna statistika: opisna statistika se ukvarja z opisovanjem in povzemanjem podatkov iz preučevanega vzorca ali populacije. Uporablja se za analizo in razumevanje podatkov, ne pa za sklepanje o celotni populaciji. Glavni cilj opisne statistike je opisati značilnosti podatkov, na primer izračunati povprečje, mediano, razpon, standardni odklon in oblikovati grafične predstavitve, kot so histogrami ali grafi. Uporablja se za izdelavo povzetkov in grafov, ki pomagajo vizualizirati podatke.



Inferenčna statistika: inferenčna statistika se ukvarja s sklepanjem o populaciji na podlagi vzorca. To pomeni, da inferenčna statistika omogoča, da se na podlagi analize vzorca sklepa o celotni populaciji. Uporablja različne statistične metode, kot so preverjanje hipotez, intervali zaupanja in regresijska analiza, da bi ugotovila, ali je mogoče rezultate opazovanega vzorca posplošiti na populacijo. Če želimo na primer ugotoviti, ali je povprečna starost v vzorcu reprezentativna za celotno populacijo, bomo uporabili inferenčno statistiko.

Inferenčna statistika

Inferenčna statistika je veja statistike, ki se osredotoča na sklepe in zaključke, ki jih lahko potegnemo iz zbranih podatkov. Njena glavna naloga je, da na podlagi analize vzorca podatkov oblikuje splošne sklepe o populaciji ali vzorcu.

Glavni cilji inferenčne statistike so:

Ocenjevanje populacijskih parametrov: inferenčna statistika nam omogoča ocenjevanje populacijskih parametrov, kot so povprečje, variance, deleži in druge značilnosti, na podlagi vzorca.

Preverjanje hipotez: inferenčna statistika se lahko uporablja za preverjanje hipotez o populaciji na podlagi vzorčenih podatkov. Pri tem gre za statistično testiranje, pri katerem vzorec primerjamo s predpostavkami o populaciji.

Oblikovanje intervalov zaupanja: inferenčna statistika nam omogoča izračun intervalov, ki vsebujejo ocenjene vrednosti populacijskih parametrov z določeno stopnjo zaupanja.

Primer inferenčne statistike: recimo, da želimo oceniti povprečno višino vseh študentov na univerzi. Ker je nemogoče preveriti vse študente, vzamemo vzorec 100 študentov in izmerimo njihovo višino.

Nato s pomočjo inferenčne statistike izračunamo interval zaupanja za povprečno višino vseh učencev. Naš vzorec ima povprečno višino 170 cm in standardni odklon 5 cm.



Ob predpostavki, da so višine učencev v populaciji **približno normalno porazdeljene**, lahko za izračun intervala zaupanja uporabimo standardno napako povprečja. Če na primer želimo 95-odstotni interval zaupanja, uporabimo standardno napako in kvantile normalne porazdelitve.

Približni 95-odstotni interval zaupanja za povprečno višino vseh študentov na univerzi bi bil:

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

To pomeni, da lahko s 95-odstotno zanesljivostjo trdimo, da je povprečna višina vseh učencev približno med 169,02 cm in 170,98 cm. Ta interval zaupanja nam omogoča, da iz celotnega vzorca sklepamo o povprečni višini vseh študentov na univerzi.

Te statistične metode skupaj logističnim podjetjem omogočajo boljše razumevanje procesov, napovedovanje prihodnjih dogodkov in sprejemanje bolj premišljenih odločitev za izboljšanje učinkovitosti in konkurenčnosti.

1.7 Korelacija in regresija

To so statistične metode, ki se uporabljajo za preučevanje odnosov med spremenljivkami in napovedovanje vrednosti. Obe metodi pomagata razumeti, kako ena spremenljivka vpliva na drugo in kako dobro lahko eno spremenljivko uporabimo za napovedovanje druge. Tukaj je razlaga vsake od teh dveh metod:



Korelacija

Korelacija se uporablja za merjenje stopnje povezanosti med dvema kvantitativnima (številčnima) spremenljivkama. Pove, ali med spremenljivkama obstaja linearna povezava in kako močna je ta povezava. Korelacija se meri s korelacijskim koeficientom, ki ima obliko **vrednosti med -1 in 1**.

Korelacijski koeficient 1 pomeni popolno pozitivno korelacijo, kar pomeni, da sta spremenljivki popolnoma povezani in se gibljeta v isti smeri.

Korelacijski koeficient -1 pomeni popolno negativno korelacijo, kar pomeni, da sta spremenljivki popolnoma obratno povezani in se gibljeta v nasprotnih smereh.

Korelacijski koeficient 0 pomeni, da med spremenljivkama ni linearne povezave.

Primer: korelacija med številom ur študija in doseženimi ocenami študentov bo pozitivna, če povečanje števila ur študija običajno pomeni višje ocene.



Regresija

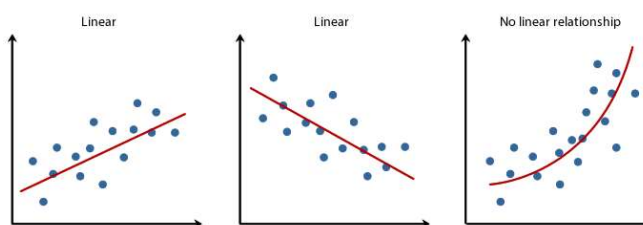
Regresija se uporablja za modeliranje in napovedovanje vrednosti ene kvantitativne spremenljivke (odvisne spremenljivke) na podlagi vrednosti druge kvantitativne spremenljivke (neodvisne spremenljivke). Obstajajo različne vrste regresije, vključno z **enostavno linearno regresijo, multiplo linearno regresijo**, logistično regresijo itd.



Enostavna linearna regresija: uporabljajo se za modeliranje razmerja med eno neodvisno spremenljivko in eno odvisno spremenljivko. Model je linearen in je običajno predstavljen z enačbo ravne črte ($y = a + bx$), kjer je a presečišče z y -osjo in b je naklon premice.

Večkratna linearna regresija: uporablja se, kadar želite modelirati razmerje med več neodvisnimi spremenljivkami in eno odvisno spremenljivko.

Primer: za modeliranje razmerja med številom opravljenih učnih nalog (neodvisna spremenljivka) in končno oceno izpita (odvisna spremenljivka) se lahko uporabi preprosta linearna regresija.



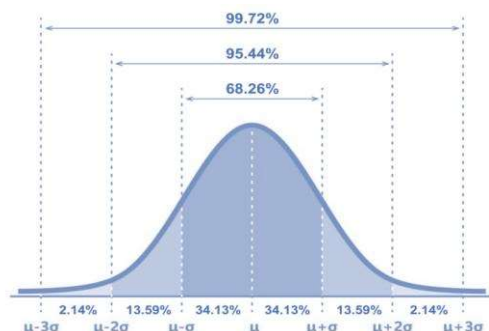
Slika 1.6 Grafi preproste linearne regresije.

1.8 Verjetnostne porazdelitve

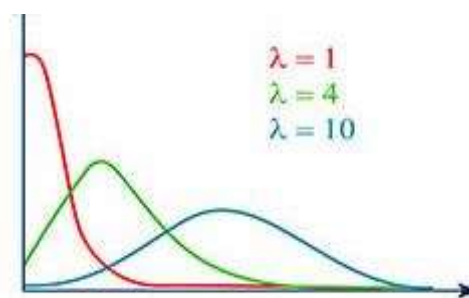
V statistiki verjetnostna porazdelitev opisuje verjetnosti različnih vrednosti, ki jih spremenljivka lahko doseže. Je matematični model, ki nam pomaga razumeti in analizirati naključne pojave ter napovedati, kako bodo vrednosti porazdeljene v določenih okoliščinah. Obstaja več različnih verjetnostnih porazdelitev, od katerih ima vsaka svoje značilnosti in se uporablja v različnih situacijah. Tukaj je nekaj najbolj znanih verjetnostnih porazdelitev v statistiki:



Normalna (Gaussova) porazdelitev: normalna porazdelitev je ena najpomembnejših in najpogosteje uporabljenih porazdelitev. Opisuje simetrično in zvonasto porazdelitev z znanima parametroma: sredino (μ) in standardnim odklonom (σ). Veliko naravnih pojavov se približuje normalni porazdelitvi.



Slika 1.8 Graf normalne porazdelitve.



Slika 1.7 Graf Poissonove porazdelitve.

Binomska porazdelitev: binomska porazdelitev se uporablja za modeliranje števila uspehov (npr. števila "glav") v danem številu neodvisnih Bernoullijevih poskusov. Ima dva parametra: število poskusov (n) in verjetnost uspeha (p).

Poissonova porazdelitev: Poissonova porazdelitev se uporablja za modeliranje števila dogodkov, ki se zgodijo v določenem časovnem ali prostorskem obdobju. Običajno se uporablja za modeliranje redkih dogodkov, kot so nesreče, klici na reševalne službe itd. Parameter porazdelitve je povprečna stopnja (λ).

Eksponentna porazdelitev: eksponentna porazdelitev je poseben primer gama porazdelitve in se uporablja za modeliranje časov do prvega dogodka v Poissonovem procesu. Parameter porazdelitve je povprečna stopnja dogodkov (λ).

Studentova t-razdelitev: Studentova t-razdelitev se uporablja za ocenjevanje intervalov zaupanja in preverjanje hipotez, kadar imamo majhen vzorec in ne poznamo standardnega odklona populacije. Pomembna je pri analizi vzorcev, kjer je predpostavka o normalni porazdelitvi lahko krhka.

Porazdelitev Chi-kvadrat: porazdelitev Chi-kvadrat se uporablja za analizo porazdelitve pogostosti v tabelah, za preverjanje neodvisnosti in preverjanje hipotez. Pogosto se uporablja pri statističnih testih, kot je test chi-kvadrat.

Porazdelitev F: porazdelitev F se uporablja pri primerjavi variabilnosti med dvema vzorcema. Uporablja se pri analizi variance (ANOVA) in drugih statističnih testih.

Te verjetnostne porazdelitve so temeljni gradniki statistike in se uporabljajo za modeliranje in analizo različnih vrst podatkov v različnih kontekstih. Izbira pravilne verjetnostne porazdelitve je ključna pri izvajanju statističnih analiz in napovedovanju rezultatov.



LITERATURA POGLAVJE 1

- 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>);
- 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.
- 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
- 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 POVEZAVE DO LITERATURE IN VIDEOPOSNETKOV NA YOUTUBU POGLAVJE 1

- <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/oe/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_aj8H0
- <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
- <https://www.youtube.com/watch?v=44MJyNTxaP8>