

Analiza varijance

Mijatović, Katarina

Undergraduate thesis / Završni rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Pula / Sveučilište Jurja Dobrile u Puli**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:137:801000>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-24**



Repository / Repozitorij:

[Digital Repository Juraj Dobrila University of Pula](#)



Sveučilište Jurja Dobrile u Puli
Fakultet ekonomije i turizma
«Dr. Mijo Mirković»

KATARINA MIJATOVIĆ

ANALIZA VARIJANCE

Završni rad

Pula, 2023.

Sveučilište Jurja Dobrile u Puli
Fakultet ekonomije i turizma
«Dr. Mijo Mirković»

KATARINA MIJATOVIĆ

ANALIZA VARIJANCE

Završni rad

JMBAG: 0303078798, redovita studentica

Studijski smjer: Ekonomija

Predmet: Statistika

Znanstveno područje: Društvene znanosti

Znanstveno polje: Ekonomija

Znanstvena grana: Kvantitativna ekonomija

Mentorica: doc. dr. sc. Katarina Kostelić

Pula, rujan 2023.



IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani _____, kandidat za prvostupnika ekonomije/poslovne ekonomije, smjera _____ ovime izjavljujem da je ovaj Završni rad rezultat isključivo mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na nedozvoljen način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student

U Puli, _____, _____ godine



IZJAVA

o korištenju autorskog djela

Ja, _____ dajem odobrenje Sveučilištu
Jurja Dobrile

u Puli, kao nositelju prava iskorištavanja, da moj završni rad pod nazivom

_____ koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama.

Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, _____ (datum)

Potpis

Sadržaj

1. Uvod	1
2. Aritmetička sredina i varijanca	3
3. Osnove inferencijalne statistike i teorijske distribucije koje se koriste pri analizi varijance	10
3.1. Normalna distribucija.....	15
3.2. Studentova t - distribucija	18
3.3. Snedecorova f – distribucija	20
4. T – test.....	22
4.1. Pretpostavke t-testa	25
4.2. Primjer provedbe t-testa	26
5. F – test.....	29
5.1. Pretpostavke f-testa	31
5.2. Primjer provedbe f-testa	31
6. Univarijatna analiza varijance	33
6.1. Jednofaktorska analiza varijance	34
6.2. Dvofaktorska analiza varijance	38
6.3. Primjeri provedbe ANOVA-e	40
7. Primjena analize varijanci u ekonomiji/poslovnoj ekonomiji	45
8. Zaključak	50
Literatura	51
Popis slika	55
Popis tablica	56
Sažetak.....	57
Summary	58

1. Uvod

Statistika je znanstvena disciplina koja proučava metode prikupljanja, analize, prikaza i tumačenja podataka. Područje statistike obuhvaća deskriptivnu i inferencijalnu statistiku. Deskriptivna statistika analizira mjere centralne tendencije (poput aritmetičke sredine, medijana i moda), mjere disperzije i zaobljenosti te koristi grafikone i tablice kako bi vizualno predstavila osnovne statističke karakteristike. S druge strane, inferencijalna statistika služi za procjene i testiranje postavljenih hipoteza (nul i alternativna) koristeći statističke testove, koeficijente i utvrđivanje značajnosti.

Analiza varijance obuhvaća širok spektar statističkih metoda koje služe za uspoređivanje varijanci i srednjih vrijednosti više skupina kako bi se otkrile statistički značajne razlike. ANOVA (ANalysis Of VAriance) model je specifični statistički model koji se koristi za provođenje analize varijance. Ovaj model pretpostavlja linearnu vezu između zavisne varijable i ispitivanih faktora te omogućuje procjenu utjecaja svakog pojedinog faktora i njihovih međusobnih interakcija na varijabilnost zavisne varijable. Bitni aspekti ANOVA modela uključuju izračune poput F-testa i p-vrijednosti kako bi se odredila statistički značajna razlika među skupinama. Budući da se često koristi, modeli analize varijance (ANOVA) su postali ključni alati za analizu kompleksnih multivarijantnih podataka i gotovo su sinonim za analizu varijance u statistici.

Pri analizi varijance primjenjuju se različite metode, ovisno o specifičnom istraživačkom dizajnu i postavljenim hipotezama. Jednofaktorska ANOVA se koristi za uspoređivanje prosjeka tri ili više neovisnih skupina s obzirom na jedan faktor. Faktor je kategorička nezavisna varijabla koja definira različite grupe ili razine koje se uspoređuju u analizi. Ta varijabla označava uvjete ili kategorije pod kojima se promatranja prikupljaju i omogućava grupiranje podataka radi usporedbe. Na primjer, ako promatramo plaće zaposlenika u tvrtki, potencijalni faktor mogla bi biti njihova radna mjesta ili završena razina obrazovanja. Kategoriziranjem zaposlenika prema njihovim radnim mjestima ili razini obrazovanja može se analizirati postoje li značajne razlike u plaćama među tim kategorijama. Općenito, ANOVA pomaže u otkrivanju statistički značajnih razlika među skupinama.

Dvofaktorska analiza se provodi kako bi ispitaio utjecaj dvaju nezavisnih faktora i njihovih međudjelovanja na zavisnu varijablu. Nadalje, postoji i višekratna analiza varijance (MANOVA). Vezane metode koje premašuju cilj ovog rada odnose se na analizu kovarijanci (ANCOVA) i opće linearno modeliranje (GLM).

Cilj ovog rada jest pružiti sveobuhvatno razumijevanje analize varijance (ANOVA) i njenih različitih aspekata, uključujući jednofaktorski i dvofaktorski model te relevantne pretpostavke. Prije nego što se detaljno razmotri analiza varijance, bit će objašnjena aritmetička sredina i njezina svojstva, a također će biti prikazane Gaussova, Studentova i Snedecorova distribucija.

Osnovna svrha rada je predstaviti analizu varijance kao važan statistički alat za uspoređivanje više grupa i procjenu utjecaja različitih faktora na varijabilnost u skupovima podataka, s osvrtom na primjenu u ekonomiji i poslovnoj ekonomiji. Stoga će se u okviru ovog rada navedena teza razraditi putem temeljitog objašnjenja koncepta ANOVA-e, uz primjere i matematičke formulacije. Metode koje će biti primijenjene obuhvaćaju teorijski pregled različitih analitičkih alata uključenih u ANOVA-u, pružanje primjera te analizu stvarnih rezultata putem tumačenja dobivenih rezultata.

2. Aritmetička sredina i varijanca

Aritmetička sredina je mjera središnje tendencije. Omogućuje određivanje središta distribucije kvantitativne varijable uzimajući u obzir sva opažanja. Prema Plackettu (1958.), koncept aritmetičke sredine potječe od grčkog astronoma Hiparha¹. Godine 1755. Thomas Simpson službeno je predložio uporabu aritmetičke sredine u pismu predsjedniku Kraljevskog društva² što je populariziralo njezinu primjenu.

Poznata je i kao prosjek koji dolazi od engleske riječi „average“ što u hrvatskom prijevodu znači prosječno. Ona je najčešća i najraširenija korištena mjera tendencije. Najlakše rečeno ona predstavlja zbroj vrijednosti svih opažanja podijeljen s ukupnim brojem opažanja. Na brojevnom pravcu aritmetička sredina je točka oko koje je zbroj pozitivnih i negativnih odstupanja jednak nuli. Mjerna jedinica aritmetičke sredine jednaka je mjernoj jedinici varijable za koju se izračunava.³

Označava se različito ovisno o tome je li riječ o populaciji ili uzorku. Aritmetička sredina populacije označava se kao μ , dok se aritmetička sredina uzorka označava kao \bar{x} .

Aritmetička sredina populacije je definirana kao:

$$\mu = \frac{\text{total populacije}}{\text{broj jedinica populacije}} = \frac{\sum_{i=1}^N x_i}{N}$$

gdje N predstavlja ukupan broj opažanja u uzorku.

Aritmetička sredina uzorka je definirana kao:

$$\bar{x} = \frac{\text{total uzorka}}{\text{broj jedinica uzorka}} = \frac{\sum_{i=1}^n x_i}{n}$$

gdje n predstavlja ukupan broj podataka u populaciji.

¹ R.L Plackett, „Studies in the history of probability and statistics. VII. The principle of the arithmetic mean“ u *Biometrika* Volume 45, 1958, str. 130–135

² Arithmetic Mean. In: The Concise Encyclopedia of Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-32833-1_12 (pristupljeno 30. kolovoza 2023.)

³ Hrvatska enciklopedija, mrežno izdanje, <http://www.enciklopedija.hr/Natuknica.aspx?ID=3839> (pristupljeno 28. travnja 2021.)

Ovisno o tomu, je li statistički niz grupiran u razrede, ili ne, izračunava se:

- jednostavna aritmetička sredina (za negrupirani statistički niz),
- ponderirana aritmetička sredina (za grupirani statistički niz).

Primjerice, pretpostavimo da imamo mjesečne plaće svih pet zaposlenika u poduzeću: 900, 1100, 1350, 1400 i 1500 eura. Za izračun aritmetičke sredine, prvo trebamo zbrojiti sve plaće (900 + 1100 + 1350 + 1400 + 1500), što daje ukupan zbroj od 6250 eura. Tada ovaj ukupan zbroj (6250) dijelimo s brojem zaposlenika, u ovom slučaju 5. Prosječna plaća zaposlenika (aritmetička sredina) u poduzeću iznosi 1250 eura. Matematički zapis ovog postupka je sljedeći:

$$\mu = \frac{\sum_{i=1}^5 x_i}{5} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{900 + 1100 + 1350 + 1400 + 1500}{5} = 1250$$

Aritmetička sredina ima pet svojstva koje imaju važnu ulogu za razumijevanje aritmetičke sredine i varijance. Prvo svojstvo vezano je uz prvi moment oko sredine. Momenti numeričkih nizova predstavljaju aritmetičke sredine odstupanja vrijednosti numeričke varijable od njezine aritmetičke sredine (ili neke druge vrijednosti) računane na neku potenciju, pri čemu je potencija naznačena u nazivu momenta oko sredine. Na primjer, za negrupirani statistički niz populacije, izračunava se pomoću sljedećeg izraza:

$$\mu_r = \frac{\sum_{i=1}^N (x_i - \mu)^r}{N},$$

pri čemu r može poprimiti 0 ili bilo koji prirodni broj. Ako r poprimi 0, tada se izračunava nulti moment oko sredine, ako je $r = 1$ tada se izračunava prvi moment oko sredine, za $r = 2$ izračunava se drugi moment oko sredine, itd.

Prvo svojstvo aritmetičke sredine govori da je *algebarski zbroj odstupanja originalnih vrijednosti numeričkog obilježja od aritmetičke sredine jednak nuli*.⁴

$$\text{Populacija: } \sum_{i=1}^N (x_i - \mu)^1 = 0$$

$$\text{Uzorak: } \sum_{i=1}^n (x_i - \bar{x})^1 = 0$$

⁴ J. Horvat i J. Mijoč, *Osnove statistike*, 2. izdanje, Naklada Ljevak, Zagreb, 2014.

U slučaju ranije danog primjera plaća zaposlenika, vrijedi:

$$\begin{aligned} & \sum_{i=1}^N (x_i - \mu)^1 \\ &= (900 - 1250) + (1100 - 1250) + (1350 - 1250) + (1400 - 1250) + (1500 - 1250) \\ &= -350 - 150 + 100 + 150 + 250 = 0. \end{aligned}$$

Zbog prvog svojstva aritmetičke sredine, prvi moment oko sredine je konstanta i jednak je nuli, jer neovisno kojim se brojem dijeli nula, rezultat će biti nula. Iako nulti moment oko sredine nije vezan uz njezina svojstva, vrijedi napomenuti da se također radi o konstanti. Naime, na prethodnom primjeru može se vidjeti da je

$$\begin{aligned} \mu_0 &= \sum_{i=1}^N \frac{(x_i - \mu)^0}{N} = \frac{(900-1250)^0 + (1100-1250)^0 + (1350-1250)^0 + (1400-1250)^0 + (1500-1250)^0}{5} = \\ & \frac{1+1+1+1+1}{5} = 1. \end{aligned}$$

Nulti moment oko sredine uvijek će biti jednak broju 1. Dakle, prvi moment oko sredine jednak je nuli, a nulti moment oko sredine jednak je jedan. Iako su to zanimljive pravilnosti, ne omogućuju otkrivanje dodatnih karakteristika statističkog niza. To nas dovodi do drugog svojstva aritmetičke sredine i drugog momenta oko sredine.

Drugo svojstvo aritmetičke sredine glasi da je *zbroj kvadrata odstupanja originalnih vrijednosti numeričkog obilježja od aritmetičke sredine jednak minimumu*. U slučaju ranije danog primjera plaća zaposlenika, vrijedi:

$$\begin{aligned} & \sum_{i=1}^N (x_i - \mu)^2 \\ &= (900 - 1250)^2 + (1100 - 1250)^2 + (1350 - 1250)^2 + (1400 - 1250)^2 + (1500 - 1250)^2 \\ &= (-350)^2 + (-150)^2 + 100^2 + 150^2 + 250^2 = 240000 \end{aligned}$$

Za pojedini niz, drugo svojstvo aritmetičke sredine predstavlja brojnik drugog momenta oko sredine. Odnosno, drugo svojstvo aritmetičke sredine može se promatrati kao brojnik drugog momenta oko sredine. Oba koncepta uključuju kvadrate odstupanja i fokusiraju se na mjerenje disperzije podataka.

Drugi moment oko sredine poznatiji je pod nazivom varijanca. *Varijanca* je prosječno kvadratno odstupanje vrijednosti niza od prosjeka. Oznaka za varijancu populacije je

σ^2 , dok se varijanca uzorka označava kao s^2 . Kvadrat je dio oznake varijance koji aludira na postupak kvadriranja korišten pri izračunu. Varijanca mjeri raspršenost ili disperziju podataka oko aritmetičke sredine te omogućuje uvid u to koliko su podaci raspršeni ili koncentrirani oko sredine niza.

$$\mu_2 = \sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \frac{(-350)^2 + (-150)^2 + 100^2 + 150^2 + 250^2}{5} = \frac{240000}{5} = 48000$$

Drugo svojstvo aritmetičke sredine i varijanca primjereni su za mjerenje disperzije jer uzimaju u obzir sve vrijednosti, neutraliziraju negativna odstupanja, teže većim odstupanjima te imaju matematičke i statističke osobine koje ih čine valjanom mjerom.

Ipak, većini ljudi je teško percipirati i razmišljati u kontekstu kvadrata te se u izvještavanju u praksi češće koristi standardna devijacija, koja se izračunava kao pozitivno rješenje drugog korijena varijance ($\sigma = +\sqrt{\sigma^2}$) i interpretira kao standardno odstupanje. U ovom slučaju, zaposlenici prosječno zarađuju 1250 eura mjesečno, a plaće standardno odstupaju 219.09 eura od prosjeka.

Na to ne umanjuje značaj varijance i njezinu primjenu u analizi podataka. Primjenjuje se u ekonomiji (i drugim disciplinama) pri testiranju hipoteza te kako bi se procijenili parametri regresijskih modela ili kako bi se minimizirala kvadratna odstupanja između predviđenih vrijednosti i stvarnih podataka.

Treće svojstvo glasi da se *aritmetička sredina uvijek se nalazi između najmanje i najveće vrijednosti numeričkog obilježja*. Aritmetička sredina uvijek će se nalaziti između najmanje i najveće vrijednosti varijable. To je zato što se aritmetička sredina računa kao prosječna vrijednost svih dostupnih vrijednosti i stoga mora biti između ekstremnih vrijednosti. Ovo svojstvo nam pomaže u razumijevanju raspona vrijednosti i prosječnih rezultata u ekonomskim analizama. Korisno je i pri provjeri izračunatih rezultata.

Četvrto svojstvo aritmetičke sredine glasi: *ako su sve vrijednosti numeričke varijable jednake konstanti C, aritmetička sredina te varijable jednaka je konstanti C*. Pretpostavimo da svih pet promatranih zaposlenika poduzeća ima jednaku plaću, 1000 eura. To znači da će i prosječna plaća tih zaposlenika biti 1000 eura. Dakle, u pitanju je konstanta.

Ovo svojstvo je korisno u ekonomiji jer nam omogućuje da utvrdimo aritmetičku sredinu cijena ili vrijednosti koje ostaju nepromijenjene tijekom vremena ili u različitim kontekstima. To nam omogućuje usporedbu cijena ili vrijednosti s tom referentnom konstantom i analizu promjena u odnosu na tu konstantu.

Peto svojstvo glasi da je *aritmetička sredina sklona ekstremima*. Ako imamo set podataka koji predstavlja plaće radnika u tvrtki, a jedan radnik ima vrlo visoku plaću u usporedbi s ostalim radnicima, ta iznimka će znatno utjecati na aritmetičku sredinu. Ovo svojstvo može biti korisno u ekonomiji jer može ukazivati na nepravilnosti ili nejednakosti u raspodjeli dohotka. Neka je plaća petog zaposlenika iz ranije uvedenog primjera 2800 eura. Tada će se prosjek izračunati na sljedeći način:

$$\mu = \frac{900 + 1100 + 1350 + 1400 + 2800}{5} = 1510$$

Novo-izračunata prosječna plaća zaposlenika promatranog poduzeća iznosi 1510 eura i viša je od ranije izračunatog prosjeka za 260 eura. Ipak, može se uočiti da je vrijednost novog prosjeka veća od prve četiri vrijednosti u nizu i bitno manja od najveće vrijednosti u nizu. Zbog karakteristika podataka, to jest izdvojenice u nizu, prosjek prestaje biti reprezentativni pokazatelj tog niza. Naime, jedan od ciljeva deskriptivne statistike je izračun nekolicine pokazatelja ili parametara koji dobro opisuju promatranu varijablu (kako ne bi svaki put morali iščitavati sve vrijednosti varijable), ali to ima smisla samo ako oni dobro opisuju danu varijablu. Ako ne opisuju dobro danu varijablu smatraju se nereprezentativnima. „U distribucijama s jednom ili više izdvojenica aritmetička sredina nije najbolji pokazatelj središta te se preporučuje korištenje drugih mjera središnje tendencije (npr. mod, medijan).“⁵

Aritmetička sredina i varijanca imaju važnu ulogu u ekonomiji i poslovnoj ekonomiji kao osnovne statističke mjere koje pružaju dublji uvid u različite aspekte financijskih i ekonomskih analiza. Aritmetička sredina se često koristi kako bi se razumjelo prosječno kretanje varijable, kao što su prihodi, cijene ili troškovi, što omogućuje identifikaciju tendencija i usporedbu među različitim skupinama ili vremenskim periodima. Također, aritmetička sredina se koristi za procjenu središnjeg mjesta distribucije podataka, što je ključno za donošenje informiranih odluka. S druge strane,

⁵ ibidem, str. 118.

varijanca mjeri disperziju podataka oko aritmetičke sredine i pruža uvid u koliko su podaci raspršeni ili koncentrirani. U ekonomiji, razumijevanje varijance pomaže u analizi rizika i neizvjesnosti te je ključno za donošenje odluka o ulaganjima, upravljanju portfeljem i procjeni financijskih performansi.

Aritmetička sredina i varijanca su ključni elementi u postupku testiranja hipoteza, koji uključuje različite statističke tehnike poput t-testova, z-testova, F-testova i analize varijance (ANOVA). U kontekstu hipoteza testiranja, aritmetička sredina ima glavnu ulogu jer predstavlja mjeru središnje tendencije skupa podataka. Osim toga, varijanca igra ključnu ulogu jer omogućuje procjenu varijabilnosti (raspršenosti) podataka unutar svake skupine.

Na primjer, prilikom primjene t-testa, uspoređujemo srednje vrijednosti dvaju skupova kako bismo utvrdili postoje li statistički značajne razlike između njih. Ovdje aritmetička sredina služi kao referentna vrijednost za usporedbu, dok varijanca pomaže procijeniti koliko su rezultati skloni varijacijama unutar svakog skupa.

F-test se često koristi kao dio analize varijance kako bi se utvrdila statistička značajnost razlika između varijanci više skupina. Aritmetička sredina i varijanca igraju ključnu ulogu u F-testu jer se koriste pri izračunu F-vrijednosti, koja uspoređuje varijaciju između skupina. Ovdje aritmetička sredina predstavlja središnji pokazatelj svake skupine, dok varijanca pruža informaciju o varijabilnosti podataka.

Aritmetička sredina i varijanca imaju ključnu ulogu u ANOVA modelima, važnom alatu u ekonomiji i poslovnoj ekonomiji. ANOVA omogućuje usporedbu više skupina kako bi se identificirale statistički značajne razlike među njima. U ovom kontekstu, aritmetička sredina igra ulogu temeljnog pokazatelja središnjeg položaja svake skupine, dok varijanca predstavlja raspršenost podataka unutar skupina. Analiza varijance koristi ove mjerne vrijednosti kako bi procijenila utjecaj različitih faktora na varijabilnost između skupina. Pomoću ANOVA-e možemo otkriti jesu li razlike između skupina rezultat slučajnosti ili postoje stvarne razlike koje se mogu povezati s varijabilnošću aritmetičke sredine i varijance. Kroz kvalitativnu interpretaciju rezultata ANOVA-e, ekonomisti i stručnjaci za poslovnu ekonomiju dobivaju dublji uvid u uzroke varijabilnosti između skupina te mogu usmjeriti svoje odluke i strategije na temelju tih

saznanja. Osim toga, analiza varijance omogućuje i identifikaciju interakcija između faktora, što dodatno obogaćuje razumijevanje ekonomske dinamike.

Osim aritmetičke sredine i varijance, za provedbu analize varijance potrebno je razumijevanje osnovnih elemenata inferencijalne statistike. U nastavku će se detaljnije predstaviti teorijske distribucije relevantne za analizu varijance.

3. Osnove inferencijalne statistike i teorijske distribucije koje se koriste pri analizi varijance

Deskriptivna statistika opisuje podatke, a inferencijalna statistika omogućuje predviđanje, odnosno zaključivanje iz tih podataka.⁶

Podaci se najčešće mogu dobiti samo iz uzoraka jer je inače preteško ili preskupo prikupiti podatke o cijeloj populaciji. Iako deskriptivna statistika može sažeti samo karakteristike uzorka, inferencijalna statistika koristi uzorak kako bi se zaključivalo o većoj populaciji.

Kod inferencijalne statistike važno je koristiti metode slučajnog uzorkovanja. Ako uzorak nije reprezentativan za populaciju, tada se ne mogu donijeti valjani statistički zaključci. Reprezentativni uzorak je uzorak koji je nasumično odabran, dovoljno velik i sadrži bitne karakteristike populacije. Slučajan način odabira ispitanika u uzorak osigurava da svaki član populacije ima jednaku vjerojatnost biti uključen u uzorak. Ovaj pristup osigurava objektivnost i smanjuje pristranost u odabiru uzorka. Pristranost u ovom kontekstu označava tendenciju ili iskrivljenost u odabiru uzorka koja može dovesti do nepravilnog ili netočnog predstavljanja populacije. Na primjer, pretpostavimo da se istražuje zadovoljstvo kupaca u određenom trgovačkom centru. Ako se odabere uzorak samo među kupcima koji se često pojavljuju u centru, preskačući povremene posjetitelje ili one koji dolaze tijekom vikenda, uzorak će biti pristran prema redovitim posjetiteljima i neće dobro predstavljati sve kupce. To bi moglo dovesti do iskrivljenih rezultata i netočnih zaključaka o zadovoljstvu svih kupaca u tom centru. Stoga, nasumični odabir uzorka pomaže izbjeći ovu pristranost i osigurava da svi tipovi kupaca budu zastupljeni u uzorku.

Veličina uzorka odnosi se na broj opažanja ili podataka koji su uključeni u uzorak uzet iz veće populacije. Veličina uzorka ključan je faktor u statističkoj analizi jer izravno utječe na pouzdanost i valjanost zaključaka izvedenih iz analize. Veći uzorak obično pruža preciznije rezultate jer pomaže smanjiti učinak slučajnih varijacija ili fluktuacija

⁶ Uvod u deskriptivnu i inferencijalnu statistiku, <https://www.unizd.hr/Portals/12/Nastavnici/Irena%20Burić/Uvod%20u%20deskriptivnu%20i%20inferencijalnu%20statistiku.pdf>, (pristupljeno 30. travnja 2021.)

koje se mogu javiti u manjim uzorcima. Ova ideja često se izražava pojmom "zakon velikih brojeva," koji sugerira da kako veličina uzorka raste, prosjek uzorka približava se prosjeku populacije. Međutim, veća veličina uzorka također može biti zahtjevnija u smislu vremena, truda i troškova potrebnih za prikupljanje i analizu podataka. Stoga određivanje prikladne veličine uzorka uključuje ravnotežu između postizanja statističke preciznosti i praktične izvedivosti. Istraživači često koriste statističke tehnike i alate (od kojih su mnogi dostupni online⁷) kako bi izračunali potrebnu veličinu uzorka.

Bitne karakteristike populacije su ključni atributi, svojstva ili varijable koje su važne za istraživanje i koje treba zastupiti u uzorku kako bi se osigurala valjana i relevantna analiza. Odabir uzorka koji sadrži bitne karakteristike populacije omogućuje da rezultati analize budu primjenjivi na cijelu populaciju. Na primjer, ako se istražuju potrošačke navike u određenom gradu, bitne karakteristike populacije u ovom slučaju mogle bi uključivati dob, spol, prihod, obrazovanje i tipične kupovine potrošača. Ako se uzorak sastoji samo od mladih odraslih osoba sa visokim prihodima, analiza ne bi bila reprezentativna za cijelu populaciju koja uključuje ljude različite dobi, prihoda i obrazovanja. Stoga bi reprezentativan uzorak trebao obuhvatiti raznoliku skupinu ljudi sa svim bitnim karakteristikama kako bi istraživanje dalo valjane rezultate koji se mogu primijeniti na sve potrošače u gradu.

„Ako se iz osnovnog skupa veličine N izaberu svi mogući uzorci veličine n , te se za svaki uzorak izračuna neki odgovarajući parametar, distribucija tih parametara naziva se sampling distribucija.“⁸ Veličina standardne pogreške aritmetičke sredine ovisi o veličini uzorka, odnosno što je veći n uzorka, standardna pogreška aritmetičke sredine je manja.

Inferencijalna statistika predstavlja skup metoda i modela koje omogućuju donošenje zaključaka o populaciji na temelju izabranog uzorka. Pritom se koriste procjene, testiranja hipoteza, određivanja veza između varijabla i predviđanja o populaciji. Dva

⁷ <https://www.calculator.net/sample-size-calculator.html>

⁸ M. Biljan-August, S. Pivac i A. Štambuk, *Statistička analiza u ekonomiji*, Ekonomski fakultet Sveučilišta u Rijeci, Rijeka, 2009., str. 43

su glavna područja inferencijalne statistike: područje procjena parametara i područje testiranja hipoteza.⁹

Inferencijalna statistika se može svesti na temelju jednadžbe da je ishod = model + pogreška. Ishod predstavlja ono što smo opazili, podatke koje smo dobili. Model je statistički model koji smo pretpostavili i usporedili s dobivenim podacima, a pogreška je posljedica činjenice da smo podatke prikupili na uzorku.¹⁰

Postupak statističke *procjene parametara* je metodologija koja se koristi kako bi se iz uzorka podataka donijele informirane zaključke o parametrima populacije, kao što su prosjek, varijanca ili drugi statistički pokazatelji. Ovaj postupak omogućuje donošenje estimacija ili procjena vrijednosti parametara temeljem uzorka, uz pridodavanje odgovarajuće razine povjerenja.

Razina povjerenja je statistički koncept koji odražava koliko smo sigurni u točnost procjene parametra populacije temeljem podataka iz uzorka. Izražava se kao postotak i označava koliko puta od 100 sličnih uzoraka možemo očekivati da će pokazatelj uzorka biti unutar određenog intervala. Razina povjerenja je često izražena kao postotak, kao što su 90%, 95% ili 99%. Što je viša razina povjerenja, to su šire procijenjeni intervali povjerenja. Na primjer, interval povjerenja od 99% bit će u pravilu širi od intervala povjerenja od 95%, jer veća razina povjerenja zahtijeva veću sigurnost u preciznost procjene. Razina povjerenja nam pomaže razumjeti koliko su pouzdane naše procjene temeljem uzorka te kako varijabilnost uzorka utječe na preciznost procjena. Na primjer, ako imamo procijenjeni interval povjerenja za prosječnu visinu populacije od 165 cm do 175 cm na razini povjerenja od 95%, to znači da u 95% uzoraka iz iste populacije očekujemo da će prosječna visina biti unutar tog intervala.

Testiranje hipoteza je statistički postupak koji se koristi kako bi se donosili zaključci o populaciji na temelju podataka iz uzorka. Ovaj postupak uključuje postavljanje nul i alternativne hipoteze, provođenje statističkog testa i donošenje zaključka o nul hipotezi. Nul hipoteza (H_0) je tvrdnja koja se ispituje i obično pretpostavlja da nema

⁹ Hrvatska enciklopedija, mrežno izdanje <https://www.enciklopedija.hr/natuknica.aspx?ID=57896>, (pristupljeno 30. travnja 2021.)

¹⁰ Uvod u deskriptivnu i inferencijalnu statistiku, <https://www.unizd.hr/Portals/12/Nastavnici/Irena%20Buric/Uvod%20u%20deskriptivnu%20i%20inferencijalnu%20statistiku.pdf>, (pristupljeno 30. travnja 2021.)

značajne razlike. Alternativna hipoteza (H_1) sugerira suprotno, tj. da postoji značajna razlika. Statistički test koristi pokazatelje uzorka kako bi se procijenila vjerojatnost odbacivanja nul hipoteze, temeljem odabrane razine značajnosti i izračuna p-vrijednosti.

Razina značajnosti je prag koji se postavlja unaprijed pri testiranju hipoteza kako bi se donosili zaključci o nul hipotezi. Ova razina određuje koliko ekstremni rezultati moraju biti da bismo mogli odbaciti nul hipotezu. Uobičajeno je postaviti razinu značajnosti na vrijednosti poput 0.05 ili 0.01. Ako je p-vrijednost (vjerojatnost dobivanja rezultata jednako ekstremnih ili ekstremnijih od onih koje smo dobili) manja od postavljene razine značajnosti, tada odbacujemo nul hipotezu. Razina značajnosti igra važnu ulogu u testiranju hipoteza jer pomaže kontrolirati vjerojatnost donošenja netočnih zaključaka. Postavljanje niže razine značajnosti znači da smo zahtjevniji u odbacivanju nul hipoteze, dok postavljanje više razine značajnosti povećava vjerojatnost odbacivanja nul hipoteze.

P-vrijednost je statistička mjerljiva vrijednost koja se koristi u testiranju hipoteza. Ona označava vjerojatnost dobivanja rezultata jednako ekstremnih ili ekstremnijih od onih koje smo dobili, ako je nul hipoteza (H_0) istinita. Na primjer, ako je p-vrijednost 0.02, to znači da, ako je nul hipoteza točna, postoji (samo) 2% vjerojatnosti da bismo dobili tako ekstremne rezultate kao što smo dobili u uzorku. Manja p-vrijednost ukazuje na veću vjerojatnost da se nul hipoteza odbaci, što sugerira veću podršku alternativnoj hipotezi (H_1). Vezana je uz pogrešku tipa I (vjerojatnost odbacivanja točne nul hipoteze), a utvrđuje se temeljem testne veličine. P-vrijednost je površina ispod oba (ili u posebnim slučajevima jednog) kraka distribucije vjerojatnosti od izračunate vrijednosti statističkog testa. Može se interpretirati kao empirijska vjerojatnost pogreške tipa I ili vjerojatnost da će se učiniti pogreška ako se odbaci nul-hipoteza. Odluka o odbacivanju ili neodbacivanju nul hipoteze temelji se na usporedbi p-vrijednosti s unaprijed postavljenom razinom značajnosti (najčešće 0.05 ili 0.01). Ako je p-vrijednost manja od te razine, odbacuje se nul hipoteza, što znači da postoji statistički značajna razlika. Ako je p-vrijednost veća od razine značajnosti, ne možemo odbaciti nul hipotezu, ali to ne znači nužno da je ona istinita. Na primjer, ako je razina značajnosti $\alpha = 0,05$ i ako je $p < 0.05$ odbacuje se nul-hipoteza, tj. rezultati su statistički značajni na 5% razini značajnosti, a ako je $p \geq 0.05$ ne odbacuje se nul-

hipoteza; tj. rezultati nisu statistički značajni na 5% razine značajnosti, odnosno nema dovoljno dokaza za odbacivanje nul-hipoteze.

Izbor veličine razine značajnosti statističkog testa (0.1; 0.05; 0.01; 0.001) je proizvoljan. Veličina razine značajnosti govori o tome u kojem postotku si istraživač dopušta načiniti grešku odbacivanja istinite nul-hipoteze. Metode inferencijalne statistike polaze iz uzorka realne i konačne populacije na čemu se baziraju njihovi rezultati i zaključci. Pritom je moguće učiniti pogreške.

U zaključivanju temeljem statističkih testova moguće su pogreške:

1. Pogreška 1. vrste – nastaje u slučaju odbacivanja istinite nul hipoteze.
2. Pogreška 2. vrste – nastaje pri neodbacivanju neistinite nul hipoteze.

Postoje parametrijski i neparametrijski statistički testove. Parametrijski statistički testovi su testovi koji, da bi se smjeli primijeniti u statističkom testiranju, postavljaju zahtjeve vezane uz distribuciju izvornih podataka. Neparametrijski testovi ne postavljaju takav tip zahtjeva, premda i oni pretpostavljaju određene uvjete pod kojima ih se smije primijeniti. Neparametrijske testove često zovu još i testovima nezavisnim o distribucijama podataka (*engl. distribution free tests*). U nastavku će se govoriti o parametrijskim testovima, kao što su Studentov ili t-test, F-test, ANOVA, koji su vezani uz specifične teorijske distribucije.

„Teorijske distribucije vjerojatnosti opisane su matematičkim modelima. To su takve distribucije koje se mogu očekivati u skladu s našim ili na temelju nekih teorijski pretpostavki“¹¹

Raspodjela ili distribucija vjerojatnosti statistička je funkcija koja se koristi za prikaz svih mogućih vrijednosti i vjerojatnosti slučajne varijable u određenom rasponu. Raspon bi bio ograničen maksimalnom i minimalnom vrijednošću, ali stvarna vrijednost ovisila bi o brojnim čimbenicima. Postoje diskretne teorijske distribucije i kontinuirane teorijske distribucije. Obje raspodjele odnose se na raspodjele vjerojatnosti, koje su temelj statističke analize i teorije vjerojatnosti.

¹¹ Analiza poslovnih podataka, http://www.efos.unios.hr/analiza-poslovnih-podataka/wp-content/uploads/sites/160/2013/04/4_teorijske-distribucije.pdf, (pristupljeno 30. travnja 2021.)

Diskretna raspodjela je raspodjela podataka koja ima diskretne vrijednosti. Diskretne vrijednosti su brojive, konačne, pozitivne cjelobrojne vrijednosti poput 1, 5, 15, ... „Najčešće korištene teorijske distribucije diskretne slučajne varijable su binomna i Poissonova distribucija.“¹²

Kontinuirana raspodjela opisuje vjerojatnosti mogućih vrijednosti kontinuirane slučajne varijable. Kontinuirana slučajna varijabla je slučajna varijabla sa skupom mogućih vrijednosti (poznata kao raspon) koja je beskonačna i nebrojiva. Kontinuirana slučajna varijabla X može poprimiti beskonačno mnogo vrijednosti.¹³ Stoga se ne računa vjerojatnost u određenoj točki, već vjerojatnost za određeni interval vrijednosti varijable X . Najpoznatije teorijske distribucije kontinuirane slučajne varijable su normalna distribucija, studentova distribucija, hi-kvadrat distribucija i f-distribucija.

3.1. Normalna distribucija

Normalna distribucija, poznata još kao i Gaussova distribucija je najvažnija i najčešća distribucija. Najlakši način prikazivanja normalne distribucije je pomoću funkcije gustoće vjerojatnosti.

Normalna distribucija je uobičajena distribucija vjerojatnosti koja je simetrična i oblika zvona. U ovoj distribuciji podaci se koncentriraju oko srednjih vrijednosti, a što se više udaljavamo od srednjih vrijednosti, manje podataka nalazimo. Aritmetička sredina, medijan i mod normalne distribucije su jednaki, a distribuciju u potpunosti karakteriziraju njezin prosjek i standardna devijacija. Standardna devijacija kontrolira oblik distribucije. Manja standardna devijacija će ukazivati na to da su podaci usko grupirani oko srednje vrijednosti i krivulja će biti šiljastija, dok veća standardna devijacija ukazuje na to da su podaci raspršeniji i da će krivulja biti šira i plosnatija.

¹² Analiza poslovnih podataka, http://www.efos.unios.hr/analiza-poslovnih-podataka/wp-content/uploads/sites/160/2013/04/4_teorijske-distribucije.pdf, (pristupljeno 30. travnja 2021.)

¹³ Kontinuirane slučajne varijable, <https://www.pmf.unizg.hr/download/repository/PREDAVANJE9.pdf>, (pristupljeno 30. travnja 2021.)

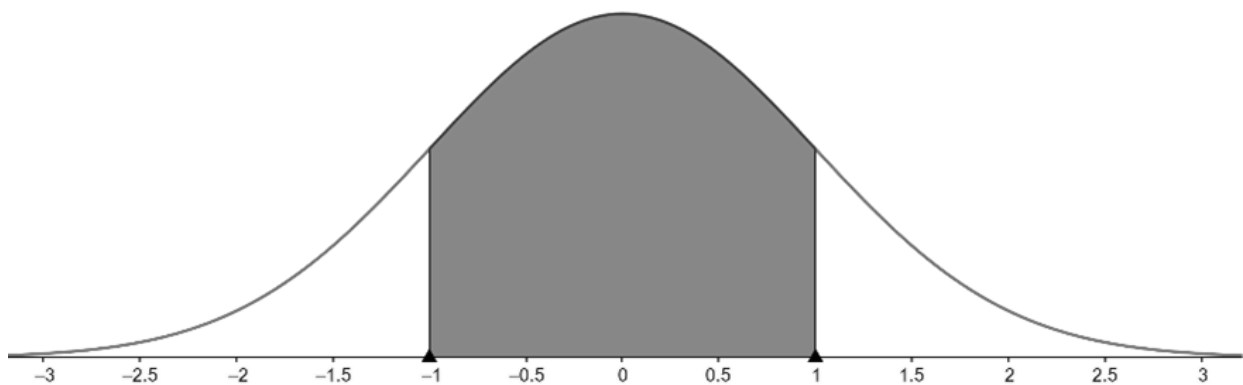
Mnoge prirodne pojave i mjerenja u područjima poput ekonomije, biologije i društvenih znanosti slijede normalnu distribuciju. To je temeljni koncept u statistici i često se koristi u različitim analitičkim i modelirajućim kontekstima.

Na primjer, zamislite veliku grupu studenata koji su polagali ispit iz ekonomije. Ako je ispit dobro osmišljen i obuhvaća široki raspon tema, distribucija rezultata vjerojatno će slijediti normalnu distribuciju. Većina studenata može postići rezultate oko prosjeka, dok manji broj studenata postiže znatno više ili niže ocjene.

Za drugi primjer, neka se promatraju dnevne prodaje jedne maloprodajne trgovine tijekom duljeg vremenskog razdoblja. Ako su prodaje rezultat različitih faktora kao što su ponašanje kupaca, ekonomske prilike i marketinški naponi, podaci o dnevnim prodajama mogli bi pokazivati normalnu distribuciju. Većinu dana, prodaja bi se mogla grupirati oko prosječne dnevne prodaje, dok se iznimno visoke ili niske prodaje događaju manje često.

Ako se vrijednosti slučajne varijable x izraze u jedinicama standardne devijacije, kreira se standardizirana normalna distribucija. Standardizirana normalna distribucija ima aritmetičku sredinu 0 i standardnu devijaciju 1, a zadaje se kao $N(0,1)$. Bilo koja normalna distribucija se može pretvoriti u standardnu putem zamjene $z = (x - \mu)/\sigma$. Ovaj postupak se zove standardizacija.

Slika 1. Standardizirana normalna distribucija



Izvor: samostalna izrada koristeći <https://www.geogebra.org/classic#probability>

Ukoliko se želi doznati proporcija mogućih odstupanja primjenjivat će se empirijsko pravilo koja glasi da ako aritmetičkoj sredini dodamo lijevo i desno po jednu standardnu devijaciju, obuhvaća se površina koja čini 68,26% cijele površine krivulje (sivo označena površina na Slici 1.). Taj broj smo dobili utvrđivanjem površine pod krivuljom u intervalu dviju standardnih devijacija. Dok 95,44% svih podataka obuhvaćeno je s dvije standardne devijacije oko aritmetičke sredine, a s tri standardne devijacije gotovo svi podaci tj. 99,73% rezultata.

Tablica 1. Usporedba karakteristika normalne i standardizirane normalne distribucije

Karakteristike normalne distribucije $N(\mu, \sigma)$	Karakteristike standardizirane normalne distribucije $Z(0, 1)$
Očekivanje, mod i medijan su jednaki	
Krivulja je simetrična oko očekivanja	
Površina ispod krivulje sa svake strane očekivanja iznosi 0,5	
Ukupna površina ispod krivulje iznosi 1	
Krakovi krivulje protežu se u beskonačnost	
Najveća točka na normalnoj krivulji nalazi se kod očekivanja	
Vrijede Empirijsko pravilo i pravilo Čebiševa	
Svaki par vrijednosti za očekivanje i standardnu devijaciju određuje drugačiju normalnu krivulju	Aritmetička sredina iznosi 0, a standardna devijacija iznosi 1, $\alpha_3 = 0, \alpha_4 = 3$
Očekivanje se nalazi u centru krivulje, a to može biti bilo koji realan broj: pozitivan, negativan ili nula	Očekivanje (aritmetička sredina) se nalazi u centru krivulje i iznosi 0
Standardna devijacija je pozitivna i ona određuje oblik normalne krivulje; što je standardna devijacija veća, to je krivulja šira i plosnatija	Standardna devijacija iznosi 1, $\alpha_4 = 3$

Izvor: e-učenje, predmet: Statistika u Ekonomiji (Katarina Kostelić)

Normalna distribucija i standardizirana normalna distribucija su ključni alati pri testiranju hipoteza u statistici. Normalna distribucija je često prisutna u stvarnim podacima i omogućava nam da razumijemo kako se vrijednosti raspoređuju oko srednje vrijednosti. Standardizirana normalna distribucija je poseban oblik normalne distribucije s prosječnom vrijednošću od 0 i standardnom devijacijom od 1, što olakšava usporedbu i izračune. Prilikom testiranja hipoteza u kojima su zadovoljeni uvjeti dovoljno velike populacije ($n > 30$), poznata standardna devijacija populacije te normalna distribucija podataka, koristi se z-test. Pritom se koristi standardizirana

normalna distribucija za izračuni p-vrijednosti. Na primjer, u ekonomiji, možemo koristiti normalnu distribuciju za testiranje hipoteza o jednakosti prosječnih plaća u dvije različite tvrtke. Ako imamo podatke o plaćama zaposlenika u obje tvrtke koji su normalno distribuirani i želimo testirati je li razlika u prosječnim plaćama statistički značajna, koristimo standardiziranu normalnu distribuciju kako bismo izračunali p-vrijednost i donijeli zaključak o nul hipotezi.

No, ako uzorak nije dovoljno veliki ili standardna devijacija populacije nije poznata, potrebno je upotrijebiti neki drugi test. Ako se radi o usporedbi parametara dviju populacija, najčešće korištena alternativa je t-test, koji se oslanja na t-distribuciju.

3.2. Studentova t - distribucija

William Gosset definirao je t-distribuciju i objavio je u časopisu *Biometrika*, 1908. godine pod pseudonimom Student.¹⁴ Zanimljivost u vezi s objavljivanjem t-distribucije jest da je tijekom tog vremena radio u pivovari Guinness u Dublinu te je objavio svoj rad pod pseudonimom "Student" zbog ograničenja tvrtke vezanih uz objavljivanje. Kasnije je ova distribucija postala poznata kao t-distribucija ili Studentova distribucija, koja je ključni metodološki odabir kada je uzorak manjeg obujma. Ova zanimljivost pokazuje praktičnu primjenu statističkih koncepta izvan akademskog okvira te koliko je Studentova distribucija postala važna u analizi podataka.

Obilježja t-distribucije:

- određena je stupnjevima slobode
- ima sličan oblik kao normalna distribucija samo što je šira i spljoštenija
- kako raste broj stupnjeva slobode oblikom je sve sličnija normalnoj distribuciji

¹⁴ D. Dizdar, *Kvantitativne metode*, Kineziološki fakultet Sveučilišta u Zagrebu, Zagreb, 2006., str. 107.

- primjenjuje se u računanju intervala pouzdanosti i testiranju hipoteza o razlici između dva uzorka:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

Distribucija t koristi se umjesto normalne raspodjele kada imamo male uzorke, odnosno kada je $n < 30$ ili nije poznata standardna devijacija populacije.

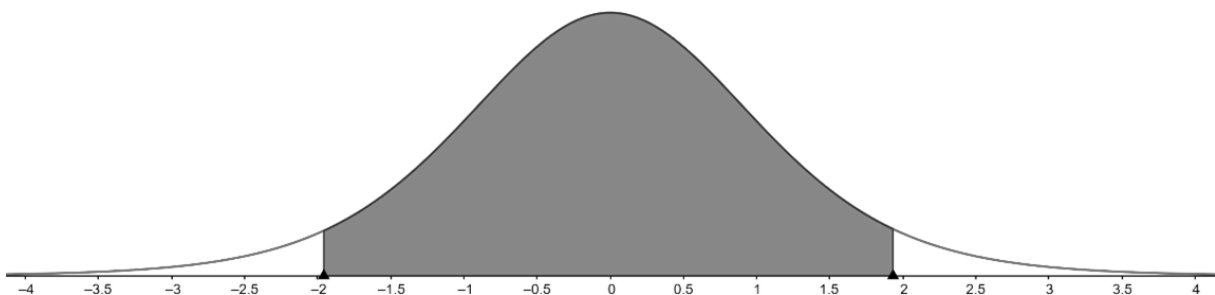
Stupnjevi slobode predstavljaju broj neovisnih opažanja u jednom uzorku i označavaju se oznakom df . Njima su zapisane vrijednosti koje slobodno variraju od uzorka do uzorka, a način njihova određivanja ovisi o statističkoj distribuciji. Za t distribuciju stupnjevi slobode se računaju :

$$df = n - 1$$

Studentova t -distribucija ima oblik sličan normalnoj distribuciji. Za $df \rightarrow \infty$, t -distribucija se oblikom približava standardiziranoj normalnoj distribuciji s parametrima

$$\mu = 0 \text{ i } \sigma = 1.^{15}$$

Slika 2. Studentova t – distribucija



Izvor: samostalna izrada koristeći <https://www.geogebra.org/classic#probability>

Kako se n povećava, Student- t raspodjela teži normalnoj $(0, 1)$ raspodjeli. Do trenutka kada je n iznosi više od 30, dvije su distribucije gotovo identične. Što je veća veličina uzorka, t - distribucija više izgleda kao normalna distribucija. Zapravo, za veličine uzoraka veće od 20, odnosno više stupnjeva slobode raspodjela je skoro slična normalnoj distribuciji.

¹⁵ ibidem, str. 109.

T-distribucija igra važnu ulogu u testiranju hipoteza, posebno kada radimo s malim uzorcima podataka. Kada imamo mali uzorak, t-distribucija nam omogućava da preciznije procijenimo raspršenost podataka i donesemo valjane zaključke o populacijskim parametrima. Primjena t-distribucije često se vidi u ekonomiji, posebno kada analiziramo učinke novih strategija na uzorku ograničene veličine. Na primjer, zamislimo istraživanje u kojem se ispituje postoji li razlika u potrošnji domaćinstava uslijed povećanja kamatnih stopa. Ako imamo samo nekoliko domaćinstava u uzorku, koristit ćemo t-distribuciju kako bismo izračunali p-vrijednost za statistički test i donijeli zaključke o statističkoj značajnosti učinka kamatnih stopa na potrošnju. T-distribucija omogućava preciznost i pouzdanost u statističkim zaključcima čak i s malim uzorkom.

3.3. Snedecorova f – distribucija

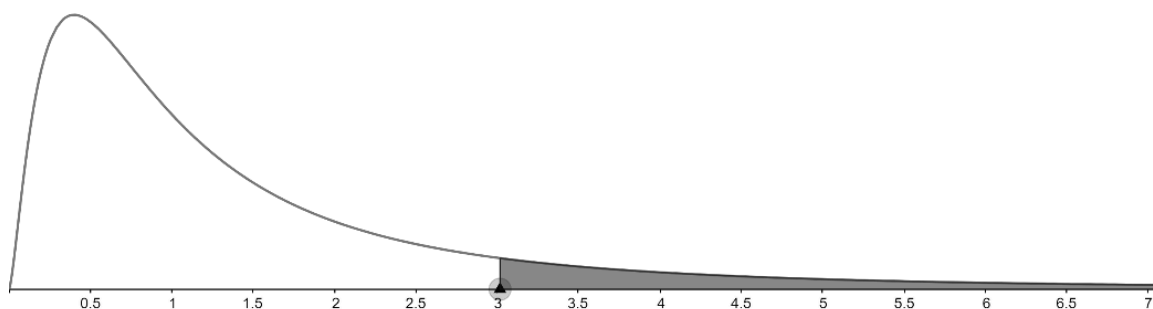
George W. Snedecor je na temelju ranijih radova R. A. Fishera, a za potrebe suvremene statističke prakse, definirao F-distribuciju. U Fisherovu čast označio ju je simbolom F. Snedecorova ili F - distribucija je kontinuirana funkcija vjerojatnosti slučajne varijable definirana u intervalu $(0, +\infty)$ ¹⁶. Slučajna varijabla F-distribucije definirana je za pozitivne stvarne vrijednosti i proizlazi iz omjera vrijednosti Hi-kvadrat distribucije s pripadajućim stupnjevima slobode. „F - distribucija je teoretska raspodjela vjerojatnosti koju karakteriziraju dva parametra, df_1 i df_2 , a oba utječu na oblik raspodjele. Budući da F-omjer uvijek mora biti pozitivan, F - distribucija je nesimetrična, nagnuta u pozitivnom smjeru.“¹⁷

F - distribucija je distribucija kvocijenata dviju χ^2 slučajnih neovisnih varijabli, pri čemu je u definiciji svaka χ^2 podijeljena sa svojim stupnjem slobode. Primjenom Cochranovog teorema, F-distribucija se može izraziti kao $X = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$. Često se koristi za testiranje jednakosti varijanci između dvije nezavisne normalno distribuirane grupe u analizi varijance, uz preduvjet da su uzorci iz normalne distribucije.

¹⁶ ibidem, str. 10.

¹⁷ Analiza varijance, <https://dwstockburger.com/Introbook/sbk21.htm>, (pristupljeno 9. svibnja 2021.)

Slika 3. F – distribucija



Izvor: samostalna izrada koristeći <https://www.geogebra.org/classic#probability>

F-distribucija je primjer omjerne distribucije, što znači da se temelji na omjerima slučajnih varijabli koje slijede druge dvije distribucije. Oblik teorijske distribucije određen je stupnjevima slobode oba uzorka: df_1 i df_2 . Oblik distribucije varira ovisno o veličini uzoraka u svakoj skupini (za razliku od Hi-kvadrat distribucije koja u obzir uzima samo jedan skup). F-distribucija sadrži pozitivne vrijednosti i pozitivno je asimetrična za manje stupnjeva slobode. Kako se stupnjevi slobode povećavaju, distribucija sve više sličí normalnoj distribuciji.

F-distribucija ima važnu ulogu u testiranju hipoteza, posebno kad je potrebno usporediti varijance između različitih grupa podataka. Ova distribucija omogućuje nam da provjerimo je li varijabilnost između grupa statistički značajno različita ili je ta razlika rezultat slučajnosti. Na primjer, neka se istražuje učinak različitih obrazovnih metoda na postignuće učenika u ekonomskom obrazovanju, tradicionalni klasični pristup i moderni interaktivni pristup. Svaka metoda obrazovanja predstavlja jedan faktor, prema kojoj se učenici, tj. njihovo postignuće dijele u dvije skupine. Nakon prikupljanja podataka i provedbe analize varijance, dobiva se F-vrijednost koja će pomoći utvrditi je li razlika u varijaciji postignuća učenika s obzirom na metodu statistički značajna ili se može pripisati slučajnosti. Ako F-vrijednost prelazi određeni prag i p-vrijednost je niža od unaprijed postavljene razine značajnosti, može se zaključiti da postoje statistički značajne razlike u postignućima između obrazovnih metoda. F-distribucija tako omogućuje donošenje informiranih odluka temeljenih na analizi varijabilnosti u ekonomskim istraživanjima.

4. T – test

Za malu veličinu uzorka ($n < 30$) odabranog iz približno normalno distribuirane populacije uz nepoznatu veličinu standardne devijacije populacije (σ) primjenjuje se statistički t-test, odnosno normalna distribucija zamjenjuje se sa studentovom, t-distribucijom.¹⁸

U Studentovom testu računa se tzv. t-parametar, t-vrijednost ili t-statistika, a to se izračunava kao omjer dvaju brojeva – u brojniku je odstupanje vrijednosti prosjeka uzorka od pretpostavljenog prosjeka populacije, a u nazivniku je standardna pogreška tog odstupanja. Što je razlika između prosječne vrijednosti uzorka i pretpostavljene prosječne vrijednosti populacije veća u odnosu na njenu standardnu pogrešku, to će t-parametar biti veći. Veća t-vrijednost ukazuje na veću statističku značajnost razlike između uzorka i pretpostavljene vrijednosti populacije, što implicira manju vjerojatnost da se takva razlika dogodila slučajno.

$$t = \frac{\text{odstupanje vrijednosti pokazatelja od referentne vrijednosti}}{\text{standardna pogreška pokazatelja}}$$

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

Postoje tri glavne vrste t-testova: jednosmjerni t-test, dvostrani t-test i uparen t-test. Jednosmjerni t-test koristi se kada želimo usporediti prosjek s poznatim prosjekom populacije ili s nekom prethodno definiranom vrijednošću (μ_0). U tom slučaju hipoteze glase: $H_0 \dots \mu \geq \mu_0$

$$H_1 \dots \mu < \mu_0$$

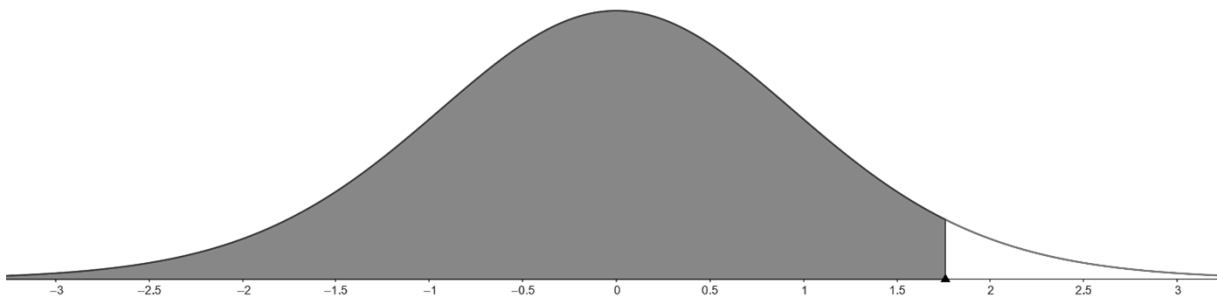
ili

$$H_0 \dots \mu \leq \mu_0$$

$$H_1 \dots \mu > \mu_0$$

¹⁸ J. Horvat i J. Mijoč, *Osnove statistike*, 2, izdanje, Naklada Ljevak, Zagreb, 2014, str. 380

Slika 4. Graf t- distribucije za df=15



Izvor: samostalna izrada koristeći <https://www.geogebra.org/classic#probability>

Slikom 4. prikazana je t-distribucija ($df = 15$), na kojoj je sivom bojom obojena površina koja predstavlja 95% podataka, a bijela površina pod krivuljom preostalih 5 % i odgovara drugom setu zapisanih hipoteza. Bijela površina vezana je uz razinu značajnosti i odbacivanje nul hipoteze, dok siva površina određuje interval neodbacivanja nul hipoteze. Ako se izračunata testna veličina nalazi u sivom području, nul hipoteza se ne odbacuje, a ako se nalazi u bijelom području, nul hipoteza se odbacuje. Ovaj test pomaže u određivanju postoji li značajna razlika između prosjeka uzorka i referentne vrijednosti.

Dvosmjerni t-test koristi se kako bi se utvrdilo postoji li značajna razlika između prosjeka dva uzoraka (μ_1, μ_2) ili između prosjeka uzorka i pretpostavljene vrijednosti (μ_0). Ovaj test omogućuje otkrivanje bilo kakvih statistički značajnih razlika, bez obzira na to koji je prosjek veći. Hipoteze koje se testiraju na ovaj način, općenito glase

$$H_0 \dots \mu_1 = \mu_2$$

$$H_1 \dots \mu_1 \neq \mu_2$$

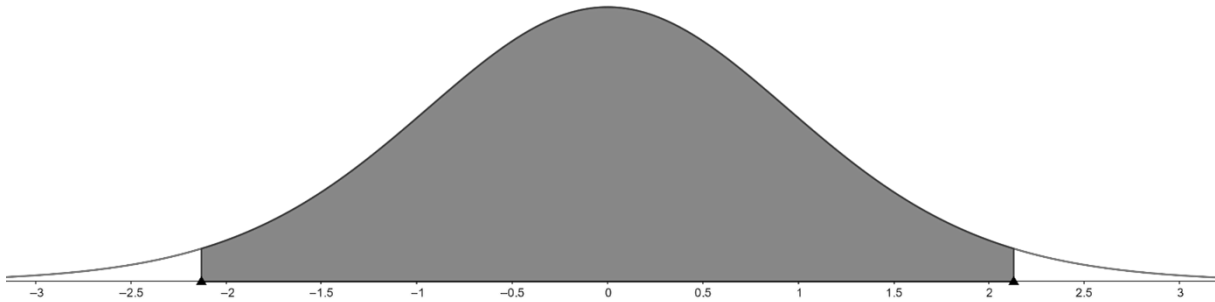
odnosno, ako je u pitanju pretpostavljena vrijednost:

$$H_0 \dots \mu = \mu_0$$

$$H_1 \dots \mu \neq \mu_0.$$

Prvi set hipoteza vezuje se i za test koji se naziva i t - test za nezavisne uzorke. Taj se test koristi kada želimo dvije različite skupine ispitanika (ili uzoraka) usporediti u nekoj varijabli koja je kvantitativna, kontinuirana. ¹⁹.

Slika 5. Dvosmjerna t - distribucija



Izvor: samostalna izrada koristeći <https://www.geogebra.org/classic#probability>

U ovom slučaju, vjerojatnost zadana razinom značajnosti dijeli se na oba kraka distribucije. Tako je uz bijelu površinu pod krivuljom vezan interval odbacivanja nul hipoteza te ako se izračunata testna veličina nađe u tom intervalu odbacuje se nul hipoteza (Slika 5). Štoviše, ako je izračunata testna veličina koja se nalazi u području odbacivanja, p-vrijednost će biti manja od razine značajnosti. Ako je testna veličina u intervalu vezanom uz sivu površinu, nul hipoteza se ne odbacuje.

Upareni t-test, odnosno t-test za uparene podatke koristi se za usporedbu srednjih vrijednosti istih subjekata ili jedinica u dva različita vremenska ili uvjetna okvira. U ovom kontekstu, često se spominje izraz zavisni ili ovisni uzorci. Izraz zavisni uzorci trebao bi ukazivati na činjenicu da rezultati drugog mjerenja na neki način ovise o rezultatima prvog mjerenja. Podaci prikupljeni na način gdje su rezultati drugog mjerenja ovisni o rezultatima prvog mjerenja zahtijevaju primjenu statističkih metoda koje uzimaju u obzir tu međusobnu ovisnost. U tu svrhu, razvijen je t-test za uparene podatke kao odgovarajuća metoda analize. T-test za uparene podatke uzima u obzir povezanost između rezultata iz dva povezana mjerenja i koristi se za usporedbu prosječnih vrijednosti između tih mjerenja. Pri tome je važno osigurati jednak broj uzoraka u oba uvjeta kako bi t-test za uparene podatke bio primjenjiv. U slučaju narušavanja ovog

¹⁹ Obrada i prikaz kvantitativnih varijabli, https://bookdown.org/gkardum1/r_znanost_obrazovanje/obrada-i-prikaz-quantitativnih-varijabli.html (pristupljeno 21. svibnja 2021.)

preduvjeta, moguća alternativna metoda je analiza varijance za ponovljena mjerenja. Hipoteze se zadaju na isti način kao pri dvosmjernom testu usporedbe prosjeka dvaju uzoraka i odluka o hipotezi donosi se na isti način.

4.1. Pretpostavke t-testa

Kad odlučimo analizirati podatke pomoću t testa, dio procesa označava provjeru mogu li se podaci koje želimo analizirati doista analizirati pomoću t testa. Ako naši podaci zadovoljavaju četiri pretpostavke potrebne za izradu jednog uzorka onda je t - test dao valjan rezultat.

1. Prva pretpostavka u vezi s t-testova se odnosi na mjernu ljestvicu. Varijabla bi se trebala sastojati od podataka izmjerenih na intervalnoj ili omjernoj mjernoj ljestvici. Neto dobit tvrtke može biti izmjerena u novčanim iznosima, što odgovara omjernoj ljestvici. Na omjernoj ljestvici, razlike između opaženih vrijednosti neto dobiti su kvantitativne i postoji apsolutna nula.

2. Druga pretpostavka je da su opažanja međusobno nezavisna, što znači da podaci proizlaze iz različitih izvora. Podaci se prikupljaju od reprezentativnog, nasumičnog odabranog dijela ukupne populacije.

3. Treća pretpostavka je da varijabla treba biti približno normalno distribuirana, kako bi podaci rezultirali normalnoj raspodjeli, odnosno krivulji u obliku zvona. Tada se može odrediti razina vjerojatnosti (alfa, razina značajnosti, p) kao kriterij za odbacivanje. U većini slučajeva odabire se vjerojatnost od 5%.

4. Konačna pretpostavka je da varijabla ne smije sadržavati izdvojenice. Izdvojenice su neobično velike ili male vrijednosti u odnosu na ostatak podataka i mogu biti rezultat pogrešaka u mjerenju, rijetki događaji ili stvarne ekstremne vrijednosti.

Ako jedna ili više pretpostavki t-testa nisu zadovoljene, nije preporučljivo provoditi t-test, jer bi to moglo dovesti do netočnih rezultata i pogrešnih zaključaka. Međutim, kada se krše pretpostavke t-testa, postoje alternativni statistički testovi koji su manje osjetljivi na te pretpostavke, a nazivaju se neparametrijski testovi.

4.2. Primjer provedbe t-testa

U hipotetskom primjeru za potrebe prikaza provedbe t-testa, istražuju se razlike u plaćama zaposlenika unutar jednog poduzeća koje ima zaposlenike u Zagrebu i Osijeku. Podaci su prikupljeni za dvije nezavisne grupe zaposlenika, svaka grupa predstavlja jedan grad. Nakon analize podataka, utvrđeno je da se radi o dvije neovisne grupe, budući da su plaće izmjerene na različitim skupinama zaposlenika u različitim gradovima. Stoga će neupareni t-test biti primijenjen za usporedbu prosječnih plaća između ove dvije grupe. Kako je cilj istraživanja utvrditi ima li razlike u plaćama između gradova, primjenjuje se dvostrani t-test.

Hipoteze koje ćemo testirati su sljedeće:

Nul hipoteza: Ne postoji značajna razlika u prosječnim plaćama između zaposlenika u Zagrebu i Osijeku

$$H_0: \mu_1 = \mu_2$$

Alternativna hipoteza: Postoji statistički značajna razlika u prosječnim plaćama između zaposlenika u Zagrebu i Osijeku

$$H_1: \mu_1 \neq \mu_2$$

Podaci su prikupljeni za 50 zaposlenika iz Zagreba i jednako toliko zaposlenika iz Osijeka, svih iz istog poduzeća. Prosječna mjesečna plaća zaposlenih u zagrebačkoj podružnici iznosila je 1200 eura, dok je prosječna plaća zaposlenika iz Osijeka iznosila 1100 eura.

Prosječna plaća za zaposlenike u Zagrebu (\bar{x}_1) iznosi 1200 eura, a prosječna plaća za Osijek (\bar{x}_2) iznosi 1100 eura. Standardne devijacije (s_1 i s_2) iznose 150 eura. Broj zaposlenih za svaki grad (n_1, n_2) je 50.

Odabir izraza za izračun testne veličine ovisi o tome jesu li varijance jednake.

Ako su jednake, onda se koristi izraz: $t = \frac{\bar{x}_1 - \bar{x}_2}{s(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, pri čemu je s_p udružena

(*engl. pooled*) standardna devijacija, prema čemu se i takav test naziva *pooled t-test*.

Pritom se udružena standardna devijacija računa na sljedeći način,

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}.$$

Ako varijance nisu jednake koristi se izraz: $t = \frac{\bar{x}_1 - \bar{x}_2}{s(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$. Stupnjevi slobode

jednaki su sumi veličine dva uzorka minus dva.

Određivanje kritične t-vrijednosti:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{49 \cdot 150^2 + 49 \cdot 150^2}{50 + 50 - 2}} = 150$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1200 - 1100}{150 \cdot \sqrt{\frac{1}{50} + \frac{1}{50}}} = \frac{100}{30} \approx 3.33$$

$$df = n_1 + n_2 - 2 = 50 + 50 - 2 = 98$$

Na temelju odabrane razine značajnosti (alfa = 0.05) i stupnjeva slobode (df = 98), pronalaze se kritične t-vrijednosti iz tablica t-distribucije (Slika 6.). Izračunata t-vrijednost (3.33) uspoređuje se s kritičnom t-vrijednosti iz tablica t-distribucije.

Slika 6. Graf t - distribucije za df = 98 i alfa = 0.05



Izvor: samostalna izrada koristeći <https://www.geogebra.org/classic#probability>

Prema tablici t-distribucije s 98 stupnjeva slobode i alfa = 0.05, teorijska t-vrijednost (dvostrani test) koja predstavlja granicu između područja odbacivanja i neodbacivanja iznosi $t_{\frac{\alpha}{2}} = \pm 1.9845$. P-vrijednost pripisana izračunatoj vrijednosti t-testa iznosi 0.00125, što je manje od razine značajnosti od 5% i upućuje na odluku o odbacivanju nul hipoteze. Zaključuje se da postoji statistički značajna razlika u prosječnim plaćama između zaposlenika u Zagrebu i Osijeku.

5. F – test

ANOVA test koristi F-vrijednosti, za određivanje statističke značajnosti rezultata, pa je zato važno upoznati se s F-testom prije prikaza ANOVA testiranja. Nakon što se dobije omjer F, samo ga je potrebno usporediti s tablicom kritičnih vrijednosti kako bi se utvrdila statistička značajnost rezultata.

F-test je statistički test koji se koristi za usporedbu omjera varijanci između dvije ili više skupina podataka. „Varijanca je prosječna suma kvadrata odstupanja vrijednosti obilježja (veličine) od aritmetičke sredine.“²⁰ Varijanca je mjera raspršenosti podataka i odražava koliko pojedinačne vrijednosti variraju od srednje vrijednosti. Ako su odstupanja manja ili su vrijednosti blizu prosjeka, varijanca će biti manja.

Kada uspoređujemo varijance između različitih skupina, ključno je razumjeti da ne možemo izravno uspoređivati vrijednosti varijanci među različitim skupinama ako se radi o varijablama mjenim u različitim mjernim jedinicama. F-test uspoređuje varijabilnost između uzoraka (mjerenu koliko se srednje vrijednosti razlikuju) s varijabilnošću unutar uzorka (mjerenu koliko su pojedinačne vrijednosti raznolike u svakoj skupini).

F-test jednakosti varijanci je statistički test koji se koristi za usporedbu varijabilnosti (ili rasipanja) opažanja dvaju ili više uzoraka kako bi se utvrdilo jesu li njihove varijance statistički jednake. Često se koristi kao prethodni korak prije izvođenja drugih statističkih analiza koje pretpostavljaju jednakost varijanci, poput t-testa ili analize varijance (ANOVA).

F-test uključuje usporedbu omjera varijanci iz različitih uzoraka. Statistička veličina testa, označena kao F, izračunava se dijeljenjem veće varijance uzorka s manjom varijancom uzorka. Ako uzorci imaju jednake varijance, očekivana vrijednost F-a bit će blizu 1.

Nul hipoteza (H_0) za F-test je da su varijance uspoređenih uzoraka jednake. Alternativna hipoteza (H_1) sugerira da varijance nisu jednake. F-test uspoređuje

²⁰ Hrvatska enciklopedija, mrežno izdanje, <https://www.enciklopedija.hr/natuknica.aspx?ID=63913>, (pristupljeno 30. svibnja 2021.)

izračunatu vrijednost F-a s kritičnom vrijednosti iz F-distribucije za odabrani stupanj značajnosti. Ako je izračunata vrijednost F-a veća od kritične vrijednosti, odbacuje se nul hipoteza, sugerirajući da varijance nisu jednake.

Formula za izračun testne veličine (F-vrijednosti) za F-test jednakosti varijanci između dva uzorka je:

$$F = \frac{s_1^2}{s_2^2},$$

gdje su s_1^2 i s_2^2 varijance prvog i drugog uzorka. Testna veličina F se uspoređuje s kritičnim vrijednostima iz F-distribucije kako bi se donijela odluka o odbacivanju ili neodbacivanju nul hipoteze. Ako se provodi analiza s više od dva uzorka, formula za izračun F-vrijednosti će biti složenija i ovisit će o broju faktora i stupnjevima slobode u analizi.

Za primjer, pretpostavimo da razmatrate investiranje u dionice kompanija A i B kako biste odabrali koja investicija ima manju varijancu u povratima. Za investitora, veća varijabilnost dionica znači veći rizik u vezi s povratima na ulaganje. Varijabilnost ili fluktuacije u povratima dionica predstavljaju promjene u vrijednosti dionica tijekom vremena. Ako je vrijednost dionica podložna velikim fluktuacijama, investitor će se suočiti s većom nesigurnošću u vezi s budućim povratima. Kada su dionice veoma varijabilne, postoji veća vjerojatnost da će investitor doživjeti velike dobitke, ali isto tako i velike gubitke. Ovo može biti posebno izazovno za one investitore koji preferiraju stabilnost i manji rizik. Ako investitor želi odabrati hoće li investirati u dionice kompanije A ili B, prvo treba prikupiti podatke o vrijednostima dionica u što dužem vremenskom periodu za obje kompanije. Potom treba izračunati varijancu za oba skupa podataka. Sljedeći korak je izračunavanje F-vrijednosti, a potom usporedba s kritičnom vrijednošću iz F-distribucije za zadani nivo značajnosti (npr. 0.05). Ako je izračunata F-vrijednost veća od kritične vrijednosti, može se zaključiti da postoji statistički značajna razlika u varijancama povrata između dionica kompanija A i B. To će utjecati na odluku o investiranju, jer veća varijanca povrata može ukazivati na veći rizik.

5.1. Pretpostavke f-testa

Da bi se podaci mogli analizirati F-testom, nekoliko pretpostavki mora biti zadovoljeno. Pretpostavke za provođenje F-testa jednakosti varijanci su sljedeće:

1. Normalna distribucija: Podaci u svim uzorcima trebaju biti izvučeni iz normalno distribuirane populacije. Ako distribucija nije normalna, rezultati testa mogu biti nepouzdana.
2. Homogenost varijanci: Varianca (rasipanje) podataka u uzorcima trebala bi biti približno jednaka za sve uspoređene skupine. Ako su varijance znatno različite, F-test može dati nepouzdan rezultate.
3. Nezavisnost uzoraka: Uzorci koji se uspoređuju trebaju biti neovisni jedni o drugima, tj. vrijednosti u jednom uzorku ne bi smjele ovisiti o vrijednostima u drugom uzorku.
4. Izjednačeni uzorci: Veličine uzoraka trebaju biti približno jednake za sve uspoređene skupine. Različite veličine uzoraka mogu utjecati na rezultate testa.

Ako navedene pretpostavke nisu ispunjene, rezultati F-testa mogu biti nepouzdana, te bi se trebalo razmotriti korištenje drugih metoda ili testova koji su manje osjetljivi na kršenje tih pretpostavki.

5.2. Primjer provedbe f-testa

U hipotetskom primjeru, pretpostavlja se da investitor razmatra investiranje u dionice kompanija A i B kako bi se odabralo koja investicija ima manju varijancu u povratima. S obzirom da se ispituje postoji li razlika između varijanci vrijednosti dionica dviju kompanija, hipoteze su zadane kao:

$$H_0 \dots s_A^2 = s_B^2$$

$$H_1 \dots s_A^2 \neq s_B^2.$$

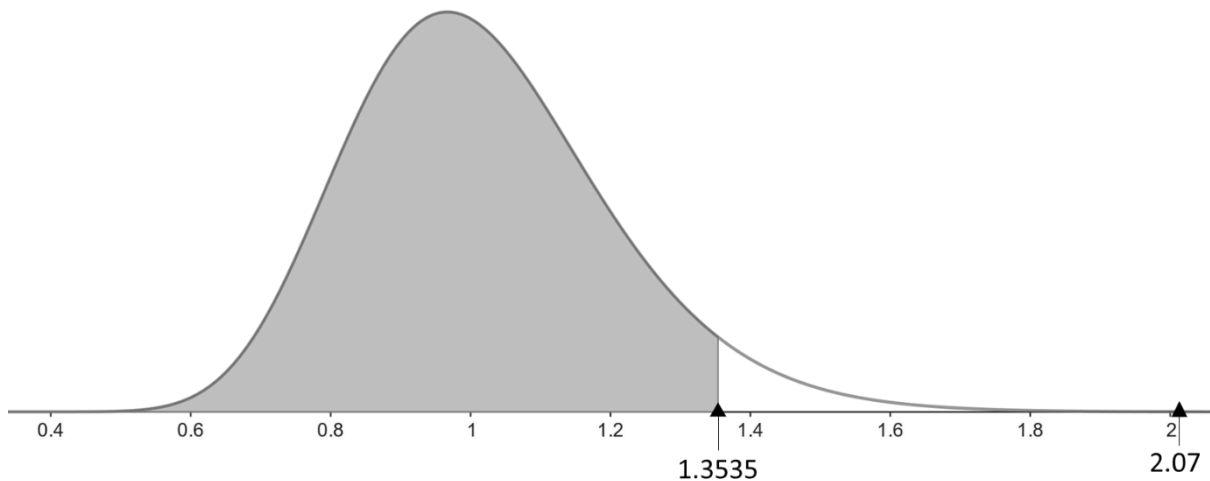
Recimo da imamo podatke o vrijednosti dionica na posljednji dan u mjesecu za posljednjih 10 godina za obje kompanije. Dakle, u svakom uzorku postoji 120

opažanja, pa je $df_A = df_B = 119$. S obzirom na veliki broj stupnjeva slobode, distribucija će oblikom biti sličnija normalnoj (u odnosu na distribucije s nižim vrijednostima stupnjeva slobode). Izračunate su varijance vrijednosti dionica kompanija; $s_A^2 = 0.029$ i $s_B^2 = 0.014$.

Sljedeći korak je izračunavanje F-vrijednosti koristeći formulu

$$F = \frac{s_1^2}{s_2^2} = \frac{0.029}{0.014} = 2.07$$

Slika 7. Graf F-distribucije s $F_\alpha = 1.3535$ i $F = 2.07$



Izvor: samostalna izrada koristeći <https://www.geogebra.org/classic#probability>

Uz razinu značajnosti 5%, teorijska F veličina iznosi $F_\alpha = 1.3535$ i predstavlja granicu između područja odbacivanja i područja neodbacivanja nul hipoteze (Slika 7.). p-vrijednost pridružena testnoj veličini je $p(f = 2.07) = 0.000045$ i manja je od odabrane razine značajnosti. Navedeno upućuje na odbacivanje nul hipoteze. Zaključuje se kako postoji statistički značajna razlika u varijancama vrijednosti dionica, odnosno postoji statistički značajna razlika u varijabilnosti vrijednosti dviju dionica, pri čemu vrijednosti dionica poduzeća A variraju znatno više od dionica poduzeća B. Dakle, ako investitor nije sklon riziku, dionice poduzeća B ($s_B^2 = 0.014$) su bolji odabir.

6. Univarijatna analiza varijance

Univarijatna analiza varijance ili ANOVA je statistički postupak koji je osmislio Ronald A. Fisher, a kojim se utvrđuje statistička značajnost razlika između aritmetičkih sredina dviju ili više grupa entiteta u jednoj varijabli.

Pri tome je moguće postaviti sljedeću nultu (H0), odnosno alternativnu (H1) hipotezu:

- H0: Uz pogrešku p ne možemo tvrditi da su razlike između aritmetičkih sredina analiziranih grupa statistički značajne.
- H1: Razlike između aritmetičkih sredina analiziranih grupa statistički su značajne uz pogrešku p .²¹

Osnovna logika univarijatne analize varijance temelji se na omjeru varijabiliteta između grupa (engl. between groups) i varijabiliteta unutar grupa (engl. within groups). Ako je varijabilitet između grupa statistički značajno veći nego varijabilitet unutar grupa, onda se grupe međusobno statistički značajno razlikuju, odnosno, ne pripadaju istoj populaciji.²²

Analiza varijance se sastoji od skupa postupaka kojima se raščlanjuje varijanca prema izvorima varijabilnosti njezinih vrijednosti. Upotrebljava se u mnogim područjima statistike kao što su analiza nacrti statističkih pokusa, testiranje hipoteze o parametru u regresijskim modelima itd.

Analiza varijance nije jednostavna metoda i uključuje niz modela:

- jednosmjerna (one-way) analiza varijance (ANOVA)
 - za zavisne i nezavisne uzorke
- višesmjerna (multi-way) analiza varijance
- multivarijatna analiza varijance (MANOVA)
- analiza varijance s kontrolom kovarijabli (ANCOVA ili MANCOVA)

²¹ Ž. Pedišić, D. Dizdar, *Priručnik za kvantitativne metode*, Kineziološki fakultet Sveučilišta u Zagrebu, Zagreb, 2010., str. 177.

²² D. Dizdar, *Kvantitativne metode*, Kineziološki fakultet Sveučilišta u Zagrebu, Zagreb, 2006., str. 150.

Analiza varijance (ANOVA) je statistička metoda koja se koristi za raščlanjivanje ukupne varijabilnosti unutar skupa podataka na dva osnovna dijela: sustavne čimbenike i slučajne čimbenike. Sustavni čimbenici predstavljaju varijabilnost koja se može pripisati neovisnoj varijabli ili faktoru, dok slučajni čimbenici predstavljaju varijabilnost koja nije povezana s neovisnom varijablom i smatra se stohastičkom ili slučajnom. ANOVA test se koristi kako bi se utvrdio statistički utjecaj neovisne varijable na zavisnu varijablu u regresijskim ili eksperimentalnim studijama.²³

Ova metoda omogućuje statističko procjenjivanje razlika između više grupa ili faktora te identificira jesu li te razlike statistički značajne. ANOVA je korisna u mnogim područjima statistike, uključujući analizu eksperimenata, istraživanje različitih faktora koji utječu na zavisnu varijablu te testiranje hipoteza o tim faktorima.

ANOVA test se koristi kao početni korak u analizi faktora koji utječu na određeni skup podataka. Nakon što je test završen, dodatna analiza se provodi kako bi se istražili metodički faktori koji mjerljivo doprinose varijabilnosti u podacima. Rezultate ANOVA testa analitičar koristi u f-testu kako bi generirao dodatne podatke koji se podudaraju s predloženim regresijskim modelima. Ovim se koracima omogućava bolje razumijevanje kako faktori povezani s neovisnom varijablom mogu kvantitativno objasniti i modelirati za bolje predviđanje budućih podataka.

6.1. Jednofaktorska analiza varijance

Jednofaktorska ANOVA, poznata i kao Fisherova analiza varijance, razvijena je 1918. godine i koristi se za ispitivanje postoji li značajna razlika između aritmetičkih sredina tri ili više nepovezanih skupina. Uglavnom se testira nul hipoteza koja glasi da ne postoji stvarna razlika u prosjecima populacija između skupina koje se proučavaju. Kod analize varijance s jednim promjenjivim faktorom cilj je ispitati odnos varijabilnosti podataka između uzoraka (grupa) i varijabilnosti podataka unutar uzoraka (grupa). Broj uzoraka određen je brojem grupa, tj. brojem modaliteta ili razina promatranog faktora.

²³ Analiza varijance, <https://hr.thelittlecollection.com/anova-12786> (pristupljeno 30. svibnja 2021.)

Analiza varijance zahtijeva odabir dviju varijabla:

1. Varijabla odgovora: Numerička varijabla koja se ispituje naziva se "varijablom odgovora" ili "zavisnom varijablom". Ovom varijablom se mjeri učinak ili rezultat.
2. Varijabla tretmana: Varijabla koja razdvaja sudionike u različite kategorije zove se "varijablom tretmana" ili "nezavisnom varijablom". To je varijabla koja može utjecati na varijabilnost varijable odgovora. Varijabla tretmana može biti nominalnog ili ordinalnog tipa i može imati proizvoljan broj kategorija, uključujući i samo dvije.²⁴

Pretpostavke jednosmjerne analize varijance (ANOVA):

1. Slučajni uzorak: Uzorci koji se koriste u ANOVA testiranju trebaju biti slučajno uzeti iz populacije. To znači da svaki član populacije ima jednaku vjerojatnost biti uključen u uzorak.
2. Normalna raspodjela: Pretpostavlja se da podaci u svakoj grupi dolaze iz populacije koja je normalno raspoređena. Ova pretpostavka je važna jer se temelji na pretpostavci da su testne statistike koje se koriste u ANOVA distribuirane normalno.
3. Homogenost varijanci: To znači da varijance u svim grupama trebaju biti približno jednake.
4. Kontinuirana zavisna varijabla: ANOVA se obično primjenjuje na kontinuirane zavisne varijable.

Prije provedbe ANOVA testa, podatke je potrebno testirati kako bi se utvrdilo zadovoljavaju li pretpostavke i to se obično provodi koristeći sljedeće testove:

- Shapiro-Wilk testom ili Kolmogorov-Smirnov testom - za testiranje normalnosti distribucije varijable;
- Leveneovim testom ili Brown-Forsythe testom - za testiranje homogenosti varijanci.

Alternativno, zaključak o normalno distribuiranoj varijabli može se izvesti iz histograma, uz izračun mjera zaobljenosti i asimetrije. Osim toga, treba imati na umu da

²⁴ One-way ANOVA, <http://www.efos.unios.hr/poslovni-informacijski-sustavi/wp-content/uploads/sites/491/2020/07/Dokument-3.pdf>, (pristupljeno 30. svibnja 2021.)

pretpostavka nezavisnosti ne smije ni u kojem slučaju biti narušena (rezultati nisu pouzdani u tom slučaju). Pretpostavka normalnosti smije biti narušena, ali samo ako je uzorak dovoljno velik. Također, pretpostavka o jednakosti varijanci smije biti narušena, ali samo u situaciji ako su veličine grupa jednake.

Pretpostavimo da imamo nezavisne uzorke od redom n_1, n_2, \dots, n_k opažanja iz k populacija. Ako su sredine populacija označene s $\mu_1, \mu_2, \dots, \mu_k$ onda je okvir za jednofaktorsku analizu varijance dizajniran za testiranje nul hipoteze.²⁵

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1: \mu_i \neq \mu_{j'} \text{ - za najmanje jedan par } \mu_i, \mu_{j'}$$

Gdje μ znači prosjek skupine, a k označava broj skupina.

Nul hipoteza pretpostavlja da su svi prosjeci skupina, označeni μ , jednaki međusobno. Odnosno, nul hipoteza tvrdi da ne postoji statistički značajna razlika između prosjeka bilo koje dvije skupine među k skupina. Alternativna hipoteza suprotna je nul hipotezi i tvrdi da postoji najmanje jedan par skupina čiji su prosjeci, μ_i i μ_j , različiti. Odnosno, barem za jedan par skupina među k skupina postoji statistički značajna razlika u prosjecima.

Kada se radi o analizi varijance s jednim faktorom i ciljem utvrđivanja statistički značajnih razlika između više grupa, prvo se izračunava varijanca između uzoraka i varijanca unutar uzoraka. Ovi rezultati se zatim koriste za izračun omjera odstupanja između i unutar skupina, što dovodi do rezultata F-testa:

$$F = \frac{MSB}{MSW}$$

pri čemu MSB označava prosječna kvadratna odstupanja između skupina, a MSW označava prosječna kvadratna odstupanja unutar skupina. Prosječna kvadratna odstupanja između skupina računaju se prema sljedećem izrazu:

$$MSB = \frac{SSB}{df_B} = \frac{\sum_{k=1}^G n_i (\bar{x}_g - \bar{x})^2}{df_B},$$

²⁵ P. Newbold, W. L. Carlson, B. Thorne, *Statistika za poslovanje i ekonomiju*, Mate d.o.o., Zagreb, 2010., str. 635

gdje je SSB suma kvadratnih odstupanja između prosjeka skupina, n_i broj opažanja u pojedinoj skupini g od ukupno G skupina, \bar{x}_g je prosjek skupine, \bar{x} je prosjek svih opažanja, a df_B označava broj stupnjeva slobode i jednak je broju skupina minus jedan. Prosječna kvadratna odstupanja unutar skupina računaju se prema formuli:

$$MSW = \frac{SSW}{df_W} = \frac{\sum_{k=1}^G \sum_{j=1}^i (x_{gi} - \bar{x}_g)^2}{df_W},$$

pri čemu je SSW suma kvadratnih odstupanja unutar grupe, x_{ij} je pojedinačno opažanje u grupi g , \bar{x}_g je prosjek grupe g .

Ako je rezultat F-testa blizu 1, to ukazuje na to da odstupanje između grupa nije značajno različito od odstupanja unutar grupa, sugerirajući da grupe vjerojatno potječu iz populacija s istim prosjekom. No, ako je rezultat F-testa značajno veći od 1, to implicira da je varijabilnost među grupama značajno veća od varijabilnosti unutar svake pojedinačne grupe. Takav rezultat sugerira da razlike među grupama vjerojatno potječu iz različitih populacija s različitim srednjim vrijednostima.

Ako ANOVA daje statistički značajan rezultat (npr. $p < 0.05$), to ukazuje na postojanje razlika između barem dvije skupine. Međutim, samo ANOVA neće odrediti koje su konkretne skupine međusobno različite. Za identifikaciju konkretnih razlika između skupina koriste se post-hoc testovi. U ovom radu se nećemo baviti post-hoc testovima jer je cilj osnovno razumijevanje ANOVA-e.

F-test kao jednofaktorska analiza varijance koristi se za procjenu razlika u očekivanim vrijednostima kvantitativne varijable između nekoliko unaprijed definiranih skupina. Na primjer, pretpostavimo da se provodi medicinsko istraživanje u kojem se uspoređuju četiri tretmana. ANOVA F-test može se koristiti kako bi se utvrdilo je li neki od tretmana u prosjeku bolji ili lošiji u odnosu na ostale.

Ako imamo samo dva uzorka, može se koristiti t-test za usporedbu načina uzorkovanja. Međutim, u slučaju više od dva uzorka, takav pristup može postati nepouzdan. Na primjer, pretpostavimo da se želi analizirati efekt razine obrazovanja na prihode u jednoj zemlji. Neka su prikupljeni podaci o godišnjim prihodima pojedinaca koji su završili različite razine obrazovanja: srednja škola, sveučilišni diplomski studij i magisterij. Pitanje je postoji li statistički značajna razlika u srednjim prihodima između

ove tri grupe obrazovanja. Provedba testa jednofaktorske analize varijance (ANOVA) na podacima o prihodima s obzirom na razinu obrazovanja rezultira F-statistikom i pripadajućom p-vrijednosti. Ako je p-vrijednost manja od odabrane razini značajnosti (npr. 0,05), odbacit će se nul-hipoteza i zaključiti da postoji statistički značajna razlika u prihodima među barem jednim parom grupa obrazovanja.

6.2. Dvofaktorska analiza varijance

Dvofaktorska analiza varijance predstavlja produžetak jednofaktorske analize varijance i koristi se za uspoređivanje srednjih vrijednosti između grupa koje su podijeljene na temelju dvaju faktora. Njezin glavni cilj je istražiti postoji li interakcija između ta dva nezavisna faktora (varijable) i kako ta interakcija utječe na zavisnu varijablu. Također, omogućuje zaključak o tome je li utjecaj jednog od tih nezavisnih faktora na zavisnu varijablu jednak za sve vrijednosti drugog nezavisnog faktora.²⁶

Ovakva analiza pomaže u razumijevanju kompleksnih odnosa između više varijabli i kako te varijable zajedno utječu na ishode. Kroz dvofaktorsku analizu varijance, istraživači mogu bolje razumjeti kako različite kombinacije faktora mogu imati različite učinke na rezultate, što može pružiti dublje uvide u prirodu tih učinaka i njihove praktične implikacije.

U postupku dvosmjernog ANOVA testa, također se izračunava F-vrijednost, slično kao i kod jednosmjernog ANOVA testa. F-vrijednost se koristi za testiranje jesu li srednje vrijednosti u potpunom dvosmjernom modelu statistički značajno različite. Postupak je sličan jednosmjernoj ANOVA analizi, s iznimkom da se koriste dva kategorijska polja umjesto jednoga.²⁷

Prije nego što se započne s dvofaktorskom analizom, podaci se trebaju testirati kako bi se provjerilo zadovoljavaju li pretpostavke odabrane metode. Varijabla odgovora (zavisna varijabla) treba biti kontinuirana, dok bi varijable tretmana (nezavisne

²⁶ Lumen Learning - Two-Way ANOVA, <https://courses.lumenlearning.com/boundless-statistics/chapter/two-way-anova/> (pristupljeno 20. svibnja 2021.)

²⁷ IBM - Analiza varijance (ANOVA), <https://www.ibm.com/docs/hr/cognos-analytics/11.1.0?topic=tests-analysis-variance-anova> (pristupljeno 20. svibnja 2021.)

varijable) trebale biti kvalitativne (kategoričke) varijable s dva ili više modaliteta obilježja. Zavisna varijabla bi se trebala distribuirati prema normalnoj raspodjeli za svaku kombinaciju skupina dviju nezavisnih varijabli, a također bi trebala ispunjavati pretpostavke o homogenosti varijanci.

Budući da dvofaktorska ANOVA uzima u obzir utjecaj dviju kategorijskih varijabli i njihov međusobni utjecaj na zavisnu varijablu, postoje tri para nul ili alternativnih hipoteza za dvofaktorsku ANOVU. Ove hipoteze obuhvaćaju utjecaj svakog faktora pojedinačno, utjecaj njihove interakcije te općenito razlike između skupina stvorenih kombinacijom razina tih faktora. Dvije neovisne varijable u dvosmjernoj ANOVA nazivaju se faktorima. Ideja je da postoje dvije varijable, čimbenici, koje utječu na zavisnu varijablu. Svaki faktor ima dvije ili više razina, pri čemu stupnjevi slobode za svaki čimbenik jedan su manji od broja razina.

Kod dvofaktorske univarijatne analize varijance, računamo prosjek kvadrata odstupanja od prosjeka (varijancu) za svaku nezavisnu varijablu, njihovu interakciju i rezidual (pogrešku). Potom provodimo tri pripadajuća F-testa.²⁸

Na primjer, želimo provesti analizu kako bismo istražili postoji li razlika u razini anksioznosti tijekom razgovora za posao s obzirom na spol i razinu dohotka ispitanika. Razina anksioznosti je varijabla odgovora, izmjerena na kontinuiranoj omjernoj ljestvici. Spol i dohodak su varijable tretmana, pri čemu je spol opisna kvalitativna varijabla s dvije razine (muški i ženski), a dohodak je ordinalna kvalitativna varijabla podijeljena u tri razine: niska, srednja i visoka. Za zaključivanje o postojanju veze i razlika u razini anksioznosti koristit ćemo odgovarajuće statističke testove, kao što su hi-kvadrat test neovisnosti ili testiranje hipoteze o koeficijentu korelacije, te ćemo interpretirati rezultate temeljem vjerojatnosti postojanja veze između varijabli.

²⁸ S. Repišti., „Problem razumijevanja varijance i kovarijance i postupaka njihovog računanja u psihometriji“, Stručni rad, Istraživanje matematičkog obrazovanja (2012.), http://www.imvibl.org/dmbl/meso/imo/imo_vol_4_2012/imo_vol_4_2012_7_31_43.pdf

6.3. Primjeri provedbe ANOVA-e

U hipotetskom primjeru provjerava se utjecaj obrazovanja na zarade radnika. Neovisni faktor je razina obrazovanja, koja ima više od dvije razine (srednja škola, preddiplomski studij, diplomski studij). Odabrana je jednofaktorska ANOVA jer omogućuje testiranje razlika među više od dvije razine obrazovanja u kontekstu kontinuirane numeričke zavisne varijable (zarade), pružajući temelj za zaključivanje o utjecaju razina obrazovanja na zarade radnika.

Neka je 50 ispitanika iz svake skupine odabrano na slučajan način i zabilježene su njihove godišnje plaće u tisućama eura. Ti su podaci (pojedinačna opažanja prikazani tablicom 2:

Tablica 2. Podaci o godišnjim plaćama ispitanika prikazanim u tisućama eura

Srednja škola	8.90	5.99	6.86	9.85	12.45	4.43	5.51	6.30	8.77	9.28
	9.37	8.07	7.49	12.25	9.62	7.35	4.39	4.99	9.07	9.21
	11.95	7.34	11.48	9.69	6.15	8.98	9.04	5.94	8.07	6.51
	6.92	5.63	5.41	9.44	8.43	6.63	8.60	9.48	4.75	10.37
	6.15	13.69	9.17	8.38	9.12	7.40	5.28	7.81	7.89	8.92
	9.79	6.72	8.87	9.29	8.10	4.99	8.94	10.16	7.51	8.34
Preddiplomski studij	18.78	18.33	2.78	14.16	14.89	16.18	18.45	12.03	13.68	7.60
	13.06	15.87	11.29	14.05	15.50	16.26	13.82	10.88	11.86	19.77
	20.51	13.59	22.15	10.63	18.25	14.66	15.31	15.81	16.85	11.80
	20.85	20.65	17.51	17.74	12.89	14.11	16.70	10.34	16.39	13.23
	13.75	19.33	24.52	21.98	11.16	10.08	13.50	16.08	18.76	20.60
	18.89	16.57	15.63	9.20	13.16	11.79	11.43	11.64	12.60	13.96
Diplomski studij	16.85	16.02	18.21	18.95	20.82	20.78	17.06	16.60	24.95	23.22
	24.47	15.40	23.75	19.25	18.85	17.10	26.75	19.38	23.15	21.21
	20.41	20.25	18.37	18.39	22.65	15.52	18.25	14.67	15.61	15.57
	17.69	20.04	19.44	23.52	22.00	18.76	22.81	20.13	21.39	24.98
	25.17	21.38	17.07	21.54	22.86	20.78	17.26	14.07	19.71	20.99
	14.12	20.87	20.74	21.41	25.72	14.64	23.02	19.84	21.92	26.11

Izvor: Samostalno generirani podaci koristeći MS Excel funkciju Random number generator

Hipoteze koje će se testirati su sljedeće:

Nul hipoteza: Ne postoji značajna razlika u prosječnim zaradama između različitih razina obrazovanja (srednja škola, preddiplomski studij, diplomski studij)

$$H_0: \mu_1 = \mu_2 = \mu_3$$

Alternativna hipoteza: Postoji značajna razlika u prosječnim zaradama između najmanje dvije razine obrazovanja (barem jedan μ različit od ostalih).

$$H_1: \mu_i \neq \mu_{j'} - \text{za najmanje jedan par } \mu_i, \mu_{j'}$$

Svaka razina obrazovanja ima po 50 opažanja. U ovom slučaju, broj grupa je 3 (različite razine obrazovanja), pa je $df_A = 3 - 1 = 2$. Ukupan broj opažanja za sve tri razine obrazovanja je 150, pa je $df_B = 150 - 3 = 147$. Distribucija F-vrijednosti u ANOVA-i bit će približno normalna ili simetrična, s obzirom na velik broj stupnjeva slobode.

Kako bi stigli do testne veličine, u prvom koraku, izračunavaju se pomoćne statističke veličine, prosjeci skupina, prosjek svih opažanja te sume kvadrata odstupanja.

$$\bar{x}_1 = \frac{277.11}{50} = 5.54$$

$$\bar{x}_2 = \frac{665.85}{50} = 13.32$$

$$\bar{x}_3 = \frac{822.49}{50} = 16.45$$

$$\bar{x} = \frac{1765.45}{150} = 11.77$$

$$SSB = \sum_{k=1}^G n_i (\bar{x}_g - \bar{x})^2$$

$$= 50 \cdot (5,54222 - 11,76965)^2 + 50 \cdot (13,31692 - 11,76965)^2 + 50 \cdot (16,44982 - 11,76965)^2 = 7125,463$$

$$SSW = \sum_{k=1}^G \sum_{j=1}^i (x_{gi} - \bar{x}_g)^2 =$$

$$= (8,8983 - 5,54222)^2 + (5,9872 - 5,54222)^2 + \dots + (26,1135 - 16,44982)^2 = 7365,665$$

$$MSB = \frac{SSB}{df_B} = \frac{7125,463}{2} = 3562,7315$$

$$MSW = \frac{SSW}{df_W} = \frac{7365,665}{147} = 50,097$$

$$F = \frac{MSB}{MSW} = \frac{3562,7315}{50,097} \approx 71,166$$

Uz izabranu razinu značajnosti od 5%, utvrđeno je da teorijska F-vrijednost iznosi $F_\alpha = 3.39$, što predstavlja granicu između područja odbacivanja i područja ne odbacivanja nul hipoteze. P-vrijednost $p(F = 71.166) < 0.001$ značajno je manja od odabrane razine značajnosti, 0.05. Na temelju ovih rezultata, nul hipoteza o nepostojanju razlika u prosječnim zaradama između različitih razina obrazovanja odbačena je. To ukazuje na značajne razlike u prosječnim zaradama između najmanje dvije razine obrazovanja.

Neka je postojeći primjer obogaćen podatkom o spolu ispitanika. U tom slučaju, postoje dva faktora. Ako svako drugo promatranje u svakoj grupi predstavlja osobu ženskog spola, tada bi u analizi varijance s dvama faktorima, osim faktora obrazovanja, također imali faktor spola. Tada bi se korištena ANOVA mogla izgledati ovako:

1. **Suma kvadrata odstupanja unutar skupina (SSW):** izračunali bi varijancu unutar svake kombinacije faktora obrazovanja (A) i faktora spola (B) (npr. muškarci s faktorom A = 1 i faktorom B = 1, žene s faktorom A = 1 i faktorom B = 1, itd.).
2. **Suma kvadrata odstupanja među razinama faktora A (SSA):** izračunali bi varijancu koja se odnosi na razlike između različitih vrijednosti faktora A (npr. 1, 2, 3). Ovo mjeri utjecaj faktora A na varijancu rezultata.
3. **Suma kvadrata među faktorima B (SSB):** ovo je analogno SSA, ali za faktor B.
4. **Interakcijska suma kvadrata (SSAB):** izračunali bi interakcijsku varijancu između faktora A i faktora B, što mjeri ima li njihova kombinacija zajednički učinak na varijancu rezultata.
5. **Suma kvadrata za faktor spola (SSSpol):** izračunali bi varijancu koja se odnosi na razlike između muških i ženskih sudionika, neovisno o faktorima A i B.

Po izračunu testne veličine moguće je zaključivanje. Rezultati analize omogućuju nekoliko zaključaka. Prvo je zaključivanje o interakciji između faktora. Ako se ustanovi značajna interakcija između faktora razine obrazovanja i spola, to ukazuje da efekt jednog faktora na ciljanu varijablu varira ovisno o razini drugog faktora. Na primjer, mogli biste zaključiti da se prosječne vrijednosti godišnje plaće razlikuju među razinama obrazovanja, ali da se ta razlika razlikuje između muškaraca i žena. To bi ukazivalo na kompleksan utjecaj oba faktora na ciljanu varijablu.

Nadalje, može se zaključivati o značajnosti faktora. Ako faktor obrazovanja ukazuje na statistički značajne razlike, znači da razlike između razina obrazovanja imaju statistički značajan utjecaj na godišnju plaću, neovisno o spolu. Slično tome, statistička značajnost faktora spola ukazuje na to da spol ima statistički značajan utjecaj na ciljanu varijablu, bez obzira na razinu obrazovanja.

Osim toga, mogu se uočiti razlike među razinama faktora. Ako se pronađe statistički značajna razlika među razinama faktora obrazovanja, to ukazuje na statistički značajnu razliku u godišnjoj plaći između različitih razina obrazovanja, bez obzira na spol. Slično tome, značajne razlike među razinama faktora spola ukazuju na statistički značajne razlike u godišnjoj plaći između muškaraca i žena, bez obzira na razinu obrazovanja.

Kako bi se ilustrirao značaj ovakvih analiza i zaključaka do kojih je ovakvom analizom moguće doći, ovdje će se ukratko predstaviti situacije u kojima bi ovakvo istraživanje, koristeći dvofaktorsku ANOVA-u bilo korisno.

Na primjer, pri upravljanju ljudskim resursima. Ako se analiza provodi na primjeru plaća u poslovnoj organizaciji, rezultati bi mogli ukazati na postojanje rodno uvjetovanih razlika u plaćama, što bi moglo ukazivati na potrebu za promjenama u politikama plaća i praksama zapošljavanja kako bi se osigurala ravnopravnost.

Nadalje, pri kreiranju marketinških strategija. Uzimanje u obzir faktora obrazovanja i spola može pomoći u ciljanju određenih demografskih skupina u marketinškim kampanjama. Na primjer, ako se utvrdi da postoje značajne razlike u preferencijama ili potrebama s obzirom na različite razine obrazovanja i/ili spol, tvrtke mogu prilagoditi

svoje marketinške strategije kako bi bolje zadovoljile potrebe tih ciljnih skupina. Također, ako se analizom utvrdi da razina obrazovanja i spol utječu na potrebe i preferencije potrošača, tvrtke mogu koristiti te informacije za razvoj proizvoda i usluga prilagođenih tim ciljnim skupinama.

Osim toga, analiza faktora koji utječu na primanja i ekonomske uvjete može pomoći u oblikovanju ekonomske politike. Ako se, na primjer, ustanovi da obrazovanje i spol igraju ključnu ulogu u ekonomskoj nejednakosti, to bi moglo potaknuti donositelje politike da poduzmu mjere za smanjenje tih nejednakosti. Također, analiza može pružiti uvid u obrazovnu strukturu i spolnu ravnotežu radne snage u određenoj industriji ili sektoru. To je korisno za planiranje obrazovnih programa i politika zapošljavanja.

U konačnici, zaključci iz dvofaktorske ANOVA analize mogu pomoći u donošenju informiranih poslovnih, ekonomskih i političkih odluka koje se odnose na različite aspekte društva i gospodarstva. Naravno, potrebno je prikupiti odgovarajuće podatke s obzirom na cilj istraživanja i postavljeno istraživačko pitanje.

7. Primjena analize varijanci u ekonomiji/poslovnoj ekonomiji

ANOVA i F-testovi se primjenjuju u ekonomiji/poslovnoj ekonomiji kako bi se provjerilo postoji li statistički značajna razlika u prosječnim vrijednostima između različitih skupina ili uvjeta. Osnovna logika ovih testova je utvrditi postoji li razlika u promatranim pojavama s obzirom na druge promatrane pojave. Važno je naglasiti da ovi statistički testovi ne impliciraju uzročnost. Razlike koje se otkrivaju pomoću ANOVA i F-testova ukazuju samo na statistički značajne razlike u prosječnim vrijednostima, ali ne ukazuju na uzročne veze između varijabli.

Kada koristimo korelacijske testove, provjeravamo povezanost između dvije ili više varijabli. Međutim, korelacija ne znači nužno uzročnost. Samo zato što postoji povezanost između dvije varijable, ne znači da jedna uzrokuje drugu.

Za istraživanje uzročnosti, koriste se regresijski modeli i eksperimentalni dizajn. Regresijski modeli pomažu identificirati uzročne veze između varijabli, kontrolirajući utjecaj ostalih varijabli. Eksperimentalni dizajn omogućuje kontrolirano manipuliranje neovisnim varijablama kako bi se utvrdio njihov uzročni utjecaj na ovisne varijable. Važno je napomenuti da postoje iznimke kada se ANOVA i F-testovi koriste za utvrđivanje statističke značajnosti regresijskog modela, podacima dobivenih anketiranjem ili u eksperimentalnom dizajnu istraživanja. Međutim, sami po sebi, ANOVA i F-testovi nisu dovoljni za zaključivanje o uzročnosti.

Ukratko, ANOVA i F-testovi koriste se za otkrivanje razlika u prosječnim vrijednostima između skupina, korelacijski testovi za povezanost između varijabli, a eksperimentalni dizajn za utvrđivanje uzročnih veza.

U radu „Mjerenje zadovoljstva klijenata uslugama u drogerijama u gradu Zagrebu“²⁹ analizirano je zadovoljstvo klijenata uslugama u drogerijama u Zagrebu. Prikupljanjem podataka putem anketa obuhvaćen je uzorak od 408 kupaca drogerija u Zagrebu. Deskriptivna statistička analiza korištena je za opisivanje demografskog profila i kupovnih navika ispitanika. Rezultati ANOVA-e su pokazali da postoje statistički

²⁹ Šušnjar, T. i Zakarija A., "Mjerenje zadovoljstva klijenata uslugama u drogerijama u gradu Zagrebu." Poslovna izvrsnost 10, br. 2 (2016): 221-239. <https://hrcak.srce.hr/170762>, (pristupljeno 28. kolovoza 2023.)

značajne razlike u zadovoljstvu klijenata s različitim aspektima usluge i među različitim lancima drogerija. Spol, različite dobne skupine i zaposlenost također su kreirali razlike u zadovoljstvu klijenata. Unatoč korisnim uvidima, postoje ograničenja. Rad se temelji na ograničenom uzorku i proveden je samo u Zagrebu, što može ograničiti generalizaciju rezultata.

U radu "Ekonomska uspješnost gradova u Hrvatskoj - statistička analiza"³⁰ primjenjuje se analiza varijance za istraživanje ekonomske uspješnosti gradova u Hrvatskoj. Naglasak je na utvrđivanju razlika u uspješnosti poslovanja gradskih samouprava i poduzetnika, uz razmatranje faktora kao što su veličina gradova, županijska pripadnost i politička struktura vlasti. Korištene statističke metode uključuju Levenov test homogenosti varijance, ANOVA-u i Sheffeov post-hoc test. Rezultati pokazuju statistički značajne razlike u ekonomskoj uspješnosti gradova, posebno u udjelu poreznih prihoda u ukupnim prihodima. Također se ističu razlike između malih i velikih gradova te gradova u različitim geografskim regijama.

U članku „Analiza ekonomskog života hrvatskih općina“³¹ provodilo se statističko ispitivanje različitih ekonomskih ishoda općina i poduzeća u 427 općina u periodu od 2007 do 2011. Ispitana je uloga veličine općina, administrativni status, geografska lokacija i dominantne stranke u općini koristeći ANOVA-u. Rezultati ukazuju da veličina općina značajno utječe na objašnjenje varijacija u fiskalnoj sposobnosti općina i prosječnim neto primanjima stanovnika. Nadalje, politička pripadnost gradonačelnika povezana je s razlikama u relativnom iznosu pomoći odobrene iz državnog proračuna, fiskalnoj sposobnosti i zaduženosti. Također, varijacije u administrativnom statusu općina doprinose nejednakostima u zapošljavanju, prosječnim neto primanjima, zaduženosti, pomoći iz državnog proračuna i izdacima za socijalnu zaštitu. Ukratko, ključni zaključak istraživanja je da se hrvatske općine značajno razlikuju u većini analiziranih ekonomskih čimbenika, naglašavajući potrebu za sveobuhvatnom

³⁰ Rašić Bakarić, I., Šimović H. i Vizek M., "Ekonomska uspješnost gradova u Hrvatskoj - statistička analiza." Ekonomski pregled 65, br. 2 (2014): 115-138. <https://hrcak.srce.hr/121452>, (pristupljeno 28. kolovoza 2023.)

³¹ Rašić Bakarić, I., Šimović H. i Vizek M., "Analiza ekonomskog života hrvatskih općina." Zbornik radova Ekonomskog fakulteta u Rijeci 32, br. 2 (2014): 285-312. <https://hrcak.srce.hr/131536>, (pristupljeno 28. kolovoza 2023.)

reformom lokalne samouprave koja bi se bavila ne samo njihovim posebnim statusom i brojem lokalnih jedinica, već i procesom fiskalne izjednačenosti.

U radu „Utjecaj osnovnih značajki korisnika na primjenu informacijskih sustava“³² provedeno je istraživanje odnosa korisnika i informacijskih sustava u Vladi Brčko distrikta BiH i Javnom poduzeću Komunalno Brčko te je analizirano kako osnovne značajke korisnika utječu na njihovo doživljavanje informacijskih sustava koje koriste. Rezultati, dobiveni korištenjem statističkih analiza poput MANOVA i ANOVA, potvrdili su statistički značajnu razliku između većine značajki korisnika s obzirom na način korištenja informacijskih sustava, osim u slučaju dobi korisnika gdje nije utvrđena statistička značajnost. Ovi rezultati pružaju bolje razumijevanje faktora koji utječu na percepciju i korištenje informacijskih sustava u radnom okruženju.

U primjeru iz članka „Mjerenje zadovoljstva klijenta uslugama na kioscima u gradu Rijeci“³³ provedeno je istraživanje temeljem prikupljanja podataka putem anketa s ciljem ocjene zadovoljstva klijenata uslugama na kioscima u gradu Rijeci. Analiza podataka uključuje metode deskriptivne statističke analize, t-test i analizu varijance. Rezultati ukazuju na visoku razinu zadovoljstva klijenata brzinom usluge i urednošću osoblja, dok su manje zadovoljni cijenama proizvoda i izgledom prodajnih mjesta. Osim toga, istraživanje pokazuje da demografske karakteristike i navike kupovanja ne utječu značajno na ocjene zadovoljstva klijenata, osim što zaposleni klijenti izražavaju veće zadovoljstvo izgledom prodajnih mjesta u usporedbi s umirovljenicima. Iako su ovi rezultati važni za razumijevanje zadovoljstva klijenata na kioscima u Rijeci, istraživanje ima ograničenja kao što su ograničen uzorak i regionalna fokusiranost. Buduća istraživanja bi trebala proširiti obuhvat na druge gradove i uključiti dodatne varijable kako bi se dobila sveobuhvatnija slika o zadovoljstvu klijenata uslugama na kioscima.

³² Šadić, S., Puška A., Beganović, A. I. i Berbić S., "Utjecaj osnovnih značajki korisnika na primjenu informacijskih sustava." *Oeconomica Jadertina* 7, br. 2 (2017): 35-50. <https://hrcak.srce.hr/191390> (pristupljeno 28. kolovoza 2023.)

³³ Raspor, S., Žunić, C. i Ribarić Aidone E., "Mjerenje zadovoljstva klijenta uslugama na kioscima u gradu Rijeci." *Zbornik Veleučilišta u Rijeci* 1, br. 1 (2013): 17-30. <https://hrcak.srce.hr/103425> (pristupljeno 28. kolovoza 2023.)

U primjeru istraživanja iz članka „Utjecaj sustava kvaliteta na zadovoljstvo i odanost studenata VŠ eMPIRICA“³⁴ analiza varijance (ANOVA) i druge multivarijantne analize koriste se za proučavanje utjecaja percepcije kvalitete na zadovoljstvo i odanost studenata visokoškolskoj ustanovi. Ispitivanje je provedeno na Visokoj školi eMPIRICA u Brčko distriktu BiH. Rezultati istraživanja ukazuju na statistički značajnu povezanost između percepcije kvalitete i zadovoljstva te odanosti studenata. Ovo istraživanje također sugerira da studenti druge godine i studenti smjera Inženjerska informatika imaju viši stupanj zadovoljstva i odanosti školi u usporedbi s prvim godinama i studentima smjera Poslovna informatika. Ovo istraživanje pokazuje kako percipirana kvaliteta obrazovnih usluga može značajno utjecati na zadovoljstvo i odanost studenata, što je relevantno za visoko obrazovanje u ekonomskom kontekstu.

U radu „Utvrđivanje odnosa između menadžerove sklonosti riziku i karijernog uspjeha“³⁵ se istražuje veza između menadžerove sklonosti riziku i njegovog karijernog uspjeha na uzorku menadžera zaposlenih u hrvatskim hotelskim poduzećima. Korištena je jednofaktorska analiza varijance za analizu ove veze. Rezultati sugeriraju da su uspješniji menadžeri skloniji riziku, posebice kada se uspjeh mjeri objektivnim kriterijima. Međutim, kada se uspjeh izražava subjektivno, razlika u sklonosti riziku nije statistički značajna. Ovaj zaključak ukazuje na važnost sklonosti riziku u menadžerskoj profesiji i njenom utjecaju na karijerni uspjeh, s potencijalnim implikacijama za organizacijski uspjeh.

U radu „Efikasnost hrvatskih trgovačkih sudova u procesuiranju stečajnih predmeta“³⁶ se koristi analiza varijance kako bi se utvrdila razlika u efikasnosti trgovačkih sudova u različitim gradovima. Rezultati pokazuju da postoje značajne razlike u efikasnosti sudova, s nekim sudovima koji su ispodprosječno efikasni, dok su drugi nadprosječno efikasni u obradi stečajnih predmeta. Također, istraživanje otkriva da postoji značajan broj dugotrajnih stečajeva, što ukazuje na potrebu za unaprjeđenjem efikasnosti stečajnog sustava u Hrvatskoj.

³⁴ Puška, A., Stanišić D. i Maksimović A., "Utjecaj sustava kvaliteta na zadovoljstvo i odanost studenata VŠ eMPIRICA." *Praktični menadžment* 6, br. 1 (2015): 15-21. <https://hrcak.srce.hr/144010> (pristupljeno 28. kolovoza 2023.)

³⁵ Vojinić, P., "Utvrđivanje odnosa između menadžerove sklonosti riziku i karijernog uspjeha." *Poslovna izvrsnost* 8, br. 1 (2014): 157-168. <https://hrcak.srce.hr/123582> (pristupljeno 30. kolovoza 2023.)

³⁶ Sajter, D. "Efikasnost hrvatskih trgovačkih sudova u procesuiranju stečajnih predmeta." *Ekonomski pregled* 65, br. 2 (2014): 155-178. <https://hrcak.srce.hr/121454> (pristupljeno 30. kolovoza 2023.)

U radu „Utjecaj komparativnog oglašavanja na stratešku adaptivnost marki stranih proizvođača u Republici Hrvatskoj“³⁷ istražuje se utjecaj izravnog i neizravnog komparativnog oglašavanja na stavove generacije Y o strateškoj adaptivnosti stranih marki proizvođača u Hrvatskoj. Analiza varijance korištena je kako bi se utvrdile statistički značajne razlike u strateškoj adaptivnosti u odnosu na vrste komparativnog oglašavanja. Rezultati ukazuju da postoji statistički značajna razlika u stavovima o strateškoj adaptivnosti s obzirom na izravno vizualno-privlačno oglašavanje i neizravno informativno komparativno oglašavanje. Međutim, treba napomenuti da se rezultati baziraju na populaciji generacije Y u dobi od 18 do 35 godina, što može ograničiti općenitost zaključaka na cijelo tržište.

U primjeru iz članka „Pojave u kretanju strukture kapitala hrvatskih poduzeća u predrecesijskom razdoblju (2002. - 2007.)“³⁸ se provodi analiza promjene u strukturi kapitala hrvatskih poduzeća od 2002. do 2007. godine, analizirajući tri financijska koeficijenta: koeficijent zaduženosti, odnos obveza prema kapitalu i faktor zaduženosti. Korištena je analiza varijance za ponovljena mjerenja na reprezentativnom uzorku od 1500 hrvatskih poduzeća podijeljenih prema djelatnosti i regijama. Rezultati ukazuju da je najčešći način promjene strukture kapitala onaj u kojem temeljni kapital ostaje nepromijenjen, dok vlastita sredstva rastu putem zadržane dobiti, istovremeno s povećanjem sredstava iz tuđih izvora. Istraživanje također ističe razlike u strukturi kapitala među sektorima i regijama, pri čemu trgovina ima najveći omjer obveza prema kapitalu, dok su usluge manje zadužene.

³⁷ Bašić, M. i Đurđević D.. "Utjecaj komparativnog oglašavanja na stratešku adaptivnost marki stranih proizvođača u Republici Hrvatskoj." Zbornik Ekonomskog fakulteta u Zagrebu 15, br. 1 (2017): 79-96. <https://doi.org/10.22598/zefzg.2017.1.79> (pristupljeno 30. kolovoza 2023.)

³⁸ Penavin, S. i Šarilja N., "Pojave u kretanju strukture kapitala hrvatskih poduzeća u predrecesijskom razdoblju (2002. - 2007.)" Ekonomski vjesnik XXIII, br. 2 (2010): 317-331. <https://hrcak.srce.hr/63246> (pristupljeno 28. kolovoza 2023.)

8. Zaključak

U ovom radu se govori o analizi varijance i njezinoj primjeni u ekonomiji. Za primjenu analize varijance, ključno je razumjeti koncepte inferencijalne statistike kao što su standardna pogreška aritmetičke sredine, standardna devijacija i tipovi pogrešaka (tip I i tip II) u kontekstu testiranja hipoteza. Održavanje ravnoteže između ovih tipova pogrešaka ključno je za valjane zaključke iz statističke analize. Nadalje, potrebno je poznavanje osnovnih karakteristika povezanih teorijskih distribucija vjerojatnosti, koje su prikazane u radu.

Upotreba i razvoj testova opisan je kroz gradaciju i kompleksnost zahtjeva koji moraju biti zadovoljeni da bi se test mogao provesti. Prikaz započinje s normalnom i standardiziranom normalnom distribucijom te vezanim z-testom. No, ako određeni uvjeti nisu ispoštovani, taj se pristup ne može koristiti. Ako su uzorci mali, t-testovi su primjerena alternativa za usporedbu prosjeka. Jednostavni t-testovi, poput onih s jednim i dva uzorka te t-testa za uparene podatke, omogućuju precizno uspoređivanje prosjeka različitih skupina. No, ako se žele uspoređivati varijance ili prosjeci više od dva uzoraka (skupina), potrebno je potražiti alternative.

U takvim situacijama, koriste se testovi vezani uz F-distribuciju. Važno je napomenuti razlike između različitih vrsta F-testova i njihovih specifičnih primjena. F-test koristi se za utvrđivanje razlika u varijancama dva uzorka. Jednofaktorska ANOVA koristi se za testiranje značajnih razlika između prosjeka tri ili više skupina, dok se dvofaktorska ANOVA primjenjuje za ispitivanje razlika s obzirom na dva faktora.

Analiza varijance ima ključnu ulogu u ekonomiji i poslovnoj ekonomiji jer omogućuje detaljno istraživanje razlika među različitim skupinama ili faktorima čija se uloga očitava u ekonomskim ishodima. Primjeri iz stvarnih istraživanja pomažu boljem razumijevanju metode i njezine primjene u konkretnim situacijama, istovremeno potičući kritičko razmišljanje i precizno definiranje istraživačkih pitanja i varijabli kako bi se osigurala valjana statistička analiza. Kroz primjere iz stvarnih istraživanja, ova metoda postaje još korisnija za analizu ekonomske složenosti i donošenje informiranih poslovnih odluka.

Literatura

Knjige:

1. Biljan-August, M., Pivac, S. i Štambuk, A., *Statistička analiza u ekonomiji*, Ekonomski fakultet Sveučilišta u Rijeci, Rijeka, 2009.
2. Dizdar, D., *Kvantitativne metode*, Kineziološki fakultet Sveučilišta u Zagrebu, Zagreb, 2006.
3. Horvat, J. i Mijoč, J., *Osnove statistike*, 2, izdanje, Naklada Ljevak, Zagreb, 2014.
4. Newbold, P., Carlson, W. L. i Thorne, B., *Statistika za poslovanje i ekonomiju*, Mate d.o.o., Zagreb, 2010.
5. Pajić, D., *Primena tehnika vizualizacije u bazičnoj statistici*, Filozofski fakultet u Novom Sadu, Novi Sad, 2020.
6. Pedišić, Ž., Dizdar, D., *Priručnik za kvantitativne metode*, Kineziološki fakultet Sveučilišta u Zagrebu, Zagreb, 2010.
7. Plackett, R.L., „Studies in the history of probability and statistics. VII. The principle of the arithmetic mean“ u *Biometrika*, vol. 45, 1958.

Web članci:

1. Bašić, M. i Đurđević D.. "Utjecaj komparativnog oglašavanja na stratešku adaptivnost marki stranih proizvođača u Republici Hrvatskoj." Zbornik Ekonomskog fakulteta u Zagrebu 15, br. 1 (2017): 79-96.
<https://doi.org/10.22598/zefzg.2017.1.79> (pristupljeno 30. kolovoza 2023.)
2. Penavin, S. i Šarilja N., "Pojave u kretanju strukture kapitala hrvatskih poduzeća u predrecesijskom razdoblju (2002. - 2007.)." Ekonomski vjesnik XXIII, br. 2 (2010): 317-331. <https://hrcak.srce.hr/63246> (pristupljeno 28. kolovoza 2023.)
3. Puška, A., Stanišić D. i Maksimović A., "Utjecaj sustava kvaliteta na zadovoljstvo i odanost studenata VŠ eMPIRICA." Praktični menadžment 6, br. 1 (2015): 15-21. <https://hrcak.srce.hr/144010> (pristupljeno 28. kolovoza 2023.)

4. Raspor, S., Žunić, C. i Ribarić Aidone E., "Mjerenje zadovoljstva klijenta uslugama na kioscima u gradu Rijeci." Zbornik Veleučilišta u Rijeci 1, br. 1 (2013): 17-30. <https://hrcak.srce.hr/103425> (pristupljeno 28. kolovoza 2023.)
5. Rašić Bakarić, I., Šimović H. i Vizek M., "Ekonomska uspješnost gradova u Hrvatskoj - statistička analiza." Ekonomski pregled 65, br. 2 (2014): 115-138. <https://hrcak.srce.hr/121452>, (pristupljeno 28. kolovoza 2023.)
6. Rašić Bakarić, I., Šimović H. i Vizek M., "Analiza ekonomskog života hrvatskih općina." Zbornik radova Ekonomskog fakulteta u Rijeci 32, br. 2 (2014): 285-312. <https://hrcak.srce.hr/131536>, (pristupljeno 28. kolovoza 2023.)
7. Repišti, S., „Problem razumijevanja varijance i kovarijance i postupaka njihovog računanja u psihometriji“, Stručni rad, Istraživanje matematičkog obrazovanja (2012.), http://www.imvibl.org/dmbl/meso/imo/imo_vol_4_2012/imo_vol_4_2012_7_31_43.pdf, (pristupljeno 22. svibnja 2021.)
8. Sajter, D. "Efikasnost hrvatskih trgovačkih sudova u procesuiranju stečajnih predmeta." Ekonomski pregled 65, br. 2 (2014): 155-178. <https://hrcak.srce.hr/121454> (pristupljeno 30. kolovoza 2023.)
9. Šadić, S., Puška A., Beganović, A. I. i Berbić S., "Utjecaj osnovnih značajki korisnika na primjenu informacijskih sustava." Oeconomica Jadertina 7, br. 2 (2017): 35-50. <https://hrcak.srce.hr/191390> (pristupljeno 28. kolovoza 2023.)
10. Šušnjar, T. i Zakarija A., "Mjerenje zadovoljstva klijenata uslugama u drogerijama u gradu Zagrebu." Poslovna izvrsnost 10, br. 2 (2016): 221-239. <https://hrcak.srce.hr/170762>, (pristupljeno 28. kolovoza 2023.)
11. Vojinić, P., "Utvrđivanje odnosa između menadžerove sklonosti riziku i karijernog uspjeha." Poslovna izvrsnost 8, br. 1 (2014): 157-168. <https://hrcak.srce.hr/123582> (pristupljeno 30. kolovoza 2023.)

Internetski izvori:

1. Analiza poslovnih podataka, http://www.efos.unios.hr/analiza-poslovnih-podataka/wp-content/uploads/sites/160/2013/04/4_teorijske-distribucije.pdf (pristupljeno 30. travnja 2021.)

2. Analiza varijance, <https://dwstockburger.com/Introbook/sbk21.htm> (pristupljeno 9. svibnja 2021.)
3. Analiza varijance, <https://hr.thelittlecollection.com/anova-12786> (pristupljeno 30. svibnja 2021.)
4. Arithmetic Mean. In: The Concise Encyclopedia of Statistics. Springer, New York, NY., https://doi.org/10.1007/978-0-387-32833-1_12 (pristupljeno 30. kolovoza 2023.)
5. Hrvatska enciklopedija, mrežno izdanje,
6. <http://www.enciklopedija.hr/Natuknica.aspx?ID=3839> (pristupljeno 28. travnja 2021.)
7. Hrvatska enciklopedija, mrežno izdanje, <https://www.enciklopedija.hr/natuknica.aspx?ID=63913> (pristupljeno 30. svibnja 2021.)
8. Hrvatska enciklopedija, mrežno izdanje <https://www.enciklopedija.hr/natuknica.aspx?ID=57896> (pristupljeno 30. travnja 2021.)
9. IBM - Analiza varijance (ANOVA), <https://www.ibm.com/docs/hr/cognos-analytics/11.1.0?topic=tests-analysis-variance-anova> (pristupljeno 20. svibnja 2021.)
10. Kontinuirane slučajne varijable, <https://www.pmf.unizg.hr/download/repository/PREDAVANJE9.pdf> (pristupljeno 30. travnja 2021.)
11. Lumen Learning - Two-Way ANOVA, <https://courses.lumenlearning.com/boundless-statistics/chapter/two-way-anova/> (pristupljeno 20. svibnja 2021.)
12. Obrada i prikaz kvantitativnih varijabli, https://bookdown.org/gkardum1/r_znanost_obrazovanje/obrada-i-prikaz-kvantitativnih-varijabli.html (pristupljeno 21. svibnja 2021.)
13. Obrada i prikaz kvantitativnih varijabli – Jednostavni t-test, https://bookdown.org/gkardum1/r_znanost_obrazovanje/obrada-i-prikaz-kvantitativnih-varijabli.html#jednostavni-t-test-ili-t-test-na-jednom-uzorku (pristupljeno 21. svibnja 2021.)

14. One-way ANOVA, <http://www.efos.unios.hr/poslovni-informacijski-sustavi/wp-content/uploads/sites/491/2020/07/Dokument-3.pdf> (pristupljeno 30. svibnja 2021.)
15. Primjer parametrijskog testa: Studentov test, <https://vub.hr/images/uploads/1471/oirus-statistika-predavanje-sat-12.pdf> (pristupljeno 21. svibnja 2021.)
16. Uvod u deskriptivnu i inferencijalnu statistiku, <https://www.unizd.hr/Portals/12/Nastavnici/Irena%20Buric/Uvod%20u%20deskriptivnu%20i%20inferencijalnu%20statistiku.pdf> (pristupljeno 30. travnja 2021.)

Popis slika

Slika 1. Standardizirana normalna distribucija	16
Slika 2. Studentova t – distribucija	19
Slika 3. F – distribucija.....	21
Slika 4. Graf t- distribucije za $df=15$	23
Slika 5. Dvosmjerna t - distribucija.....	24
Slika 6. Graf t - distribucije za $df = 98$ i $\alpha = 0.05$	27
Slika 7. Graf F-distribucije s $F\alpha = 1.3535$ i $F = 2.07$	32

Popis tablica

Tablica 1. Usporedba karakteristika normalne i standardizirane normalne distribucije	17
Tablica 2. Podaci o godišnjim plaćama ispitanika prikazanim u tisućama eura	40

Sažetak

Na početku rada detaljno se razmatraju osnovni pojmovi analize varijance (ANOVA), uključujući aritmetičku sredinu i njena svojstva te njihova veza s momentima oko sredine, posebno s varijancom kao drugim momentom oko sredine. Nadalje, obraća se pažnja na inferencijalnu statistiku te se objašnjavaju teorijske distribucije vjerojatnosti koje se primjenjuju u okviru analize varijance, s posebnim fokusom na karakteristike normalne distribucije, Studentove t-distribucije i Snedecorove f-distribucije. Rad dalje istražuje primjenu t-testa, proučavajući njegove pretpostavke i podjelu na jedan i dva uzorka. Prikazuju se primjeri primjene t-testa kako bi se olakšalo razumijevanje njegove svrhe i koristi. Također, istražuje se f-test, analizirajući njegove pretpostavke i pružajući primjere primjene. Ovaj dio istraživanja ima za cilj razjasniti ulogu f-testa u analizi varijance. U daljnjem tijeku rada, obrađuje se univarijatna analiza varijance, uključujući jednofaktorsku i dvofaktorsku analizu varijance, s naglaskom na identificiranje i procjenu varijacija između skupova podataka. U radu su prikazani primjeri provedbe ANOVA-e kako bi se ilustrirala njena praktična primjena u analizi statističkih podataka i donošenju zaključaka temeljenih na rezultatima istraživanja. Također, detaljno je objašnjena uloga analize varijance u ekonomiji i poslovnoj ekonomiji kroz primjere iz stvarnih istraživanja, kako bi se dobio uvid u primjenu ove metode u praksi.

Ključne riječi: analiza varijance, studentova distribucija, snedecorova distribucija, t-test, f - test, jednofaktorska analiza, dvofaktorska analiza

Summary

The beginning of the study thoroughly examines the fundamental concepts of Analysis of Variance (ANOVA), including the arithmetic mean and its properties, as well as their connection to moments around the mean, particularly the variance as the second moment around the mean. Furthermore, attention is given to inferential statistics and the explanation of theoretical probability distributions applied within the context of ANOVA, there is a special focus on the characteristics of the normal distribution, Student's t-distribution, and Snedecor's F-distribution. The research further explores the application of the t-test, studying its assumptions and the differentiation between one-sample and two-sample t-tests. Examples of t-test applications are provided to facilitate understanding of its purpose and benefits. Similarly, the research delves into the f-test, analyzing its assumptions and presenting practical examples of its application. This part of the research aims to clarify the role of the f-test in analysis of variance. In the subsequent sections, univariate analysis of variance is discussed, encompassing both one-way and two-way analysis of variance, with a focus on identifying and estimating variations among data sets. This study includes examples of conducting ANOVA to illustrate its practical application in the analysis of statistical data and drawing conclusions based on research results. Furthermore, the role of analysis of variance in economics and business economics is explained in detail through examples from real research to gain insight into the practical application of this method in practice.

Key words: analysis of variance, student's distribution, snedecor's distribution, t-test, f-test, one-factor analysis, two-factor analysis