STATISTIČKE METODE ZA ANALIZU LOGISTIČKIH PODATAKA

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

Autori:
Sanja Bojić
Kristijan Brglez
Maja Fošner
Roman Gumzej
Rebeka Kovačič Lukman
Benjamin Marcen
Marinko Maslarić
Boško Matović
Dejan Mirčetić

Sufinancira Europska unija

Financira Europska unija.
Međutim, izraženi stavovi i mišljenja pripadaju samo autoru(-ima) i ne moraju nužno biti odražavaju one Europske unije ili Europske izvršne agencije za obrazovanje i kulturu (EACEA).
Ni Europska unija ni EACEA ne mogu se smatrati odgovornima za njih.

Izdavač:

Politechnika Poznańska

Poznanj, Poljska

Recenzenti:
dr. sc. Ajda Fošner, Sveučilište Primorska, Fakultet za menadžment
dr. Nikša Alfirević, Sveučilište u Splitu, Ekonomski fakultet

Tehnički urednik:
Kristijan Brglez

Prijevodi:
Prof. dr.sc. Sanja Bojić
Kristijan Brglez
Prof. dr. Sc. Maja Fošner
Prof. dr. sc. Roman Gumzej
Prof. dr. sc. Rebeka Kovačič Lukman
Doc. dr. sc. Benjamin Marcen
Prof. dr. sc. Marinko Maslarić
Doc. dr. sc. Boško Matović
dr. Dejan Mirčetić
Doc. dr. sc. Jelena Franjković

Tiskano izdanje

Online izdanje

ISBN: 978-83-62285-72-3

Monografija je nastala u okviru projekta Business Analytics Skills for the Future-proof Supply Chains (BAS4SC) [2022-1-PL01-KA220-HED-000088856], financiranog iz programa Erasmus+. Ovaj projekt financiran je uz potporu Europske komisije. Ova publikacija odražava samo stajališta autora, a Komisija se ne može smatrati odgovornom za bilo kakvu upotrebu informacija sadržanih u njoj.

Monografija je besplatno dostupna na: enauka.put.poznan.pl

Predgovor

Područje logistike sve se više oslanja na uvide temeljene na podacima kako bi se optimiziralo poslovanje, smanjili troškovi i osigurala učinkovitost u opskrbnim lancima. Ovaj udžbenik, *Statističke metode za analizu logističkih podataka*, služi kao ključni izvor za studente i stručnjake, pružajući im potrebne vještine za snalaženje u složenosti modernog upravljanja opskrbnim lancem. Razvijena kao dio projekta Business Analytics Skills for Future-proof Supply Chains (BAS4SC), ova knjiga nudi opsežan pregled statističkih metoda, upravljanja podacima i naprednih analitičkih tehnika eksplicitno prilagođenih logističkoj industriji.

Kroz pažljivo istraživanje, ovaj udžbenik rješava nedostatke u trenutnoj ponudi obrazovanja kombinirajući teoretsko znanje s praktičnom primjenom. Deset poglavlja uključivalo je detaljno istraživanje ključnih tema kao što su predviđanje potražnje, simulacijsko modeliranje, regresijska analiza i integracija umjetne inteligencije i strojnog učenja u logističke operacije. Korištenjem široko priznatih alata kao što su SPSS, R i SQL, sadržaj ovog udžbenika osmišljen je da premosti jaz između akademskog obrazovanja i potreba industrije.

Nadamo se da će ovaj udžbenik pružiti temeljno znanje o poslovnoj analitici za logistiku i potaknuti inovacije u tom području, potičući buduće lidere koji su spremni uhvatiti se u koštac s izazovima opskrbnih lanaca koji se brzo razvijaju.

Prof. dr. sc. Sanja Bojić
Kristijan Brglez
Prof. dr. Sc. Maja Fošner
Prof. dr. sc. Roman Gumzej
Prof. dr. sc. Rebeka Kovačič Lukman
Doc. dr. sc. Benjamin Marcen
Prof. dr. sc. Marinko Maslarić
Doc. dr. sc. Boško Matović
dr. Dejan Mirčetić

SADRŽAJ

UVOD

Ovaj udžbenik, pod naslovom *Statističke metode za analizu logističkih podataka*, treći je u nizu udžbenika razvijenih u sklopu projekta *Business Analytics Skills for Future-proof Supply Chains* (BAS4SC). Provedeno je nekoliko preliminarnih istraživanja kako bi se odredio sadržaj ovog udžbenika. Prvo je provedeno opsežno istraživanje kako bi se ispitalo postojeći kolegiji poslovne analitike, njihov sadržaj i vještine koje prenose studentima logistike diljem Europske unije, Sjedinjenih Američkih Država i Ujedinjenog Kraljevstva. Ova je analiza otkrila jaz između logističkog znanja i statističkih vještina potrebnih na terenu te onih koji se trenutno nude studentima. Na temelju dubinskih intervjua sa sveučilišnim nastavnim osobljem, studentima i stručnjacima iz industrije, više od 100 vještina poslovne analitike identificirano je kao neophodno. Koristeći ABC metodu klasifikacije rangiranja, 33 vještine odabrane su za uključivanje u ovu knjigu, prvenstveno usmjerene na matematiku, informatiku, menadžment, primijenjenu matematiku i statistiku. Kombinacija ovih utvrđenih potreba i vještina dovela je do razvoja deset poglavlja sadržaja koja se bave najkritičnijim vještinama potrebnih u tom području.

Prvo poglavlje pokriva *Uvodnu statistiku* i pruža sveobuhvatan pregled statističkih koncepata i njihovih primjena, posebice unutar analize opskrbnog lanca. Započinje naglašavanjem ključne uloge statistike u optimizaciji opskrbnih lanaca. Koristi deskriptivnu statistiku kao što je srednja vrijednost, medijan i standardna devijacija za analizu vremena isporuke, razine zaliha i troškova. Poglavlje predstavlja prediktivne tehnike poput regresije i analize vremenskih nizova za prognoziranje potražnje i zaliha. Dalje istražuje važnost varijabli, razlikujući kvalitativne i kvantitativne tipove, te zadire u temeljne statističke mjere kao što su srednja vrijednost, medijan, mod, varijanca i standardna devijacija.

Osim toga, pokriva metode grafičkog predstavljanja podataka, kao što su histogrami i raspršeni dijagrami, te ističe razliku između deskriptivne i inferencijalne statistike. Konačno, uvodi ključne koncepte korelacije, regresije i distribucije vjerojatnosti, nudeći alate za razumijevanje odnosa između varijabli i modeliranje slučajnih pojava u podacima. Ove statističke tehnike pomažu poboljšati donošenje odluka u opskrbnom lancu, učinkovitost i upravljanje rizikom.

Drugo poglavlje, *Statistika za poslovnu analitiku*, istražuje bitne statističke koncepte i tehnike za dobivanje uvida iz poslovnih podataka. Započinje uvođenjem važnosti analize podataka u poslovnom odlučivanju. Objašnjava temeljnu ulogu normalne distribucije, koja služi kao osnova za brojne statističke metode. Poglavlje se bavi standardnom devijacijom, naglašavajući njenu važnost u mjerenju varijabilnosti podataka. Također pokriva sampling-distribuciju i centralni granični teorem, objašnjavajući kako oni zaključuju o parametrima populacije iz podataka uzorka. Teme kao što su testiranje hipoteza, Z-rezultati i t-rezultati istražuju se kao pomoć pri donošenju odluka i izračunima vjerojatnosti. Poglavlje završava raspravom o standardnoj pogrešci i intervalima pouzdanosti, koji pomažu kvantificirati nesigurnost koja okružuje procjene. U konačnici, ovo poglavlje čitateljima daje statističke alate potrebne za poslovnu analitiku, omogućujući im donošenje informiranih odluka vođenih podacima.

Poglavlje o *Upravljanju podacima* istražuje različite aspekte upravljanja podacima u logistici, usredotočujući se na formate podataka, organizaciju i tehnologije. Počinje s ulogom elektroničke razmjene podataka (engl. *Electronic Data Interchange* - EDI) u razmjeni informacija unutar opskrbnog lanca korištenjem standardiziranih alfanumeričkih formata. Objašnjava koncept informacija, podataka i znanja, raspravljajući o tome kako se podaci digitaliziraju i organiziraju u baze podataka, skladišta i baze znanja. Poglavlje se bavi logističkim podacima, posebice upotrebom crtičnih kodova (bar kodova) i RFID oznaka za identifikaciju i praćenje u logistici. Također predstavlja tehnike organizacije podataka, u rasponu od proračunskih tablica do relacijskih baza podataka, objašnjava ključne koncepte kao što su primarni i strani ključevi, normalizacija i podatkovni upitni jezici poput SQL-a. Osim toga, u poglavlju se raspravlja o najboljim praksama za filtriranje podataka i sprječavanje pogrešaka tijekom unosa podataka. Na kraju, raspravlja se o razlikama između skladišta podataka i baza znanja, ističući njihove uloge u poslovnoj analizi i donošenju odluka.

Poglavlje o *Simulacijskom modeliranju i analizi* (engl. *Simulation Modelling and Analysis* - SMA) fokusirano je na stvaranje digitalnih modela za simulaciju sustava u stvarnom svijetu za optimizaciju i donošenje odluka. Započinje objašnjenjem Conant-Ashby teorema, koji sugerira da simulacijski model mora odgovarati složenosti svog pandana iz stvarnog svijeta kako bi se učinkovito regulirao. SMA optimizira opskrbne lance i prometne mreže u logistici, omogućujući menadžerima da simuliraju i procijene različite scenarije. Poglavlje opisuje ključne metodologije simulacije, kao što je simulacija diskretnih događaja (engl. *Discrete Event Simulation* - DES) za analizu usmjerenu na proces, sustavna dinamika (engl. *System Dynamics*

- SD) za performanse sustava visoke razine, simulacija temeljena na agentima (engl. *Agent-Based Simulation* - ABS) za modeliranje ponašanja pojedinačnih entiteta i simulacija mreže (engl. *Network Simulation* - NS) za analizu mrežnih tokova. Svaka metoda pruža uvid u različite aspekte logistike, od proizvodnih ciklusa do optimizacije prometa. Poglavlje završava pregledom simulacija logističkih projekata, strukturiranih oko Dizajna za šest sigma (engl. Six Sigma) i Demingovog ciklusa poboljšanja, koji pomažu u planiranju, izvođenju i poboljšanju složenih logističkih sustava.

Poglavlje *Linearna regresija s jednim i više regresora* uvodi regresijsku analizu za razumijevanje odnosa između zavisnih i nezavisnih varijabli. Započinje jednostavnom linearnom regresijom, gdje jedna nezavisna varijabla, poput izdataka za oglašavanje, predviđa ishod kao što je prodaja. U poglavlju se objašnjava konstrukcija regresijskog modela i regresijske jednadžbe koja se koristi za predviđanje zavisne varijable na temelju podataka uzorka. Također pokriva metodu najmanjih kvadrata, ključnu tehniku za procjenu regresijske linije minimiziranjem pogrešaka predviđanja. Zatim uvodi koeficijent determinacije (R^2) za mjerenje koliko dobro regresijski model odgovara podacima. Zatim se poglavlje bavi višestrukom regresijom, gdje dvije ili više neovisnih varijabli predviđaju zavisnu varijablu, nudeći sveobuhvatniju analizu. Primjeri uključuju predviđanje vremena putovanja na temelju udaljenosti i broja dostava.

Poglavlje *Uvod u operacijska istraživanja* usredotočeno je na korištenje analitičkih metoda za poboljšanje donošenja odluka, posebice u logistici i upravljanju opskrbnim lancem. Operacijska istraživanja koriste tehnike modeliranja, statistike i optimizacije za pronalaženje optimalnih rješenja za složene probleme, omogućavajući učinkovito upravljanje resursima, kontrolu zaliha i optimizaciju procesa. Poglavlje naglašava strateško logističko planiranje, koje uključuje metode kao što su Six Sigma i Just-in-Time proizvodnja za poboljšanje operativne učinkovitosti. Poslovna inteligencija (engl. *business intelligence* - BI) i poslovna analitika (engl. *business analytics* - BA) ključni su u analizi podataka, omogućujući tvrtkama donošenje informiranih odluka korištenjem prognoziranja, prediktivne analitike i vizualizacije podataka. Poglavlje također uvodi višekriterijsko odlučivanje (engl. *multi-criteria decision-making* - MCDM), koje pomaže u ocjenjivanju i odabiru optimalnih rješenja na temelju različitih kriterija. Konačno, sustavi za podršku odlučivanju (engl. *decision support systems* - DSS) i inženjerstvo temeljeno na znanju (engl. *knowledge-based engineering* - KBE) pomažu u integraciji znanja i podataka

u procese donošenja odluka, dodatno poboljšavajući operativnu učinkovitost i strateško planiranje.

Poglavlje o *Statističkoj obradi podataka sa SPSS-om* predstavlja IBM-ov softver SPSS kao moćan alat za automatizaciju složene statističke analize, povećanje pouzdanosti i olakšavanje donošenja odluka. Objašnjava kako SPSS omogućuje uvoz, manipulaciju i pripremu podataka putem korisničkog sučelja. Poglavlje pokriva ključne funkcije poput deskriptivne statistike, izrade grafikona i vizualizacije podataka. Također predstavlja temeljne statističke testove - T-testove, korelaciju, hi-kvadrat i ANOVA - vodeći čitatelje kroz postavljanje i tumačenje svakog testa. Osim toga, istražuje alate za upravljanje podacima kao što su spajanje, razdvajanje i izračunavanje varijabli u skupovima podataka, pokazujući kako SPSS poboljšava statističku analizu u logistici i drugim domenama.

Poglavlje 8 istražuje *Poslovnu analitiku (BA)* i njezinu primjenu putem alata kao što su R i SQL za rješavanje poslovnih problema. BA ima za cilj poboljšati donošenje odluka i uspješnost tvrtke korištenjem metoda vođenih podacima. Uključuje deskriptivne, prediktivne i preskriptivne platforme za analizu podataka i donošenje informiranih odluka. R je predstavljen kao robustan alat za statističku analizu i vizualizaciju otvorenog koda, dok je SQL neophodan za upravljanje i postavljanje upita velikim bazama podataka. U poglavlju se detaljno opisuje integracija R-a i SQL-a za učinkovitu poslovnu analitiku, naglašavajući kako se podaci pohranjeni u SQL-u mogu analizirati pomoću R skripti za automatizaciju zadataka. Praktični primjeri, poput postavljanja upita bazi podataka Chinook, ilustriraju kako R i SQL rade zajedno za generiranje uvida, kao što je identificiranje najprodavanijih albuma. Ova sinergija između BA, R i SQL poboljšava sposobnost upravljanja i analize dinamičkih poslovnih podataka.

Poglavlje 9 fokusirano je na prognoziranje potražnje u opskrbnom lancu, vizualizaciju i inženjering značajki. Prognoziranje potražnje predviđa potrebe kupaca, mijenja cijeli opskrbni lanac i smanjuje logističke troškove. U poglavlju se obrađuju ključni koraci za prognoziranje potražnje, uključujući definiranje problema, prikupljanje podataka, analizu trendova, odabir modela i njihovu procjenu. Vizualizacija pomaže u prepoznavanju obrazaca kao što su sezonska kretanja i trendovi, koji mogu utjecati na odabir modela. S-ARIMA (engl. Seasonal Autoregressive Integrated Moving Average) ističe se kao učinkovit model za rukovanje složenim vremenskim serijama podataka, posebno sa sezonskim uzorcima potražnje. Primjer u prehrambenoj industriji pokazuje sposobnost modela S-ARIMA da predvidi potražnju i

usmjeri donošenje odluka. Konačno, poglavlje pokriva kako testirati i potvrditi modele predviđanja kako bi se osigurala njihova učinkovitost u aplikacijama u stvarnom svijetu, koristeći metrike kao što su RMSE i MAPE za procjenu učinka.

Posljednje poglavlje istražuje ulogu umjetne inteligencije (engl. *artificial intelligence* - AI) i strojnog učenja (engl. *machine learning* - ML) u opskrbnim lancima, počevši s pregledom razvoja AI-a od simboličkih sustava do modernih pristupa ML-u. AI se odnosi na automatizaciju zadataka koji obično zahtijevaju ljudsku inteligenciju, pri čemu je ML podskup AI-a koji se fokusira na učenje obrazaca iz podataka. U poglavlju se ističe kako se AI/ML modeli, poput nadziranog i nenadziranog učenja, primjenjuju za rješavanje poslovnih problema, uključujući prognoziranje potražnje, upravljanje zalihama i optimizaciju. Bitna studija slučaja uključuje primjenu AI i ML algoritama u centralnom skladištu tvornice hrane, optimizirajući korištenje viličara. Uključivanjem sustava za podršku odlučivanju (DSS), AI/ML modeli pomažu menadžerima u odabiru optimalnog broja viličara, poboljšavajući operativnu učinkovitost. Poglavlje naglašava kako AI/ML može obuhvatiti stručno znanje, smanjiti troškove i poboljšati procese donošenja odluka u upravljanju opskrbnim lancem.

1. Uvodna statistika

1.1 Uloga i važnost statistike u analizi podataka u opskrbnim lancima

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

Istaknimo neke od važnih uloga statistike u analizi opskrbnog lanca.

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

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

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

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

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

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

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

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

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

1.2 Osnovni pojmovi statistike

Varijable

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

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

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

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

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

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

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

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

1.3 Osnovni statistički koncepti s primjerima

Prosjek (srednja vrijednost)

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

Izračunavanje prosjeka:

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

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

Primjer:

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

Prosjek = (80 + 85 + 90 + 75 + 95) / 5 = 425 / 5 = 85

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

Medijan

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

Kako izračunati medijan?

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

Primjer:

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

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

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

Mod

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

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

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

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

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

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

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

Raspon varijacije

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

Izračun raspona varijacije je jednostavan:

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

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

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

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

Varijanca i standardna devijacija

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

Izračunavanje standardne devijacije:

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

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

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

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

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

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

Varijanca je srednja vrijednost ovih vrijednosti:

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

Na kraju, izračunavate standardnu devijaciju uzimajući kvadratni korijen varijance:

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

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

Kvantili

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

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

1.4 Prikaz statistike

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

Evo nekoliko uobičajenih načina za prikaz statistike:

Tablice

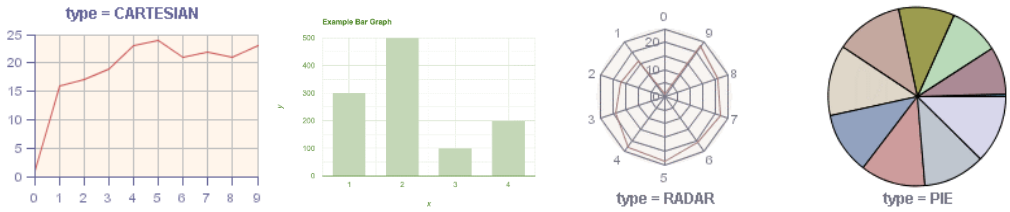
Tablice su osnovna metoda za prikaz podataka. Primjeri uključuju tablice frekvencija, koje pokazuju broj pojavljivanja za različite vrijednosti, i podatkovne tablice, koje prikazuju više informacija o podacima.

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

Slika 1.1 Primjer tablice.

Grafički prikazi

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



Slika 1.2 Primjeri grafičkih prikaza podataka.

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

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

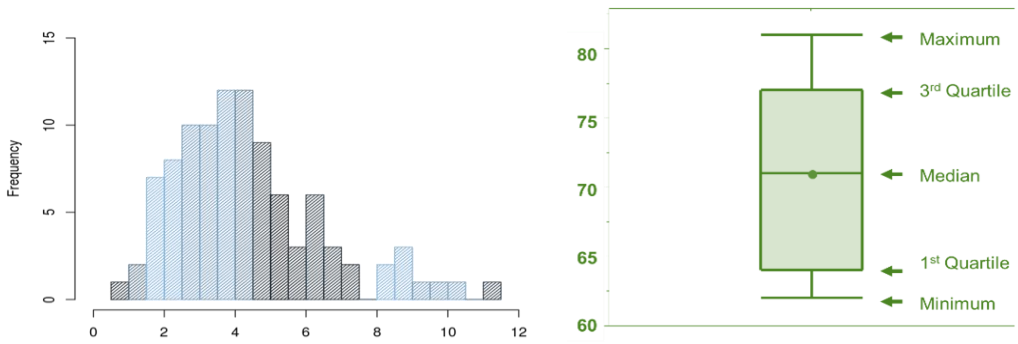
Stupčasti grafikoni koriste se za usporedbu količina u različitim kategorijama, što ih čini idealnim za predstavljanje diskretnih podataka. Posebno su učinkoviti za isticanje razlika, sličnosti i trendova među skupinama. Stupčasti grafikoni obično se koriste kada trebate prikazati frekvencije, postotke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

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

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

Histogrami

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



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

Dijagram pravokutnika (Box plot)

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

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

1.5 Distribucija frekvencija

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

Proces stvaranja distribucije frekvencija uključuje sljedeće korake:

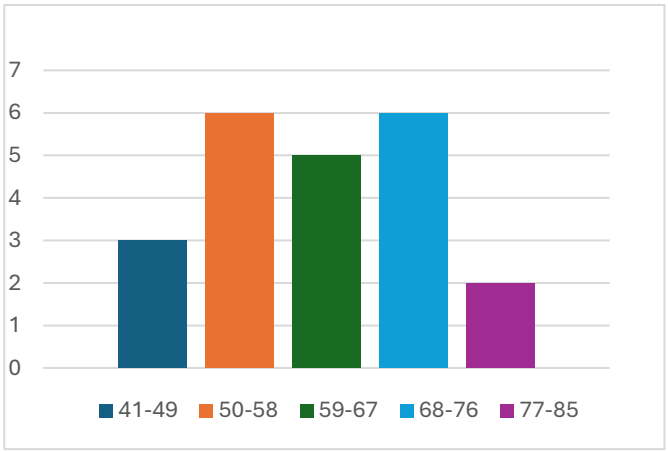
- Prikupljanje podataka: prvo prikupite podatke za koje želite izraditi distribuciju frekvencija.
- Identificirajte različite vrijednosti: identificirajte različite vrijednosti koje se pojavljuju u vašim podacima. Ovo su kategorije ili diskretne vrijednosti koje želite analizirati.
- Brojanje pojavljivanja: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

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

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

Slika 1.22 Tablica distribucije frekvencija.

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



Slika 1.23 Grafikon distribucije frekvencija.

1.6 Deskriptivna i inferencijalna statistika

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

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

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

Inferencijalna statistika

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

Glavni ciljevi inferencijalne statistike su:

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

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

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

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

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

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

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

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

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

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

1.7 Korelacija i regresija

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

Korelacija

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

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

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

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

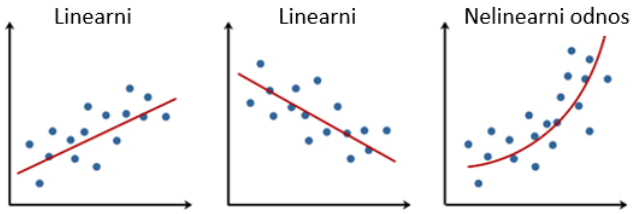
Primjer: korelacija između broja sati učenja i ocjena koje studenti postižu bit će pozitivna ako povećanje broja sati učenja obično odgovara višim ocjenama.

Regresija

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

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

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



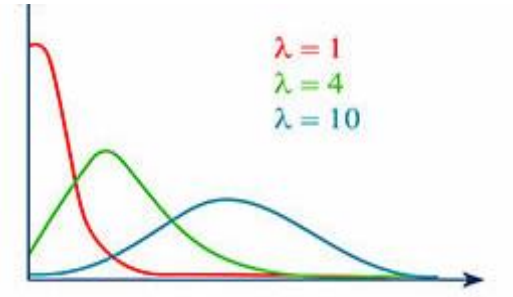
Slika 1.24 Grafikon jednostavne linearne regresije.

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

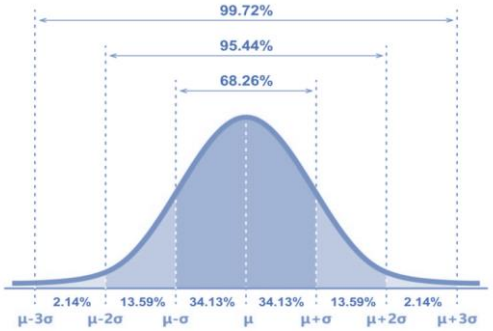
1.8 Distribucija vjerojatnosti

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

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



Slika 1.26 Grafikon Poissonove distribucije.



Slika 1.25 Grafikon normalne distribucije.

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

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

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

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

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

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

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

Literatura 1. poglavlja

- *Introductory Statistics.* Bentham Science Publishers, Kahl, A. (Publish 2023). DOI:10.2174/9789815123135123010l
- Introductory Statistics 2e, Openstax, Rice University, Houston, Texas 77005, Jun 23, senior contributing authors: Barbara Illowsky and Susan dean, De anza college, Publish Date: Dec 13, 2023, (https://openstax.org/details/books/introductory-statistics-2e);
- Introductory Statistics 4th Edition, Susan Dean and Barbara Illowsky, Adapted by Riyanti Boyd & Natalia Casper (Published 2013 by OpenStax College) July 2021, (http://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf);
- Introductory Statistics 7th Edition, Prem S. Mann, eastern Connecticut state university with the help of Christopher Jay Lacke, Rowan university, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, 2011
- Journal of the Royal Statistical Society 2024**,** A reputable journal publishing cutting-edge research and articles on various aspects of statistics, including theoretical advancements and practical applications. Recent issues have featured studies on sampling and hypothesis testing.
- Introduction to statistics, made easy second edition, Prof. Dr. Hamid Al-Oklah Dr. Said Titi Mr. Tareq Alodat, March 2014
- Statistics for Business and Economics, Thirteenth Edition, David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran, 2017, 2015 Cengage Learning®

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

Dodatne poveznice na literaturu i Youtube videozapise 1. poglavlja

- https://open.umn.edu/opentextbooks/textbooks/196
- https://www.scribbr.com/category/statistics/
- https://stats.libretexts.org/Bookshelves/Introductory_Statistics
- https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
- https://saylordotorg.github.io/text_introductory-statistics/
- https://drive.uqu.edu.sa/_/mskhayat/files/MySubjects/20178FS%20Elementary%20Statistics/Introductory%20Statistics%20(7th%20Ed).pdf
- https://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf
- https://www.geeksforgeeks.org/introduction-of-statistics-and-its-types/
- https://onlinestatbook.com/Online_Statistics_Education.pdf
- https://www.researchgate.net/profile/Tareq-Alodat-2/publication/340511098_INTRODUCTION_TO_STATISTICS_MADE_EASY/links/5e8de3dc4585150839c7b58a/INTRODUCTION-TO-STATISTICS-MADE-EASY.pdf
- https://byjus.com/maths/statistics/
- https://www.khanacademy.org/math/statistics-probability
- https://www.youtube.com/watch?v=XZo4xyJXCak
- https://www.youtube.com/watch?v=LMSyiAJm99g
- https://www.youtube.com/watch?v=VPZD_aij8H0
- https://www.youtube.com/watch?v=TLwp5DwcqD4
- https://www.youtube.com/watch?v=fpFj1Re1l84
- https://youtube.com/playlist?list=PLqzoL9-eJTNAB5st3mtP_bmXafGSH1Dtz&si=z-IXQ1iKbw2-ieJW
- https://www.youtube.com/watch?v=44MJyNTxaP8

2. Statistika za poslovnu analitiku

Dobrodošli u svijet poslovne statistike, gdje se podaci pretvaraju u značajne uvide, usmjeravajući donošenje odluka i otkrivajući skrivene istine. U ovom sveobuhvatnom istraživanju krećemo na putovanje kako bismo demistificirali bitne statističke koncepte i tehnike koji podupiru rigoroznu analizu poslovnih podataka. Od razumijevanja zamršenosti distribucija do primjene testiranja hipoteza i konstruiranja intervala pouzdanosti, svako poglavlje otkriva novi aspekt statističke pismenosti.

U srcu statističke analize leži normalna distribucija, krivulja u obliku zvona koja prožima bezbrojne pojave u prirodi i ljudskom ponašanju. U ovom dijelu ulazimo u bit normalne distribucije, razotkrivajući njena svojstva i značaj u statističkom zaključivanju. Kroz vizualizaciju primjera iz stvarnog svijeta, rasvjetljavamo sveprisutnost ove temeljne distribucije i njezinu ulogu kao kamena temeljca statističke teorije.

Standardna devijacija služi kao kompas u statističkom krajoliku, vodeći nas kroz varijabilnost svojstvenu skupovima podataka. U ovom poglavlju rastavljamo koncept standardne devijacije, otkrivajući njezinu važnost u kvantificiranju disperzije i procjeni raspršenosti podataka. Opremljeni dubljim razumijevanjem standardnih odstupanja, kretat ćete se podacima s povjerenjem, precizno uočavajući uzorke i netipične vrijednosti.

Varijable čine građevne blokove statističke analize, a svaka posjeduje različite karakteristike i implikacije. Ovo poglavlje pojašnjava dihotomiju između kontinuiranih i diskretnih varijabli, bacajući svjetlo na njihovu ulogu u modeliranju i interpretaciji podataka. Shvaćanjem nijansi tipova varijabli, iskoristit ćete puni potencijal statističkih tehnika prilagođenih različitim strukturama podataka.

Sampling-distribucija služi kao temelj statističkog zaključivanja, premošćujući jaz između promatranja uzorka i parametara populacije. U ovom poglavlju razotkrivamo koncept sampling-distribucije, razjašnjavajući njegovu relevantnost u izradi vjerojatnosti o karakteristikama populacije. Kroz konkretne primjere razvit ćete intuitivno razumijevanje uloge sampling-distribucije uzorkovanja u robusnoj statističkoj analizi.

Centralni granični teorem je ključni koncept u statistici koji nam pomaže da shvatimo nesigurnost. Ovo poglavlje objašnjava centralni granični teorem na jednostavan način,

pokazujući kako čini prosjeke uzorka predvidljivijima i pomaže u testiranju hipoteza. Razumijevanjem ovog koncepta moći ćete izvući smislene zaključke iz podataka.

Razumijevanje testiranja hipoteza bitno je za donošenje odluka na temelju podataka. Omogućuje nam da utvrdimo jesu li uočeni obrasci u podacima smisleni ili su jednostavno slučajni. Primjenom testiranja hipoteza možemo procijeniti pretpostavke, usporediti grupe i procijeniti statistički značaj rezultata, što ga čini vitalnim alatom u znanstvenom istraživanju, poslovnoj analizi i mnogim drugim područjima.

Z-standardizirana vrijednost i z-tablice služe kao navigacijska pomoć u moru standardne normalne distribucije, olakšavajući standardizirane usporedbe i izračune vjerojatnosti. Ovo poglavlje pojašnjava zamršenost z-standardiziranih vrijednosti, osnažujući vas da tumačite standardizirane rezultate i koristite Z-tablice za statističku analizu. Uz vještinu o z-standardiziranim vrijednostima, kretat ćete se ogromnim prostranstvom normalne distribucije s povjerenjem i preciznošću.

U situacijama kada su veličine uzorka male ili su standardne devijacije populacije nepoznate, t-rezultati i t-tablice pojavljuju se kao nezamjenjivi alati za statističku analizu. Ovo poglavlje razotkriva misterije t-rezultata, vodeći vas kroz njihov izračun i tumačenje pomoću t-tablica. Naoružani ovim znanjem, lako će te se snalaziti u nijansama t-distribucija, osiguravajući zaključivanje u različitim statističkim scenarijima.

Normalna i t-distribucija predstavljaju stupove teorije vjerojatnosti, a svaka posjeduje jedinstvene karakteristike i primjene. U ovom poglavlju razjašnjavamo razlike između ovih distribucija, omogućujući vam da razlučite kada svaku od njih upotrijebiti u statističkoj analizi. Kroz praktične primjere i komparativne analize, razvit ćete razumijevanje normalne i t-distribucije, obogaćujući svoj skup statističkih alata.

Intervali pouzdanosti pružaju uvid u neizvjesnost oko parametara populacije, omogućujući nam da kvantificiramo preciznost naših procjena. U ovom poglavlju istražujemo konstrukciju intervala pouzdanosti za srednje vrijednosti i proporcija, razotkrivajući metodologiju i tumačenje iza ovih bitnih statističkih alata. Savladavanjem intervala pouzdanosti, transparentno i kritički ćete prenijeti neizvjesnost koja je svojstvena vašim nalazima.

Dok p-vrijednosti nude pristup statističkim zaključivanjima, njihovo pogrešno tumačenje može dovesti do pogrešnih zaključaka i pogrešno informiranih odluka. Ovo poglavlje ispituje

potencijalne zamke pretjeranog oslanjanja na p-vrijednosti, naglašavajući važnost konteksta i veličine učinka u statističkoj analizi. Kroz kritičko ispitivanje i praktične uvide, pažljivo ćete se kretati kroz složenost p-vrijednosti, osiguravajući integritet svojih statističkih zaključaka.

Unutar ovih stranica leže ključevi za otključavanje misterija statističke analize, što vam omogućuje da pouzdano i precizno upravljate složenošću podataka. Dok zajedno krećemo na ovo putovanje, neka nam znatiželja bude kompas, a istraživanje naše svjetlo vodilja, osvjetljavajući put prema dubljem razumijevanju i djelotvornim uvidima.

2.1 Normalna distribucija

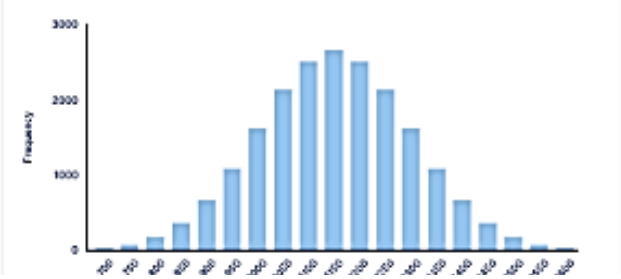
U središtu statističke analize nalazi se normalna distribucija, sveprisutna distribucija vjerojatnosti koja služi kao mjerilo za mnoge statističke tehnike. Udubit ćemo se u njezine karakteristike, njezinu simetričnu krivulju u obliku zvona i značaj u razumijevanju distribucije podataka.

Normalna distribucija nalazi primjenu u raznim područjima, uključujući financije, psihologiju, inženjerstvo i biologiju. Od modeliranja cijena dionica do razumijevanja distribucije ljudske visine, normalna distribucija služi kao svestran alat za analizu i tumačenje podataka.

Kroz ovo poglavlje zadubit ćemo se u matematička svojstva normalne distribucije, istražujući kako izračunati vjerojatnosti, percentile i z-središnje vrijednosti. Raspravljat ćemo o praktičnim tehnikama za vizualizaciju i interpretaciju normalnih distribucija pomoću histograma, dijagrama gustoće i funkcija kumulativne distribucije.

Do kraja ovog poglavlja duboko ćete cijeniti normalnu distribuciju i njen značaj u statističkoj analizi. Bit ćete dobro opremljeni za rješavanje naprednijih statističkih koncepata i njihovu primjenu na skupove podataka u stvarnom svijetu. Krenimo na ovo putovanje kako bismo zajedno razotkrili misterije normalne distribucije.

Normalna distribucija, također poznata kao Gaussova distribucija ili zvonolika krivulja, pokazuje simetričnu distribuciju podataka bez asimetrije. Kada su grafički prikazani, podaci tvore krivulju u obliku zvona, s većinom vrijednosti koje se skupljaju oko središta i smanjuju kako se udaljavaju od njega.

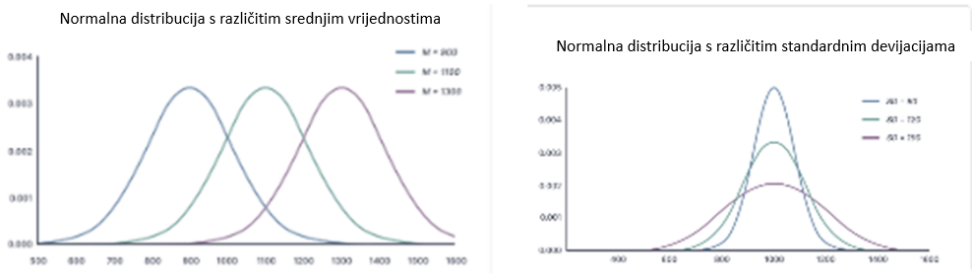


Slika 2.1 Primjer Gaussove distribucije ili zvonolike krivulje

Različite varijable u prirodnim i društvenim znanostima obično pokazuju normalnu distribuciju ili su joj blizu. Primjeri uključuju visinu, porođajnu težinu, sposobnost čitanja, zadovoljstvo poslom i SAT rezultate. Zbog učestalosti normalno raspodijeljenih varijabli, brojni statistički testovi prilagođeni su takvim populacijama. Vještina u razumijevanju karakteristika normalne distribucije osnažuje pojedince da koriste inferencijalnu statistiku za usporedbu grupa i generiranje procjena populacije iz uzoraka.

Normalne distribucije imaju ključne karakteristike koje je lako uočiti na grafikonima:

- Srednja vrijednost, medijan i mod su potpuno isti.
- Distribucija je simetrična u odnosu na srednju vrijednost - polovica vrijednosti nalazi se ispod, a polovica iznad srednje vrijednosti.
- Distribucija se može opisati s dvije vrijednosti: srednjom vrijednošću i standardnom devijacijom.



Slika 2.28 Normalna distribucija s različitim srednjim vrijednostima i različitim standardnim devijacijama.

Srednja vrijednost služi kao lokacijski parametar koji diktira središte vrha krivulje. Podešavanje srednje vrijednosti pomiče krivulju u skladu s tim: povećanje pomiče krivulju udesno, dok smanjenje pomiče krivulju ulijevo. U međuvremenu, standardna devijacija funkcionira kao parametar razmjera, utječući na širenje ili širinu krivulje.

Standardna devijacija rasteže ili stišće krivulju. Mala standardna devijacija rezultira uskom krivuljom, dok velika standardna devijacija dovodi do široke krivulje.

2.2 Empirijsko pravilo

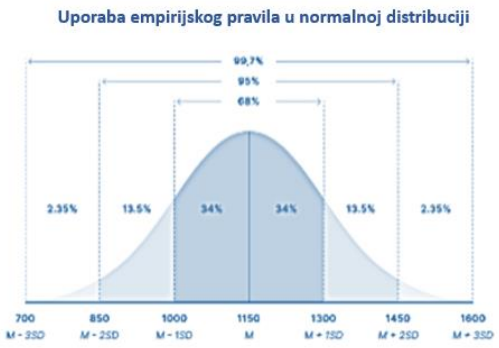
Empirijsko pravilo, također poznato kao pravilo 68-95-99.7, daje uvid u distribuciju vrijednosti unutar normalne distribucije:

- Otprilike 68% vrijednosti pada unutar 1 standardne devijacije od srednje vrijednosti.
- Otprilike 95% vrijednosti nalazi se unutar 2 standardne devijacije od srednje vrijednosti.
- Oko 99,7% vrijednosti obuhvaćeno je unutar 3 standardne devijacije od srednje vrijednosti.

Na primjer, razmotrite scenarij u kojem se prikupljaju rezultati SAT-a od učenika u novom tečaju pripreme za ispit, a podaci su u skladu s normalnom distribucijom sa srednjom ocjenom (M) od 1150 i standardnom devijacijom (SD) od 150.

Primjena empirijskog pravila daje sljedeće uvide:

- Oko 68% rezultata spada u raspon od 1000 do 1300, što odgovara 1 standardnoj devijaciji iznad i ispod prosjeka.
- Otprilike 95% rezultata je unutar raspona od 850 do 1450, što predstavlja 2 standardne devijacije iznad i ispod prosjeka.
- Gotovo svi rezultati, oko 99,7%, leže u rasponu od 700 do 1600, obuhvaćajući 3 standardne devijacije iznad i ispod prosjeka.



Slika 2.54 Empirijsko pravilo u normalnoj distribuciji.

Empirijsko pravilo nudi brzu metodu za procjenu podataka, omogućujući otkrivanje outliera ili netipičnih vrijednosti koje odstupaju od očekivanog obrasca. U slučajevima kada podaci iz malih uzoraka značajno odstupaju od ovog obrasca, alternativne distribucije kao što je t-distribucija mogu biti prikladnije. Identificiranje distribucije varijable omogućuje primjenu relevantnih statističkih testova.

2.3 Formula normalne krivulje

Za konstruiranje normalne krivulje na temelju dane srednje vrijednosti i standardne devijacije, može se upotrijebiti funkcija gustoće vjerojatnosti, čime se točno predstavlja distribucija podataka.

Slika 2.74 Normalna krivulja prilagođena podacima SAT rezultata.

Unutar funkcije gustoće vjerojatnosti, područje ispod krivulje predstavlja vjerojatnost. S obzirom da normalna distribucija služi kao distribucija vjerojatnosti, kumulativna površina ispod krivulje uvijek iznosi 1 ili 100%. Iako se formula za normalnu funkciju gustoće vjerojatnosti može činiti zamršenom, njeno korištenje samo zahtijeva poznavanje srednje vrijednosti populacije i standardne devijacije. Zamjenom ovih parametara u formulu, može se odrediti gustoća vjerojatnosti povezana s bilo kojom danom vrijednošću x.

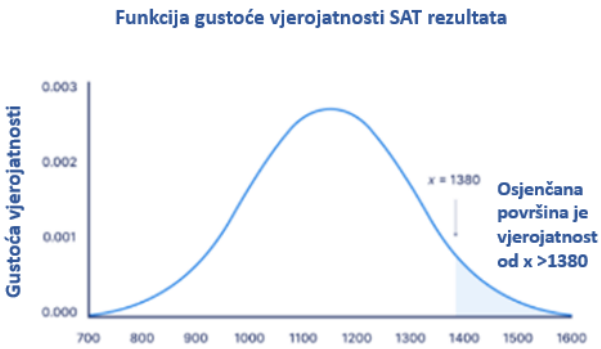
- $f(x)$ = vjerojatnost
- x = vrijednost varijable
- μ = srednja vrijednost
- σ = standardna devijacija
- σ^2 = varijanca

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Primjer:

Koristeći funkciju gustoće vjerojatnosti, želite znati vjerojatnost da SAT rezultati u vašem uzorku premašuju 1380.

Na vašem grafikonu funkcije gustoće vjerojatnosti, vjerojatnost je osjenčano područje ispod krivulje koja se nalazi desno od mjesta gdje je vaš SAT rezultat jednak 1380.



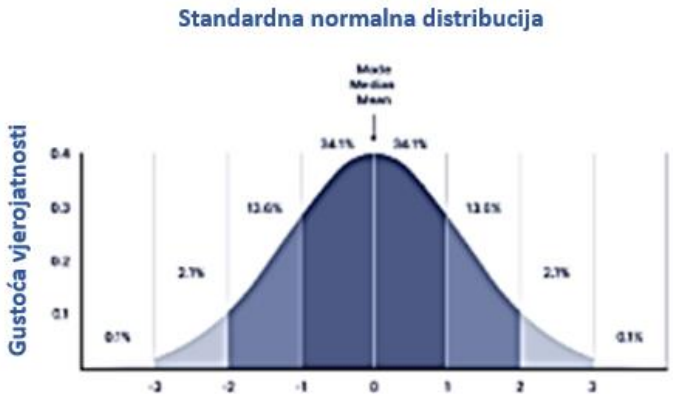
Slika 2.100 Grafikon funkcije gustoće vjerojatnosti SAT rezultata.

Vrijednost vjerojatnosti ovog rezultata možete pronaći pomoću standardne normalne distribucije.

2.4 Standardna normalna distribucija

Standardna normalna distribucija, poznata kao **z-distribucija** , razlikuje se po tome što ima srednju vrijednost od 0 i standardnu devijaciju od 1. Svaka normalna distribucija može se promatrati kao transformacija standardne normalne distribucije, koja prolazi kroz prilagodbe u mjerilu, položaju ili oba.

U kontekstu z-distribucije, pojedinačna opažanja, koja se obično označavaju kao x u normalnim distribucijama, nazivaju se z-standardizirane vrijednosti ili z-skorovi. Ovi z-skorovi predstavljaju broj standardnih devijacija za koje svaka vrijednost odstupa od srednje vrijednosti. Posljedično, pretvaranje vrijednosti iz bilo koje normalne distribucije u z-skorove olakšava usporedbu i analizu unutar okvira standardne normalne distribucije.



Slika 2.120 Grafikon standardne normalne distribucije.

Trebate znati samo srednju vrijednost i standardnu devijaciju vaše distribucije da biste pronašli z-skor vrijednosti.

Objašnjenje formule z-skora

$$z = \frac{x - \mu}{\sigma}$$

- x = pojedinačna vrijednost
- μ = srednja vrijednost
- σ = standardna devijacija

Normalne distribucije pretvaramo u standardnu normalnu distribuciju iz nekoliko razloga:

- kako bismo pronašli vjerojatnost opažanja u distribuciji koja pada iznad ili ispod zadane vrijednosti;
- kako bismo pronašli vjerojatnost da se srednja vrijednost uzorka značajno razlikuje od poznate srednje vrijednosti populacije.
- za usporedbu rezultata na različitim distribucijama s različitim srednjim vrijednostima i standardnim odstupanjima.

2.5 Određivanje vjerojatnosti korištenjem z-distribucije

Svaki z-rezultat odgovara vjerojatnosti, koja se često naziva p-vrijednost, koja ukazuje na vjerojatnost opažanja vrijednosti ispod tog specifičnog z-skora. Transformacijom pojedinačne

vrijednosti u z-skor, može se odrediti vjerojatnost da se sve vrijednosti do te točke pojave unutar normalne distribucije.

Na primjer, razmotrite scenarij u kojem želite utvrditi vjerojatnost da će SAT rezultati u vašem uzorku premašiti 1380. U početku izračunavate z-skor koristeći srednju vrijednost i standardnu devijaciju distribucije. Uz srednju vrijednost od 1150 i standardnu devijaciju od 150, z-skor otkriva broj standardnih devijacija za koje 1380 odstupa od srednje vrijednosti.

Izračun formule

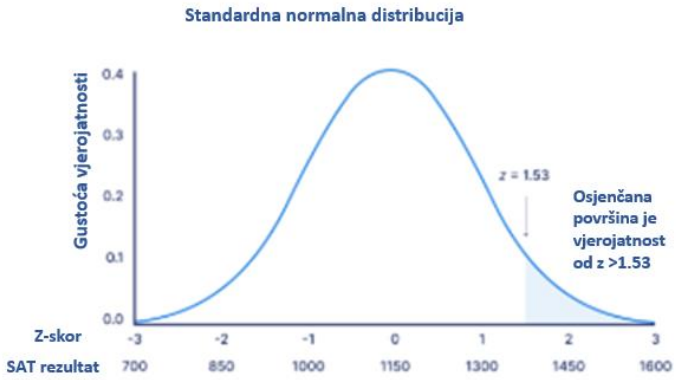
$$z = \frac{x - \mu}{\sigma} = \frac{1380 - 1150}{150} = 1.53$$

Za *z* -skor od 1,53, *p* -vrijednost je 0,937. Ovo je vjerojatnost da će SAT rezultati biti 1380 ili manje (93,7%), a to je područje ispod krivulje lijevo od osjenčanog područja.

Da biste pronašli osjenčano područje, oduzmite 0,937 od 1, što je ukupna površina ispod krivulje.

Vjerojatnost x > 1380 = 1 – 0,937 = **0,063**

To znači da je vjerojatno da samo 6,3% SAT rezultata u vašem uzorku prelazi 1380.



Slika 2.128 Standardna normalna distribucija s naznačenim SAT

2.6 Sampling-distribucija

Sampling-distribucije čine okosnicu statističkog zaključivanja, omogućujući nam izvođenje zaključaka o populacijama na temelju podataka uzorka. Udubit ćemo se u zamršenost

sampling-distribucija, razumijevajući kako odražavaju varijabilnost statistike uzorka i njihovu ključnu ulogu u testiranju hipoteza.

Sampling-distribucija odnosi se na distribuciju statističkih podataka, kao što je srednja vrijednost uzorka ili proporcija uzorka, dobivenih iz više uzoraka iste veličine izvučenih iz populacije. Pruža uvid u ponašanje statistike uzorka i njihovu varijabilnost u različitim uzorcima.

2.7 Centralni granični teorem i sampling-distribucija

Centralni granični teorem (engl. Central Limit Theorem - CLT) temeljni je koncept u statistici koji podupire ponašanje sampling-distribucije. Navodi se da se sampling-distribucija srednje vrijednosti uzorka približava normalnoj distribuciji kako se veličina uzorka povećava, bez obzira na oblik distribucije populacije. Ovaj teorem nam omogućuje da napravimo čvrste zaključke o parametrima populacije iz uzorka podataka.

Središnji granični teorem služi kao kamen temeljac razumijevanja normalnih distribucija u statistici. U uvjetima istraživanja, dobivanje točne procjene srednje vrijednosti populacije često uključuje prikupljanje podataka iz brojnih nasumičnih uzoraka unutar populacije. Te pojedinačne srednje vrijednosti uzoraka zajedno tvore ono što je poznato kao sampling-distribucija srednje vrijednosti.

Centralni granični teorem ocrtava dva ključna principa:

1. **Zakon velikih brojeva:** kako se veličina uzorka ili broj uzoraka povećava, srednja vrijednost uzorka nastoji se približiti srednjoj vrijednosti populacije.
2. **Normalnost sampling-distribucije:** usprkos izvornoj distribuciji varijable, kada se radi s višestrukim velikim uzorcima, sampling-distribucija srednje vrijednosti teži približnoj normalnoj distribuciji.

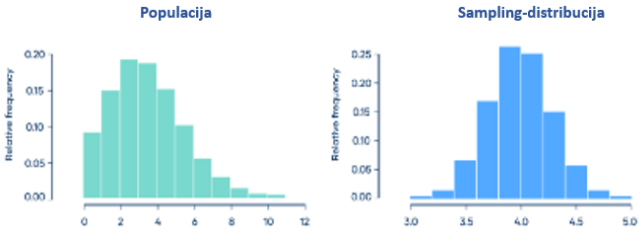
Parametarski statistički testovi konvencionalno pretpostavljaju da su uzorci izvedeni iz normalno distribuiranih populacija. Međutim, centralni granični teorem uklanja nužnost ove pretpostavke za dovoljno velike uzorke. S velikim uzorcima, parametarski testovi mogu se primijeniti bez obzira na distribuciju populacije, pod uvjetom da su zadovoljene druge relevantne pretpostavke. Veličina uzorka od 30 ili više obično se smatra dovoljno velikom.

Nasuprot tome, za male uzorke, osiguravanje pretpostavke normalnosti je ključno zbog nesigurnosti koja okružuje sampling-distribuciju srednje vrijednosti. Točni rezultati zahtijevaju

potvrdu da se populacija pridržava normalne distribucije prije korištenja parametarskih testova s malim uzorcima.

Ilustrativno, centralni granični teorem tvrdi da će dobivanjem dovoljno velikih uzoraka iz populacije, srednje vrijednosti tih uzoraka pokazati normalnu distribuciju, čak i ako temeljna distribucija populacije odstupa od normalnosti.

Primjer: Razmotrite populaciju prema Poissonovoj distribuciji (prikazano na lijevoj slici). Nakon izvlačenja 10 000 uzoraka iz ove populacije, od kojih se svaki sastoji od 50 opažanja, distribucija srednjih vrijednosti uzorka blisko je usklađena s normalnom distribucijom, u skladu s centralnim graničnim teoremom (kao što je ilustrirano na desnoj slici).



Slika 2.155 Primjer populacije u Poissonovoj distribuciji i normalnoj distribuciji.

Centralni granični teorem ovisi o pojmu sampling-distribucije, koja predstavlja distribuciju vjerojatnosti statistike izračunate iz brojnih uzoraka izvučenih iz populacije.

Konceptualizacija eksperimenta može pomoći u shvaćanju sampling-distribucije:

- Zamislimo izvlačenje slučajnog uzorka iz populacije i izračunavanje statistike, kao što je srednja vrijednost.
- Nakon toga se izvlači još jedan nasumični uzorak identične veličine, a srednja vrijednost se ponovno izračunava.
- Ovaj se proces ponavlja mnogo puta, što rezultira mnoštvom srednjih vrijednosti, od kojih svaka odgovara uzorku.

Združivanje ovih srednjih vrijednosti uzoraka predstavlja primjer sampling-distribucije. Prema centralnom graničnom teoremu, sampling-distribucija srednje vrijednosti teži prema normalnoj distribuciji kada je veličina uzorka dovoljno velika. Nevjerojatno, bez obzira na distribuciju

populacije - bila ona normalna, Poissonova, binomna ili neka druga – sampling-distribucija srednje vrijednosti pokazuje normalnost.

Srećom, ne treba opetovano uzorkovati populaciju da bi se razaznao oblik sampling-distribucije. Umjesto toga, parametri sampling-distribucije srednje vrijednosti ovise o parametrima same populacije.

- Srednja vrijednost sampling-distribucije je srednja vrijednost populacije.

  $\mu_{\bar{x}} = \mu$

- Standardna devijacija sampling-distribucije je standardna devijacija populacije podijeljena s kvadratnim korijenom veličine uzorka.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Sampling-distribuciju srednje vrijednosti možemo opisati pomoću ove oznake:

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

gdje:

- $\bar{X}$ je sampling-distribucija srednjih vrijednosti uzorka
- $\sim$ znači "slijedi distribuciju"
- N je normalna distribucija
- μ je srednja vrijednost populacije
- σ je standardna devijacija populacije
- n je veličina uzorka.

Veličina uzorka, označena kao n, predstavlja broj opažanja izvučenih iz populacije za svaki uzorak, održavajući ujednačenost u svim uzorcima. Veličina uzorka značajno utječe na sampling-distribuciju srednje vrijednosti u dva ključna aspekta.

1. Veličina uzorka i normalnost:

- Veći uzorci obično daju sampling-distribucije koje su bliske normalnoj distribuciji.

- Suprotno tome, s malim uzorcima, sampling-distribucija srednje vrijednosti može odstupati od normalnosti. Ovo odstupanje nastaje jer valjanost centralnog graničnog teorema ovisi o "dovoljno velikoj" veličini uzorka.
- Uobičajeno, veličina uzorka od 30 ili više smatra se "dovoljno velikim".
- Kada je n<30, centralni granični teorem se ne primjenjuje, a sampling-distribucija odražava distribuciju populacije. Stoga je sampling-distribucija normalna samo ako je distribucija populacije normalna.
- Nasuprot tome, kada je n≥30, centralni granični teorem vrijedi, a sampling-distribucija približava se normalnoj distribuciji.

2. Veličina uzorka i standardna devijacija:

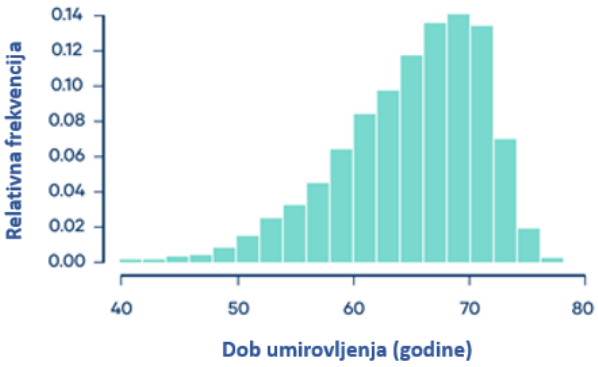
- Veličina uzorka izravno utječe na standardnu devijaciju sampling-distribucije, odražavajući varijabilnost ili raspršenost distribucije.
- S manjim uzorcima, standardna devijacija obično je viša, što ukazuje na veću varijabilnost među srednjim vrijednostima uzorka zbog njihove neprecizne procjene srednje vrijednosti populacije.
- Suprotno tome, veći uzorci odgovaraju nižim standardnim devijacijama, što ukazuje na manju varijabilnost među srednjim vrijednostima uzorka zahvaljujući njihovoj točnijoj procjeni srednje vrijednosti populacije.

Važnost centralnog graničnog teorema:

Parametarski testovi kao što su t-testovi, ANOVA i linearna regresija imaju veću statističku snagu u usporedbi s većinom neparametarskih testova. Ova povećana statistička snaga proizlazi iz pretpostavki o distribuciji populacija, koje su utemeljene na centralnom graničnom teoremu.

Kontinuirana distribucija

Razmotrimo dob za odlazak u mirovinu pojedinaca u Sjedinjenim Američkim Državama. Stanovništvo se sastoji od svih umirovljenih Amerikanaca, a distribucija ovog stanovništva može se predstaviti na sljedeći način:



Slika 2.182 Grafikon kontinuirane

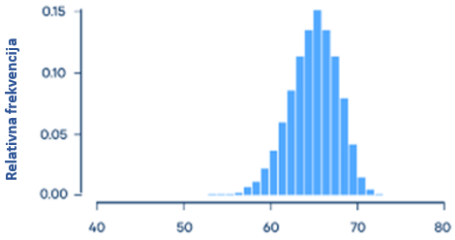
Distribucija dobi za umirovljenje iskrivljena je ulijevo, pri čemu većina odlazi u mirovinu unutar približno pet godina od prosječne dobi za umirovljenje od 65 godina. Međutim, postoji prošireni rep pojedinaca koji odlaze u mirovinu puno ranije, primjerice s 50 ili čak 40 godina. Populacija pokazuje standardnu devijaciju od 6 godina.

Zamislite provođenje malog uzorkovanja ove populacije. Nasumično se odabire pet umirovljenika i bilježi se njihova dob za odlazak u mirovinu. Na primjer: 68, 73, 70, 62, 63.

Srednja vrijednost ovog uzorka služi kao procjena srednje vrijednosti populacije, iako s ograničenom preciznošću zbog male veličine uzorka od 5. Na primjer: Srednja vrijednost = (68 + 73 + 70 + 62 + 63) / = 67,2 godine

Sada pretpostavimo da se ovaj proces uzorkovanja ponovi 10 puta, a svaki uzorak uključuje pet umirovljenika. Izračunava se srednja vrijednost svakog uzorka, što rezultira distribucijom poznatom kao sampling-distribucija srednje vrijednosti. Na primjer: 60,8, 57,8, 62,2, 68,6, 67,4, 67,8, 68,3, 65,6, 66,5, 62,1

Budući da se ovaj proces ponavlja mnogo puta, histogram koji prikazuje srednje vrijednosti ovih uzoraka približno će odgovarati normalnoj distribuciji.



Slika 2.208 Normalna distribucija srednjih vrijednosti

Unatoč tome što sampling-distribucija pokazuje nešto normalniji oblik u usporedbi s populacijom, još uvijek zadržava blagi zaokret ulijevo. Osim toga, evidentno je da je varijabilnost u sampling-distribuciji uža od one populacije.

Prema centralnom graničnom teoremu, sampling-distribucija srednje vrijednosti nastoji se približiti normalnoj distribuciji kako se veličina uzorka povećava. Međutim, trenutna sampling-distribucija srednje vrijednosti odstupa od normalne zbog relativno male veličine uzorka.

2.8 Testna statistika

Testna statistika predstavlja brojčanu vrijednost izvedenu iz testiranja statističke hipoteze koja ukazuje na stupanj usklađenosti između vaših opaženih podataka i distribucije očekivane prema nultoj hipotezi tog testa.

Ova statistika igra ključnu ulogu u izračunavanju p-vrijednosti vaših nalaza, olakšavajući odluku o prihvaćanju ili odbacivanju vaše nulte hipoteze.

Ali što točno čini testnu statistiku?

Testna statistika artikulira sličnost između distribucije vaših podataka i distribucije predviđene prema nultoj hipotezi korištenog statističkog testa. Distribucija podataka razjašnjava učestalost svakog opažanja, koju karakterizira centralna tendencija i varijabilnost oko nje. Budući da različiti statistički testovi predviđaju različite vrste distribucije, odabir odgovarajućeg testa usklađen je s vašom hipotezom.

Testna statistika sažima vaše opažene podatke u jedinstvenu brojku, koristeći mjere kao što su centralna tendencija, varijabilnost, veličina uzorka i broj varijabli predviđanja u vašem statističkom modelu.

Tipično, testna statistika proizlazi iz vidljivih obrazaca u vašim podacima (npr. korelacije između varijabli ili odstupanja među grupama), podijeljenih s varijancom podataka (tj. standardnom devijacijom).

Razmotrite ovaj primjer:

Istražujete povezanost između temperature i datuma cvjetanja kod određene vrste stabla jabuke. Analizirajući opsežan skup podataka koji obuhvaća 25 godina, prateći temperaturu i datume cvjetanja nasumičnim uzorkovanjem 100 stabala godišnje s eksperimentalnog polja.

• Nulta hipoteza (H_0): Ne postoji korelacija između temperature i datuma cvjetanja.

• Alternativna hipoteza (H_A ili H_1): Postoji korelacija između temperature i datuma cvjetanja.

Da biste ispitali ovu hipotezu, provodite regresijski test, dajući t-vrijednost kao testnu statistiku. Ova t-vrijednost suprotstavlja uočenu korelaciju između varijabli naspram nulte hipoteze koja pretpostavlja da nema korelacije.

2.9 Vrste testne statistike

U nastavku je prikazan sinopsis prevladavajućih testnih statistika, zajedno s njihovim odgovarajućim hipotezama i kategorijama statističkih testova u kojima se koriste. Iako različiti statistički testovi mogu koristiti različite metodologije za izračunavanje ovih statistika, temeljne hipoteze i tumačenja testne statistike ostaju dosljedni.

Testna statistika	Nulta i alternativna hipoteza	Statistički testovi
***t* vrijednost**	**Nulta:** Srednje vrijednosti dviju grupa su jednake. **Alternativna:** Srednje vrijednosti dviju skupina nisu jednake.	• *T* test • Regresijski testovi
***z* vrijednost**	**Nulta:** Srednje vrijednosti dviju grupa su jednake. **Alternativna:** Srednje vrijednosti dviju skupina nisu jednake.	• *Z* test
***F* vrijednost**	**Nulta:** Varijacija između dvije ili više grupa veća je ili jednaka varijaciji između grupa. **Alternativna:** Varijacije između dvije ili više grupa su manje od varijacija između grupa.	• ANOVA • ANCOVA • MANOVA
***X*²-vrijednost**	**Nulta:** Dva su uzorka neovisna. **Alternativna:** Dva uzorka nisu neovisna (tj. koreliraju).	• Hi-kvadrat test • Neparametarski korelacijski testovi

U scenarijima iz stvarnog svijeta, obično ćete izračunati svoju testnu statistiku koristeći statistički softverski paket kao što je R, SPSS ili Excel, koji će također dati p-vrijednost povezanu sa testnom statistikom. Unatoč tome, formule za ručno izračunavanje ovih statistika mogu se pronaći na internetu.

Na primjer, u testiranju vaše hipoteze o temperaturi i datumima cvjetanja, provodite regresijsku analizu. Regresijski test daje: • regresijski koeficijent od 0,36 • t-vrijednost koja uspoređuje ovaj koeficijent s očekivanim rasponom regresijskih koeficijenata pod nultom hipotezom nepostojanja veze.

Rezultirajuća t-vrijednost iz regresijskog testa od 2,36 predstavlja vašu testnu statistiku.

2.10 Standardna pogreška

Standardna pogreška srednje vrijednosti (engl. *standard error of the mean* - SE ili SEM) služi kao pokazatelj vjerojatne razlike između srednje vrijednosti populacije i srednje vrijednosti uzorka. Nudi uvid u stupanj varijabilnosti koji bi se očekivao u srednjoj vrijednosti uzorka ako bi se studija replicirala koristeći svježe uzorke izvučene iz iste populacije.

Dok je standardna pogreška srednje vrijednosti najčešće citirani oblik standardne pogreške, slične mjere postoje za druge statističke parametre kao što su medijan ili proporcije. Standardna pogreška funkcionira kao prevladavajuća mjera pogreške uzorkovanja, prikazujući nejednakost između parametra populacije i statistike uzorka.

Kako bi se ublažila standardna pogreška, preporučuje se povećanje veličine uzorka. Korištenje velikog, nasumičnog uzorka služi kao najučinkovitija strategija za smanjenje pristranosti uzorkovanja i povećanje pouzdanosti nalaza.

Standardna pogreška i standardna devijacija mjere su varijabilnosti:

- **Standardna devijacija** opisuje varijabilnost **unutar jednog uzorka**.
- **Standardna pogreška** procjenjuje varijabilnost **u višestrukim uzorcima** populacije.

Standardna devijacija služi kao deskriptivna statistika izvedena izravno iz podataka uzorka, dok standardna pogreška predstavlja inferencijalnu statistiku, obično procijenjenu, osim ako nije poznat točan parametar populacije.

2.11 Formula standardne pogreške

Standardna pogreška srednje vrijednosti određena je primjenom standardne devijacije uz veličinu uzorka. Kroz formulu postaje očito da su veličina uzorka i standardna pogreška u obrnutom odnosu. Jednostavnije rečeno, kako se veličina uzorka povećava, standardna pogreška se smanjuje. Do ovog fenomena dolazi jer veći uzorak ima tendenciju dati statističke podatke uzorka bliže parametru populacije.

Koriste se različite formule na temelju toga je li poznata standardna devijacija populacije. Ove formule su primjenjive na uzorke koji sadrže više od 20 elemenata (n > 20).

Kada su poznati parametri populacije

Kada je poznata standardna devijacija populacije, možete je koristiti u donjoj formuli za točan izračun standardne pogreške.

Formula	Obrazloženje
$SE = \frac{\sigma}{\sqrt{n}}$	• *SE* je standardna pogreška • σ je standardna devijacija populacije • n je broj elemenata u uzorku

Kada su parametri populacije nepoznati

Kada je standardna devijacija populacije nepoznata, možete koristiti donju formulu samo za procjenu standardne pogreške. Ova formula uzima standardnu devijaciju uzorka kao procjenu standardne devijacije populacije.

Formula	Obrazloženje
$SE = \frac{s}{\sqrt{n}}$	• SE je standardna greška • s je standardna devijacija uzorka • n je broj elemenata u uzorku

Primjer: Korištenje formule standardne pogreške za procjenu standardne pogreške za rezultate SAT-a iz matematike. Slijedite sljedeća dva koraka.

Najprije pronađite kvadratni korijen veličine uzorka (*n*).

Formula	Izračun
$n = 200$	$\sqrt{n} = \sqrt{200} = 14.1$

Zatim podijelite standardnu devijaciju uzorka s brojem koji ste pronašli u prvom koraku.

Formula	Izračun		
$SE = \frac{s}{\sqrt{n}}$	$s = 180$	$\sqrt{n} = 14.1$	$\frac{s}{\sqrt{n}} = \frac{180}{14.1} = 12.8$

Standardna pogreška rezultata SAT iz matematike je 12,8.

Možete predstaviti standardnu pogrešku uz srednju vrijednost ili je uključiti u interval pouzdanosti kako biste prenijeli nesigurnost koja okružuje srednju vrijednost.

Na primjer: Prikaz srednje vrijednosti i standardne pogreške. Srednji rezultat SAT-a iz matematike za slučajni uzorak ispitanika je 550 ± 12,8 (SE).

Izvještavanje o standardnoj pogrešci unutar intervala pouzdanosti je poželjno jer eliminira potrebu čitatelja za izvođenje dodatnih izračuna kako bi dobili smisleni raspon.

Interval pouzdanosti označava raspon vrijednosti gdje se očekuje da će nepoznati parametar populacije najčešće biti ako bi se studija ponovila s novim slučajnim uzorcima.

Na razini pouzdanosti od 95%, očekuje se da će 95% svih srednjih vrijednosti uzorka pasti unutar intervala pouzdanosti koji obuhvaća ± 1,96 standardnih pogrešaka srednje vrijednosti uzorka. Ovaj interval služi kao procjena unutar koje se vjeruje da se stvarni parametar populacije nalazi unutar 95% pouzdanosti.

Na primjer: Konstruiranje intervala pouzdanosti od 95% Vi konstruirate interval pouzdanosti od 95% (CI) da biste procijenili srednju vrijednost matematičke SAT ocjene populacije. S obzirom na normalno raspodijeljenu karakteristiku kao što su SAT rezultati, otprilike 95% svih srednjih vrijednosti uzorka pada unutar približno 4 standardne pogreške srednje vrijednosti uzorka.

Formula intervala pouzdanosti

$CI = \bar{x} \pm (1{,}96 \times SE)$

$\bar{x}$ = srednja vrijednost uzorka = 550

SE = standardna pogreška = 12,8

Donja granica	**Gornja granica**
$\bar{x} - (1{,}96 \times SE)$	$\bar{x} + (1{,}96 \times SE)$
550 − (1,96 × 12,8) = **525**	550 + (1,96 × 12,8) = **575**

S nasumičnim uzorkovanjem, 95% CI [525 575] govori vam da postoji vjerojatnost od 0,95 da je srednja vrijednost matematičkog SAT rezultata populacije između 525 i 575.

Literatura 2. poglavlja

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

- Statistics for Business and Economics, Thirteenth Edition, David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran, 2017, 2015 Cengage Learning®
- Statistics for Business, First edition, Derek L Waller, 2008 Copyright © 2008, Derek L Waller, Published by Elsevier Inc. All rights reserved

Dodatne poveznice na literaturu i Youtube videozapise 2. poglavlja

- https://open.umn.edu/opentextbooks/textbooks/196
- https://www.scribbr.com/category/statistics/
- https://stats.libretexts.org/Bookshelves/Introductory_Statistics
- https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
- https://saylordotorg.github.io/text_introductory-statistics/
- https://drive.uqu.edu.sa/_/mskhayat/files/MySubjects/20178FS%20Elementary%20Statistics/Introductory%20Statistics%20(7th%20Ed).pdf
- https://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf
- https://www.geeksforgeeks.org/introduction-of-statistics-and-its-types/
- https://onlinestatbook.com/Online_Statistics_Education.pdf
- https://www.researchgate.net/profile/Tareq-Alodat-2/publication/340511098_INTRODUCTION_TO_STATISTICS_MADE_EASY/links/5e8de3dc4585150839c7b58a/INTRODUCTION-TO-STATISTICS-MADE-EASY.pdf
- https://byjus.com/maths/statistics/
- https://www.khanacademy.org/math/statistics-probability

3. Upravljanje podacima

U B2B elektroničkoj razmjeni podataka (engl. *Electronic Data Interchange* - EDI), poruke koje sadrže kodove proizvoda ili usluga, identifikaciju transportnih jedinica, kao i pravne dokumente razmjenjuju se među partnerima u opskrbnom lancu. Obično imaju oblik alfanumeričkih nizova.

3.1 Informacije-Podaci-Znanje

U računalnim informacijskim sustavima prikaz informacija varira ovisno o njihovoj namjeni i upotrebi. Može biti alfanumerički ili binarni s obzirom na to predstavlja li tekst, sliku, zvuk ili izvršni program. Da bi mogao pohraniti, dohvatiti, obraditi i prenijeti podatke različitih vrsta (npr. brojeva, znakova, datuma, valuta itd.).) moraju biti ispravno kodirani. Alfanumerički formati predstavljanja podataka potječu iz ASCII (ask-key) abecede, nakon što su se razvili u smislu nacionalnih skupova znakova (npr. standardi 8859-1, Latin 1 i 8859-2, Latin 2) i konačno napredovali u međunarodne UTF-8 i UTF-16 formati. Stoga omogućuju zajedničko tumačenje podataka od strane poslovnih partnera koji pripadaju različitim etničkim skupinama i geografskim sredinama. Dok alfanumerički nizovi ovise o njihovom kodiranju, numerički se podaci uglavnom razlikuju po veličini i/ili preciznosti.

Proces pretvaranja podataka iz izvornog analognog u digitalni oblik popularno se naziva digitalizacija. Odgovarajuće dizajnirani pomoćni i aplikacijski programi koji prihvaćaju podatke iz različitih izvora (npr. optički skeneri, električni senzori, EDI, itd.) omogućuju organizacijama da automatiziraju svoje prikupljanje podataka, pohranjivanje, obradu i prijenos unutar i između svojih računalnih informacijskih sustava.

Kada se prikupe, podaci različitih vrsta mogu se spojiti i organizirati u tablice podataka, baze podataka, skladišta podataka (kronološki redoslijed) ili baze znanja (konceptualni redoslijed). Više razine organizacije podataka omogućuju automatizirano razvrstavanje, zaključivanje i predstavljanje time akumuliranog znanja u analitičke svrhe.

3.2 Logistički podaci

U logistici se EDI koristi za prijenos transakcijskih podataka između poslovnih partnera. Budući da mogu koristiti različite jezike i aplikacije, potrebna je njihova konverzija u zajednički format (npr. XML, JSON) kako bi ih mogli interpretirati različiti informacijski sustavi partnera (W3schools, 2023). Za brzu identifikaciju i manipulaciju osmišljeni su bar kodovi i RFID oznake.

Kako bi se omogućila međunarodna suradnja, trebalo je definirati globalno prihvaćene formate podataka za potrebe logistike. Formati logističkih podataka odgovaraju oznakama usluga i proizvoda, identifikaciji transportnih jedinica i kodovima transakcija, obično u obliku alfanumeričkih nizova s vremenskim žigom. Radi jednostavnosti rukovanja i brzine obrade, ovi su kodovi standardizirani i kodirani kao optički čitljivi bar-kodovi ili elektromagnetski čitljivi radiofrekvencijski identifikacijski (RFID) kodovi.

Bar-kod (EAN/UCC) je višesektorski i međunarodni oblik numeriranja stavki (POS EAN-8 i EAN-13, promjenjivi EAN-128, podatkovna traka, pakiranje ITF-14, QR, podatkovna matrica itd.) . Koriste se za identifikaciju proizvoda, serija proizvoda ili pošiljaka (1D kodovi) kao i usluga (2D kodovi). Među 2D bar kodovima QR kod je najprisutniji, čitljiv i pametnim telefonima, što povećava njihovu upotrebljivost u različitim područjima primjene.

RFID kodovi se prvenstveno koriste na isti način kao bar kodovi. Oni jedinstveno identificiraju artikle ili usluge. Obično, osim izbornog crtičnog koda, RFID naljepnice nose još više informacija na čipu veličine glave pribadače. Osim identifikacije, RFID oznake omogućuju bilježenje podataka o praćenju, što je često potrebno u logističkim aplikacijama. Za razliku od crtičnih kodova, RFID omogućuje njihovo skeniranje bez izravnog vidnog polja, kao i skeniranje više naljepnica odjednom.

GS1 standardom EPC Gen2 (ISO/IEC 18000-6:2013) uspostavljen je tehnološki standard koji određuje komunikaciju između RFID tagova i čitača. Slično bar kodovima, EPCglobal standardi povezuju RFID tehnologiju s EPC (engl. *Electronic Product Code*) označavanjem proizvoda, logističkih transportnih jedinica, lokacija, zaliha, povratnih artikala, dokumenata itd. za izravnu, automatiziranu identifikaciju i praćenje logističkih jedinica unutar opskrbnog lanca.

EPCglobal standardi također predstavljaju osnovu GDSN-a (engl. *Global Data Synchronization Network*). Omogućuje automatizirano prikupljanje i razmjenu specifikacijskih podataka o

proizvodima i njihovoj ambalaži, čime se poduzećima omogućuje centralizirano upravljanje tim podacima kako bi ih oni i njihovi partneri naizmjenično koristili.

Tablica 3.1 sažima različite identifikacijske tehnologije s njihovim primjenama. Otkriva niz jednodimenzionalnih i dvodimenzionalnih bar kodova, kao i različite klase RFID kodova s njihovim mogućnostima.

Tablica 3.1Tehnologije označavanja.

Tehnologija	**Primjena**
Bar kod 1D	Maloprodajni artikli i komponente proizvoda
Bar kod 2D	Usluge (npr. UPS, zrakoplovne karte), veleprodajni artikli koji zahtijevaju praćenje
RFID klasa 1 (pasivno, R-oznake)	Predmeti koji zahtijevaju masovnu identifikaciju, kontrolu pristupa
RFID klasa 2 (pasivno, RW-tagovi)	Stavke koje zahtijevaju praćenje
RFID klasa 3 (poluaktivan, RW-tagovi)	Kontrola pristupa s dodanim informacijama za praćenje
RFID klasa 4 (aktivan, RW-tagovi)	Trasiranje i praćenje zatvorenog prostora
RFID klasa 5 (aktivne oznake/ispitivači)	Praćenje otvorenog prostora, blizina usluge s omogućenim uređajima, usluge temeljene na lokaciji

Budući trendovi u označavanju, praćenju i sljedivosti slijede dva glavna smjera: minijaturizacija i raznolikost. Bar kodovi (1D) također moraju omogućiti označavanje minijaturnih predmeta (npr. medicinskih kapsula). Novi kodovi podatkovne matrice (2D) ne samo da će omogućiti ispravljanje pogrešaka tijekom skeniranja, već i šifriranje podataka.

RFID se nastavlja širiti na druga područja upotrebe kao što je identifikacija od strane pružatelja usluga (npr. željeznička kartica, prijava na posao, itd.), beskontaktna plaćanja (npr. bežični

prijenos novca, plaćanja na automatima) i pametna rješenja (npr. upravljanje pametnim domom, daljinski upravljani pametni uređaji itd.) kao i e-valute.

3.3 Organizacija podataka

Osim što imaju određeni format, podaci se mogu organizirati na različite načine kako bi se olakšalo njihovo upravljanje, obrada i prezentacija. Iako su podaci na svom ulazu uglavnom nestrukturirani, njihovim pohranjivanjem, prijenosom i obradom povećava se njihova organiziranost. U nastavku su prikazani uobičajeni oblici organizacije podataka po rastućoj složenosti od polustrukturiranih (npr. CSV) do strukturiranih (npr. proračunske tablice, baze podataka, itd.) formata.

Proračunske tablice

Prvi oblik organizacije podataka su dvodimenzionalni nizovi polja, koji se također nazivaju tablice ili proračunske tablice. Obično prvi redak proračunske tablice označava značenja vrijednosti pohranjenih u temeljnim stupcima, nakon čega slijede redovi podataka.

Polje ili ćelija tablice je najmanja jedinica podataka. Ima određeni tip (broj, datum, valuta itd.). Njegov sadržaj se može adresirati oznakama retka i stupca (npr. A1, koji predstavlja prvi redak stupca A).

Svaki redak tablice je grupa povezanih polja, koja predstavljaju zapis (npr.: transakcija, studentski zapis, podaci o proizvodu itd.). Budući da svi reci tablice imaju istu strukturu, tip zapisa možemo definirati kao popis atributa (npr. podaci o studentu (ime, prezime, datum rođenja, mjesto rođenja, ID...)) odgovarajućih tipova podataka.

Baze podataka

Datoteka ili tablica baze podataka zbirka je zapisa iste vrste. Baza podataka (engl. *Data Base* - DB) sastoji se od više međusobno povezanih tablica. Dakle, ANSI definicija baze podataka:

- Podaci baze podataka su međusobno povezani i sortirani
- Bazu podataka može istovremeno koristiti više korisnika
- Podaci u bazi podataka se ne ponavljaju
- Baza podataka je pohranjena u računalu

Iz gornje definicije mogu se izvući neki zaključci o klijentsko-poslužiteljskoj arhitekturi gdje poslužitelj drži DB, kojem pristupaju njegovi klijenti. Naravno, za pristup DB-u mora biti uspostavljena komunikacijska mreža između poslužitelja i njegovih klijenata. DB poslužitelj se obično naziva njegov "*back-end*", dok klijenti predstavljaju njegov "*front-end*". Sustav za upravljanje bazom podataka (engl. *Data Base Management System* - DBMS) na poslužitelju omogućuje svojim klijentima pristup podacima pohranjenim u DB-u putem svog aplikacijskog programskog sučelja (API) i DBM funkcija. DBM funkcije su mehanizmi koji omogućuju unos, dohvaćanje, obradu i prezentaciju podataka u DB-u. Za pozivanje ovih funkcija definirani su standardni jezici upita (engl. *Standard Query Languages* - SQL).

Model relacijske baze podataka

Postoje različiti oblici organizacije DB-a, a najčešći je relacijski model (RDB). Osnovna ideja iza ovog modela je činjenica da korisnik ne može unaprijed znati sve moguće upotrebe podataka pohranjenih u DB-u. Budući da obično ne postoje fiksni putevi pretraživanja kroz datoteke baze podataka, osmišljeni su različiti upitni jezici za dohvaćanje i manipulaciju podacima. RDB model temelji se na konceptu entiteta i odnosa:

- Entitet je osoba/stvar/koncept koji se može jedinstveno identificirati i ima atribute.
- Relacija predstavlja način povezivanja dva ili više entiteta.

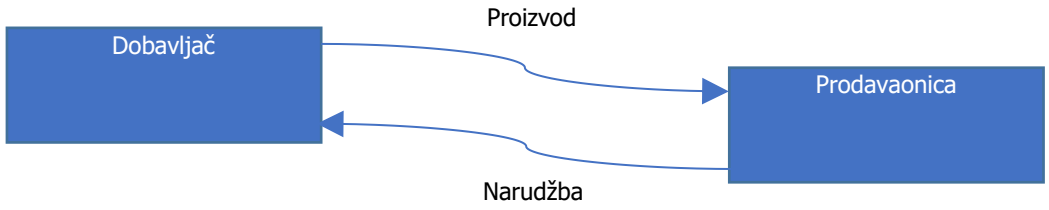
RDB tablice, koje predstavljaju entitete ili relacije, međusobno su povezane pomoću ključeva. Skup atributa koji jedinstveno identificiraju entitet naziva se njegov primarni ključ. Kada se primarni ključ pojavljuje kao polje u drugoj tablici s ciljem ispunjavanja relacije s izvornom tablicom, naziva se sekundarnim ili stranim ključem. Dok tablica može sadržavati samo jedan primarni ključ za jedinstvenu identifikaciju svojih zapisa, može sadržavati više sekundarnih ključeva.

Općenito, postoje dva pristupa konstruiranju RDB-a: analitički i sintetički.

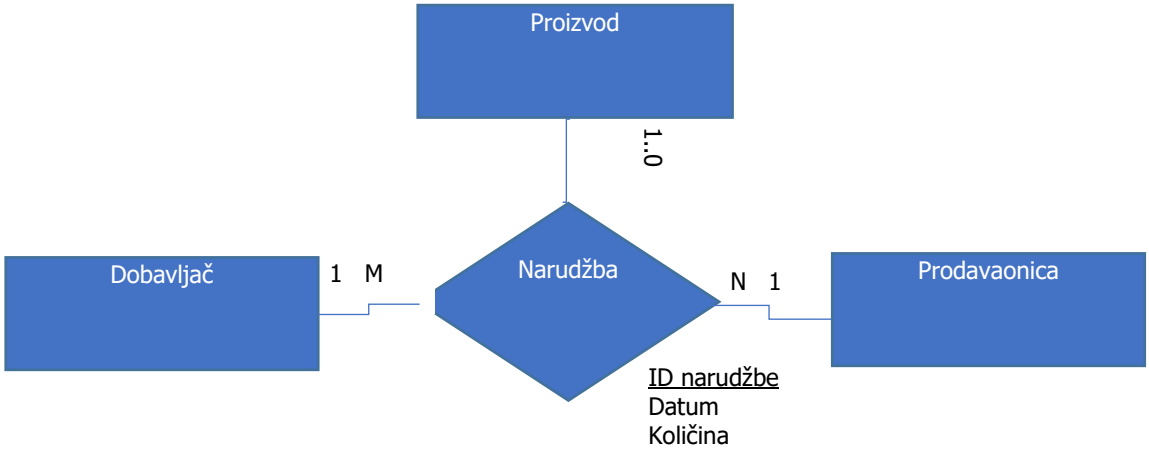
Analitički pristup izradi RDB-a sastoji se od sljedeća četiri koraka:

1. Analiza stvarnog svijeta - globalni model
2. Određivanje entiteta i odnosa – konceptualni model (npr. E-R dijagram)
3. Određivanje logičkog modela – relacijska shema
4. Izgradnja baze podataka (DBMS) – fizički model

Kako bismo ilustrirali pristup, razmotrimo primjer trgovačkog lanca i njegovih dobavljača (slika 3.1). Svaka prodavaonica ima više dobavljača. Svaki dobavljač može opskrbljivati različite prodavaonice. Ciklus nadopune započinje narudžbom iz prodavaonice. Zauzvrat dobavljač isporučuje robu u prodavaonicu. Narudžbe su transakcije u kojima se spajaju podaci o prodavaonici, dobavljaču i isporučenoj robi (slika 3.2).



Slika 3.1 Globalni model.



Slika 3.13 Konceptualni model.

DOBAVLJAČ (Dobavljač_ID#, Dobavljač_naziv, Dobavljač_kontakt)
PRODAVAONICA (Prodavaonica_ID#, Prodavaonica_naziv, Prodavaonica_adresa)
NARUDŽBA (Dobavljač_ID#, Prodavaonica_ID#, Narudžba_ID#, Datum, Količina, EPC)
PROIZVOD (EPC#, Proizvod_naziv, Proizvod_cijena)

Slika 3.14Logički model.

Logički model predstavlja RDB tablice prema vrstama njihovih zapisa. U logičkom modelu (slika 3.3) određeni atributi sadrže simbol ljestve (engl. *hashtag*) (#). To znači da polje koje oni predstavljaju jest ili pripada (kompozitnom) primarnom ključu. Neka su polja podcrtana.

Nazivaju se sekundarnim ili stranim ključevima, budući da referenciraju primarne ključeve povezanih tablica.

Sintetički pristup RDB-u sastoji se od sljedeća tri koraka:

1. Analiza podataka – popis svih relevantnih atributa
2. Određivanje logičkog modela normalizacijom – relacijska shema
3. Fizički model (DBMS)

Normalizacija ili kanonička sinteza (Kent, 1983) osigurava da se obrnutim inženjeringom iz relevantnih atributa formira DB koji ispunjava uvjete RDB-a. Dok početni normalni oblik (0NF) atributa predstavlja tablicu neuređenih atributa, naknadni normalni oblici, kako ih definira (Codd, 1970), predstavljaju više razine organizacije podataka. Može se tvrditi da je dostizanjem 3. normalne forme postignuta shema koja ispunjava zahtjeve za RDB logički model.

Tablica je u 1NF, ako predstavlja relaciju. Time je osigurano da se sve grupe podataka koje se ponavljaju čuvaju odvojeno i stoga se ne ponavljaju.

Tablica u 2NF je u 1NF. Dodatno, nijedan atribut ključa ne smije djelomično funkcionalno ovisiti o primarnom ključu. Ovime se svi ključevi koji jedinstveno identificiraju određene atribute čuvaju odvojeno. Ovo je uglavnom kako bi se osiguralo da se samo atributi koji ovise o svim (dijelovima) primarnog ključa čuvaju u jednoj tablici.

Tablica u 3NF je u 2NF. Osim toga, nijedan atribut koji nije ključ nije tranzitivno ovisan o primarnom ključu. To znači da se svi ne-ključni atributi koji mogu predstavljati ključ za određene druge ne-ključne atribute čuvaju u zasebnoj tablici, pri čemu se samo njihov ključ održava u izvornoj tablici kao strani ključ.

Slijedeći korake normalizacije od 1NF do 3NF, završava se s logičkim modelom koji odgovara propisima relacijske baze podataka (RDB). Viši oblici normalizacije uglavnom su za optimizaciju RDB modela.

Tablica u Boyce-Codd NF (BCNF) je u 3NF; dodatno, svaka odrednica je ključ. Ovo uklanja sve odnose koji nisu obuhvaćeni postojećim redoslijedom ključeva iz izvorne tablice, čime se formiraju dodatne tablice za svaki ključ kandidata. Tablica u 4NF je u 3NF i BCNF. Osim toga, svaki atribut s više vrijednosti koji djelomično ovisi o ključu nalazi se u vlastitoj tablici. 4NF je namijenjen uklanjanju svih mogućih preostalih atributa s više vrijednosti iz originalne tablice.

Tablica je 5NF je u 4NF; dodatno, svaka JOIN operacija je predviđena ključevima. Tablica u 6NF je u 5NF; dodatno, uzimaju se u obzir sve netrivijalne JOIN-ovisnosti.

Može se uočiti da kod viših oblika organizacije broj tablica raste sa svakim korakom. Stoga je razumno promatrati fragmentaciju podataka kako bi se spriječilo stvaranje nepotrebnih tablica kojima se rijetko pristupa.

Tablica 3.2Primjer normalizacije RBD-a.

<table>
<tr><td>ONF

NARUDŽBA

• Dobavljač_ID *
• Ime_dobavljača
• Dobavljač_kontakt
• Prodavaonica_ID *
• Naziv_prodavaonice
• Adresa_prodavaonice
• ID_narudžbe*
• Datum
• EPC
• Količina
• Naziv_proizvoda
• Cijena_proizvoda</td><td>1NF

NARUDŽBA

• Dobavljač_ID *
• Ime_dobavljača
• Dobavljač_kontakt
• Prodavaonica_ID *
• Naziv_prodavaonice
• Adresa_prodavaonice
• ID_narudžbe*
• Datum
• EPC
• Količina
• Naziv_proizvoda
• Cijena_proizvoda</td></tr>
<tr><td colspan="2">* Ključevi kandidata za ponavljajuću grupu atributa, koji jedinstveno identificiraju narudžbu</td></tr>
<tr><td>2NF

DOBAVLJAČ

• Dobavljač _ID #
• Ime_dobavljača
• Dobavljač_kontakt

PRODAVAONICA

• Prodavaonica _ID #
• Naziv_prodavaonice
• Adresa_prodavaonice</td><td>

NARUDŽBA

• Narudžba_ID#
• Dobavljač_ID#
• Prodavaonica _ID #
• EPC
• Naziv_proizvoda
• Cijena_proizvoda
• Datum
• Količina</td></tr>
<tr><td>3NF

NARUDŽBA

• Dobavljač _ID #
• Prodavaonica _ID #
• Narudžba_ID#</td><td>

PROIZVOD

• EPC #
• Naziv_proizvoda</td></tr>
</table>

• Datum • Količina • ID_proizvoda	• Cijena_proizvoda

Upitni jezici

Uglavnom su dvije vrste: Structured Query Language (SQL) i Query by Example (QBE). Dok se SQL smatra programskim jezikom i API-jem DBMS-a, QBE se uglavnom koristi s DBMS-om izravno za upravljanje bazom podataka i skladištenje podataka.

Standardni SQL (ISO/IEC 9075, 1986-2016) je programski jezik četvrte generacije za manipulaciju bazom podataka. Omogućuje pretraživanje, dodavanje, mijenjanje kao i brisanje zapisa podataka. Unatoč njegovoj standardizaciji, postoje male razlike u njegovoj implementaciji s različitim sustavima za upravljanje bazama podataka (DBMS).

U nastavku je jezik ukratko predstavljen s najčešćim opcijama. Prema konvenciji, SQL ključne riječi pišu se velikim slovima i svaka rečenica završava točkom sa zarezom. Rečenice su predstavljene s poveznicama na povezane referentne materijale koji nude dodatne informacije.

Svaka manipulacija bazom podataka počinje njezinim stvaranjem. Rečenica

CREATE DATABASE *baza podataka_naziv* ;

stvara novu praznu bazu podataka s navedenim imenom.

Kao što je gore objašnjeno, podaci unutar baza podataka organizirani su u tablice zapisa podataka određenog tipa gdje svi redovi dijele zajedničku strukturu. Za izradu tablice koristi se sljedeća rečenica:

CREATE TABLE *tablica_naziv* (
stupac1 tip1,
stupac2 tip2, stupac3 tip3,
….);

Svaki imenovani stupac predstavlja atribut s određenim tipom podataka. Na primjer, u:

CREATE TABLE *Prodavaonica* (Prodavaonica _ *ID* int NOT NULL PRIMARY KEY,…);

CREATE TABLE *Narudžba* (Narudžba_ *ID int* NOT NULL PRIMARY KEY, …, Product_Id int FOREIGN KEY REFERENCES *Proizvod* (*EPC*));

kreiraju se dvije tablice. Prva sadrži podatke o kupcima, dok druga sadrži podatke o njihovim narudžbama, pozivajući se na prvu tablicu preko broja kupca kao stranog ključa.

Dok su podaci u tablici već sortirani po primarnom ključu, mogu se dodatno sortirati po drugim atributima, pod uvjetom da su indeksirani. Možemo ga indeksirati stvaranjem indeksa na danom atributu(ima) sljedećom rečenicom:

CREATE INDEX *indeksa_naziv* ON *tablica_naziv* (*naziv_stupca*);

Svaka manipulacija podacima na indeksiranoj tablici traje malo dulje, budući da za njezinu konzistentnost ne samo da treba provjeriti podatke koje daju ključevi i pravilno poredati podatke, već i druge atribute iz navedenog indeksa.

Najčešća operacija u bazi podataka je upit podataka omogućen naredbom SELECT :

SELECT *stupac1, stupac2* , …
FROM *tablica_ naziv*;

Ovaj podatkovni upit vraća podatke u stupcu1, stupcu2 itd. iz tablice. Rečenice upita obično se formiraju pružanjem dodatnih opcija, filtriranjem podataka, ispunjavanjem navedenih uvjeta:

WHERE navodi uvjet koji određuje kriterije odabira zapisa.

GROUP BY spaja zapise, imajući zajedničko svojstvo za omogućavanje skupnih funkcija.

HAVING specificira agregatne funkcije na grupama definiranim naredbama GROUP BY.

ORDER BY specificira atribute prema kojima su povratni zapisi poredani.

Na primjer:

```
SELECT "Prodavaonica" . " Prodavaonica_naziv " , "Proizvod" . " Proizvod_naziv " , "
Narudžba" . "Količina " FROM "Narudžba" , "Proizvod" , "Dobavljač" , " Prodavaonica "
WHERE "Narudžba" . " ID_proizvoda " = " Proizvod" . "EPC " AND "Narudžba" . "
Dobavljač_ID " = "Dobavljač" . "Dobavljač_ID" AND "Narudžba" . " Prodavaonica _ID" =
"Prodavaonica"."Store_ID" ORDER BY "Prodavaonica"."Prodavaonica_naziv" ASC
```

vraća popis prodavaonica s njihovim naručenim proizvodima i količinama, poredanih prema nazivu prodavaonice.

Najvažnija operacija u procesu selekcije je operacija JOIN. Često zamjenjuje uvjet WHERE kao JOIN ON, nakon čega slijedi uvjet. Uspoređuje vrijednosti stupaca i na temelju usporedbe određuje treba li ih uključiti u rezultat ili ne. U LEFT JOIN zapis se vraća ako su kriteriji ispunjeni u lijevoj tablici i obrnuto u operaciji RIGHT JOIN. Kao što je gore navedeno, uvjet mora biti ispunjen u obje tablice kako bi bio u skladu s operacijom INNER JOIN ili FULL JOIN. Budući da se potonji najčešće koristi , može se koristiti JOIN kao sinonim. Pozivajući se na uvjete 5 i 6 normalne forme, ovo je ista JOIN operacija, koju je potrebno ispuniti da bi se ispunili uvjeti odgovarajućeg NF-a.

Za unos novih podataka u tablicu koristi se operacija INSERT INTO:

INSERT INTO *tablica_naziv* (stupac*1* , [stupac*2* , ...])
VALUES (*vrijednost1* , [*vrijednost2* , ...]);

Da bi bile uspješne, vrijednosti u operaciji trebaju ispuniti sve uvjete atributa označenih nazivima stupaca. Ne treba navesti nazive stupaca u slučaju da su sve vrijednosti navedene. U slučaju da su u tablici predviđene neke DEFAULT vrijednosti, ne treba ih navoditi, osim ako se razlikuju.

Kada se podaci unesu, mogu se modificirati naredbom UPDATE :

UPDATE *tablica_naziv*
SET *stupac1=vrijednost1, stupac2=vrijednost 2* ,...
WHERE *neki_stupac = neka_vrijednost*;

U izjavi su dane nove vrijednosti za polja u navedenim stupcima. Kriteriji odabira retka označeni su specifikatorom WHERE, koji određuje sve vrijednosti stupca na koje se odnosi naredba UPDATE. Kako bi se spriječile neželjene promjene, potreban je dodatni oprez pri formuliranju kriterija odabira.

Zapis ili više zapisa može se izbrisati iz tablice operacijom DELETE :

DELETE FROM *tablica_naziv*
WHERE *neki_stupac = neka_vrijednost*;

Kao i s naredbom UPDATE, specifikator WHERE koristi se za određivanje svih redaka koje treba izbrisati.

Naravno, upravljanje bazom podataka ne završava ovdje. Svaki element baze podataka također se može ukloniti, izmijeniti i/ili zamijeniti novim. U slučaju da se indeks, tablica ili baza podataka trebaju ukloniti, mogu se primijeniti sljedeće izjave:

DROP INDEX *indeks_ime* ON *iablic_ime*;

DROP TABLE *tablica_naziv*;

DROP DATABASE *baza podataka_naziv*;

Ako netko samo želi ukloniti podatke iz tablice, može se koristiti naredba TRUNCATE:

TRUNCATE TABLE *tablica_naziv* ;

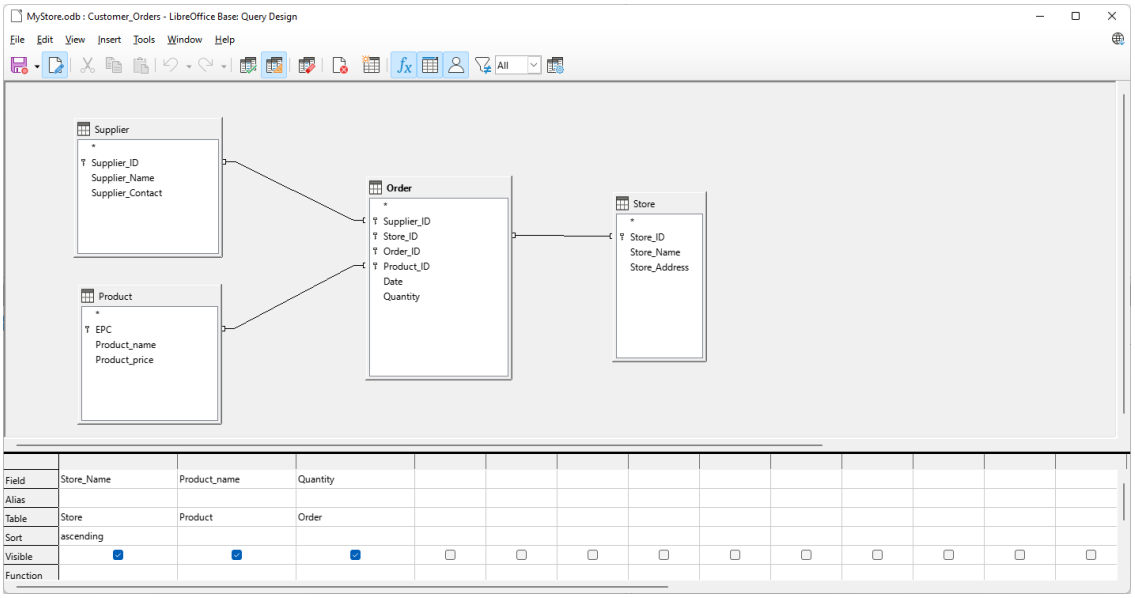
U slučaju da želite dodati ili ukloniti atribut (stupac) u/iz tablice, to možete učiniti naredbom ALTER:

ALTER TABLE *tablica_naziv* ADD stupac_*naziv vrsta podataka*;

ALTER TABLE *tablica_naziv* DROP COLUMN stupac_*naziv;*

Ovo zaključuje ovaj kratki pregled SQL jezika i njegovih najčešćih scenarija upotrebe. SQL se obično koristi u klijent-poslužitelj arhitekturama s DBMS-om koji je na glavnom računalu poslužitelja. Da bi mu se pristupilo, izdaju se SQL naredbe, bilo od strane klijentskog aplikacijskog programa ili web-sučelja DBMS poslužitelja.

S druge strane, QBE se također često koristi s relacijskim DBMS-om s grafičkim korisničkim sučeljem (engl. *Graphical User Interface* - GUI), kao što su MS Access ili LibreOffice Base. S QBE-om se baza podataka i njezine tablice stvaraju puno interaktivnije, a njihovu strukturu lakše je održavati. Kao što mu ime govori, nudi i jednostavniji oblik unosa i otkrivanja podataka. Za izvođenje pretraživanja potrebno je sastaviti sve tablice koje se koriste u pretraživanju i zatim uspostaviti uvjete kao uzorke u poljima stupaca kako bi se filtrirali relevantni podaci (Slika 3.4). Formulacija upita dopunjena je postojećim relacijama između tablica. Kao i obično, rezultat takvog upita je još jedna tablica s rezultirajućim podacima, koji se kasnije mogu dalje obraditi. Na ovaj način također se mogu formirati kaskadni ili višefazni upiti.



Slika 3.15QBE upit ekvivalentan gornjem SQL upitu.

Podatkovni filtri i maske

Kako bi se spriječili pogrešni ili nepotpuni podaci, koji bi mogli ometati njihovu obradu i tumačenje, potrebno je primijeniti dodatne mjere opreza:

1. Filtriranje praznih redaka i stupaca
2. Primjena snažnog tipiziranja podataka za sprječavanje računalnih pogrešaka
3. Definiranje maski za unos kako bi se spriječio unos pogrešnih podataka
4. Pristranost podataka kako bi se spriječili pogrešni rezultati

Prazni redovi i stupci čest su izvor pogrešaka koje uglavnom potječu od loših sučelja u aplikacijama za prikupljanje podataka. Otkazane ili nedovršene transakcije obično rezultiraju praznim redovima ili nedostajućim podacima u zapisnicima transakcija. Oni se samo djelomično mogu riješiti aplikacijama proračunskih tablica gdje se prazni reci i podaci koji nedostaju mogu otkriti provjerom ukupnog broja u odnosu na broj podataka koji nisu nula u recima/stupcima. Mogu se filtrirati uklanjanjem praznih redaka/stupaca, no to možda nije uvijek željena radnja jer bismo mogli izgubiti i neke vrijedne podatke. Najbolji način da se to spriječi je korištenjem DBMS-a koji bi s jedne strane omogućio unos samo kompletnih transakcijskih podataka, dok bi s druge strane također spriječio prazne redove/stupce, budući da ih u bazama podataka nema.

Slabo upisivanje podataka još je jedan uobičajeni izvor pogrešaka. Ako se umjesto numeričkih podataka unesu alfanumerički podaci, poput datuma ili iznosa valute, to bi rezultiralo greškama prilikom obrade tih podataka. U proračunskim tablicama, kao i u bazama podataka, pojedinačnim podatkovnim ćelijama, koje predstavljaju vrijednosti atributa u slogovima podataka, mogu se dodijeliti tipovi podataka, što bi nas alarmiralo pri unosu podataka u krivom formatu. Stoga se ovom mjerom može spriječiti da pogrešni podaci ometaju njihovu obradu.

Prilikom konstruiranja baza podataka, ograničenja se mogu primijeniti na podatkovna polja koja predstavljaju atribute entiteta ili relacije. Osim što im se može dodijeliti odgovarajuća vrsta, maske za unos mogu se definirati čime se omogućuje unos podataka, poput datuma, valuta, EAN kodova itd., samo u određenom formatu. To obično rješava mnoge pogrešne predodžbe koje bi inače mogle nastati tijekom obrade podataka.

Još jedan čest izvor pogrešaka su nepristrani podaci, koji predstavljaju podatke koji su reda veličine veći ili manji od očekivanog. Opet, mogli bi ometati našu obradu, dajući pogrešne rezultate. Teže ih je otkriti i mogu se filtrirati samo gledanjem podataka. U aplikacijama za proračunske tablice dobra uobičajena praksa bila bi određivanje minimalnih, maksimalnih i srednjih vrijednosti podataka u odgovarajućim stupcima kako bi se otkrila moguća odstupanja. Ako se otkriju, mogu se istaknuti i ručno obraditi, ako ih je malo, ili filtrirati i modificirati upitom u tablici baze podataka ako ih je puno. U svakom slučaju treba ih pažljivo procijeniti, kako se situacija ne bi još više pogoršala, a u tom slučaju bi bilo bolje ukloniti te podatke.

Skladišta podataka i baze znanja

Podaci u skladištima podataka prikupljaju se iz RDB-a i katalogiziraju kronološkim redom. Obično se na podacima provode neke poslovne analize, a analitički odjel također pohranjuje rezultate za kasnije potrebe. Nakon što se pohrane u skladište, ti se podaci obično ne mijenjaju kako bi se očuvala njihova dosljednost.

Osim kronološkog reda, prilikom izgradnje baza znanja uzimaju se u obzir i kontekstualni redovi. Ovdje su subjekti predstavljeni kao derivati entiteta najviše razine ili njegovih podružnica. Njihovi se odnosi uspostavljaju slobodnije jer su namijenjeni ažuriranju i nadogradnji kako se koriste. Uspostavljeni su u obliku pravila, temeljenih na svojstvima entiteta. Stoga je oblik u kojem su pohranjeni nešto drugačiji. Često se pohranjuju u obliku ontologija koje sadrže dublje znanje o prikupljenim podacima. Slično pohranjivanju rezultata

upita u skladištima podataka, upiti u bazama znanja također se pohranjuju za kasniju upotrebu za prikaz trenutnih rezultata, kako se mijenjaju entiteti, relacije i instance podataka.

Za razliku od baza podataka i skladišta, koji su specifični za aplikaciju, baze znanja mogu biti neovisne o aplikaciji i često ih različite aplikacije koriste među domenama. Primjer je prikazan u (Gumzej i dr., 2023).

3.4 Zaključak

U ovom poglavlju obrađeni su različiti aspekti upravljanja podacima u logistici. Osim prikaza podataka i standarda za pohranu podataka, prikazana je organizacija podataka i mehanizmi pronalaženja. Konačno, neke uobičajene pogreške u automatiziranoj obradi podataka su riješene kako bi svjesni čitatelj ostao na oprezu. Osim navedenih primjera, više se može otkriti u pridruženim materijalima za učenje.

Literatura 3. poglavlja

- Codd E.F. (1970). A relational model of data for large shared data banks. Communications. ACM 13, 6, pp. 377–387.
- Gumzej, R., Kramberger, T., Dujak, D. (2023). A knowledge base for strategic logistics planning, Proceedings of the 23rd International Scientific Conference Business Logistics in Modern Management: October 5-6, 2023, Osijek, Croatia, Dujak, Davor (ed.) Osijek: Josip Juraj Strossmayer University of Osijek, Faculty of Economics and Business, pp. 317-330. [available at: https://blmm-conference.com/past-issues/, access November 3rd, 2023]
- GS1 (2023). GS1 Standards. [available at: https://www.gs1.org/standards, access October 27th, 2023]
- Kent, W. (1983). A Simple Guide to Five Normal Forms in Relational Database Theory, Communications of the ACM, vol. 26, pp. 120-125.
- W3schools (2023). XML Tutorial [available at: https://www.w3schools.com/xml/, access November 3rd, 2023]
- W3schools (2023). JSON - Introduction [available at: https://www.w3schools.com/js/js_json_intro.asp, access November 3rd, 2023]

- W3schools (2023). Database Normalization [available at: https://www.w3schools.in/DBMS/database-normalization/, access December 7th, 2023]

4. Simulacijsko modeliranje i analiza

Simulacijskim modeliranjem i analizom (engl. *simulation modelling and analysis* - SMA) nastoje se ispuniti zahtjevi Conant–Ashby teorema (Conant i Ashby, 1970), definirajući model sustava kao dobrog regulatora, koji ima onoliko ručki, dijelova i stanja koliko i njegov izvorni fizički pandan. Time pruža mogućnost izgradnje svog digitalnog modela i uspostavljanja digitalnog laboratorija koji će omogućiti njegovo istraživanje, prilagodbu i optimizaciju. Rezultirajući simulacijski modeli su apstraktni, dinamički i u većini slučajeva stohastični, budući da su njihove varijable sustava modelirane distribucijama vjerojatnosti.

4.1 Simulacija u logistici

U logistici SMA može pružiti vrijedne inpute za optimizaciju opskrbnog lanca (engl. *supply chain* - SC) i prometne mreže (engl. *traffic network* - TN). Simulacijsko modeliranje može se koristiti za grafičku vizualizaciju vremenskih tokova kroz složene SC i TN procese i resurse, omogućujući predviđanje i kvantifikaciju mogućih ishoda iz različitih scenarija. Ovo pomaže SC i TN entitetima da steknu dragocjene uvide i razumiju učinke svojih potencijalnih odluka na SC i TN izvedbu uključujući SC vremena dostave (engl. *lead-times*), TN vrijeme putovanja i troškove. Stoga, SMA u SC i TN modeliranju može doprinijeti SC i TN analizi i poboljšanju njihovih dizajna prema postizanju veće učinkovitosti i održivosti.

Postoje mnogi aspekti SC-a, koji predstavljaju različite perspektive upravljanja opskrbnim lancem. Pogled menadžera proizvodnje na opskrbni lanac razlikuje se od pogleda menadžera marketinga, koji se opet razlikuje od pogleda menadžera nabave, itd. Dakle, korišteni modeli su različiti, čak i za istu tvrtku, a osobito za cijeli SC.

Prilikom rješavanja problema SMA u logistici, menadžeri trebaju donositi odluke na strateškoj, taktičkoj i operativnoj razini, ovisno o njihovom učinku na SC ili TN u cjelini. Zbog svoje međuovisnosti, menadžeri često nisu u stanju riješiti probleme ni na jednoj razini. Istodobno, također je teško promatrati sve tri razine iz perspektive bilo koje pojedinačne cjeline. Iz perspektive SMA, SC ili TN se mogu promatrati na dvije razine:

1. Makro razina

- samoorganizacija,
- koevolucija entiteta,
- ovisnost o vezama/transportnim rutama.

2. Mikro razina

- višestruki i heterogeni entiteti,
- lokalne interakcije među entitetima,
- strukturirani entiteti,
- adaptivni entiteti.

Iako se izvode u stvarnom vremenu, vremenski aspekt SC operacija je donekle dvosmislen. Ovisno o razini i perspektivi, trajanje operacija može se mjeriti u danima, tjednima ili čak mjesecima kada su u pitanju među-organizacijske aktivnosti, dok se s druge strane, unutar-organizacijske operacije mjere u satima ili čak sekundama. Ovisno o prirodi modeliranog problema, trajanje najkraće operacije ili maksimalna učestalost dolaznih/odlaznih zahtjeva određuje ne samo prikaz vremena u SMA modelu, već i njegovu granularnost. Što je kraće minimalno trajanje najkraće operacije ili što je veća učestalost zahtjeva, to je finija granularnost vremena, odnosno preciznost vođenja vremena u modelu. Ovo je važno za modelara, budući da vrijeme reakcije modela ne može biti kraće od unaprijed definirane vremenske granularnosti. Stoga je potrebno unaprijed procijeniti trajanje svih operacija i vremena između dolazaka dolaznih/odlaznih signala kako bi se mogle ispravno odrediti vremenske jedinice modela sustava.

U simulacijskom modelu vrijeme može napredovati kritičnim događajima od transakcije do transakcije ili kontinuirano. U potonjem slučaju, tijek vremena u modelu ne ovisi o učestalosti operacija. S protokom vremena pokrenutog kritičnim događajem, operacije se pozivaju prema vremenu njihovog pojavljivanja, odnosno kritičnih događaja. Prednost SMA je u tome što se tijekom simulacije može ubrzati tijek vremena u modelu, tako da se procesi izvode brže nego u stvarnom vremenu. Stoga se mogu napraviti rana predviđanja sljedećih događaja.

Vremena između dolaznih simulacijskih jedinica i vremena njihove obrade/tranzita mogu proizaći iz promatranja i mjerenja. Ako ne variraju, onda su deterministički. Međutim, obično

su po svojoj prirodi stohastične. Stoga, potrebno je uvođenje konstrukata koji modeliraju njihove funkcije distribucije vjerojatnosti (npr. trokutaste, uniformne, eksponencijalne, itd.).

4.2 Simulacija diskretnog događaja

Simulacija diskretnog događaja (engl. *discrete event simulation* – DES) nudi voditelju proizvodnje najdetaljniji uvid u logistički (proizvodni) proces po konzistentnom i koherentnom modelu. Stoga je DES visoko cijenjen alat za određivanje ponašanja i iskorištenja resursa u stvarnom vremenu u procesnoj industriji, uključujući logistiku.

Konstrukti:

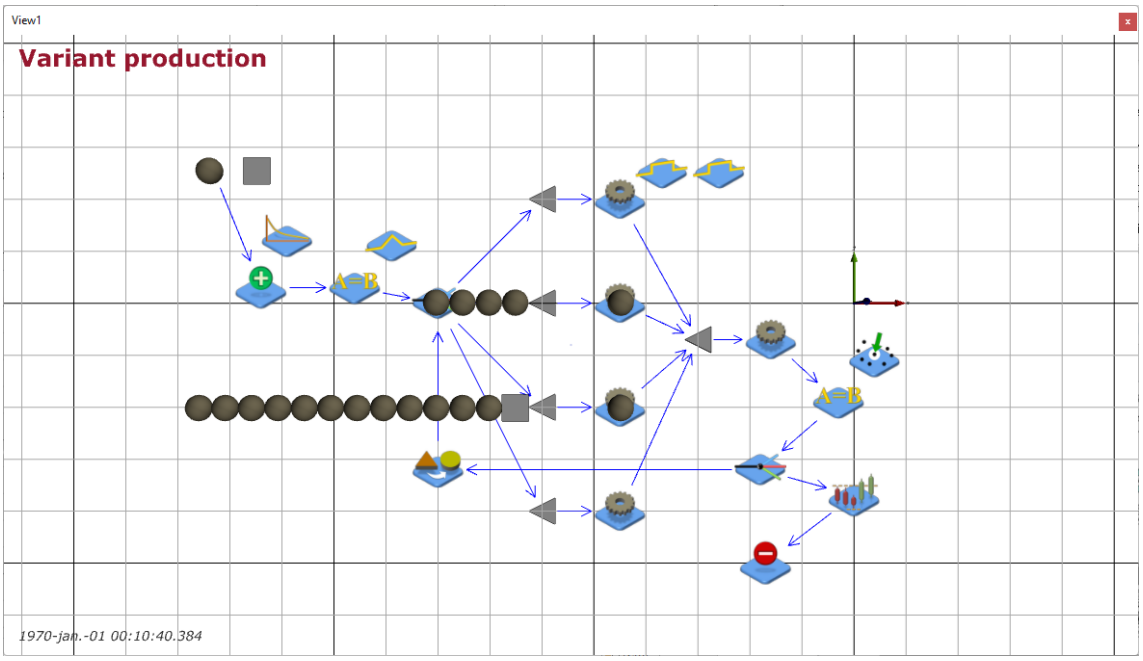
- Jedinice toka predstavljaju jedinice simulacije (npr. narudžbe, materijali itd.) koje ulaze u sustav na ulazu(ima) i napreduju kroz model sustava.
- Procesori predstavljaju mobilne (npr. ljudi, viličari itd.) i fiksne (npr. strojevi, proizvodne linije itd.) resurse koji obrađuju simulacijske jedinice.
- Redovi čekanja pohranjuju jedinice toka do njihovog prijelaza na sljedeći dostupni procesor.
- Konektori definiraju promicanje jedinica kroz model sustava.

Svojstva:

- Orijentiran na proces.
- Fokusira se na detaljno modeliranje procesa.
- Heterogeni entiteti.
- Mikro-entiteti su pasivni objekti.
- Događaji unose dinamiku u sustav.
- Diskretna vremenska progresija; od jednog (vremenskog) događaja do sljedećeg.
- Fleksibilnost se postiže promjenom strukture modela; struktura sustava tijekom simulacije je fiksna.

Primjer

Primjer DES-a (Slika 4.1, iz simulacijskog okruženja JaamSim (JaamSim Development Team, 2023)) uključuje model varijante proizvodnje, gdje se proizvode četiri različita proizvoda (Gumzej i Rakovska, 2020). Prema planu proizvodnje, proizvodi se 10, 30, 40, 20% proizvoda vrste 1, 2, 3, odnosno 4. Odabir vrste proizvoda potaknut je trokutastom raspodjelom između 1 i 4 s modulom na 3. Svaka vrsta proizvoda ima namjensku proizvodnu liniju. Proizvodni nalozi se ispunjavaju prema eksponencijalnoj distribuciji oko srednje vremenske vrijednosti od 30 s. Proizvodnja svakog pojedinog proizvoda traje 100-120s prema ravnomjernoj distribuciji. Nakon što su finalizirani, kvaliteta proizvoda se provjerava na namjenskom ispitnom mjestu. Provjera kvalitete traje 10 s. Iz iskustva tvrtke, u prosjeku svaki 1 od 10 proizvoda ne prođe kontrolu. Proizvodi nedovoljne kvalitete transportiraju se natrag na izvornu proizvodnu liniju. Njihova ponovna obrada traje 120-130s prema ravnomjernoj raspodjeli. Trajanje proizvodnje i kontrole kvalitete te ponovne obrade ne ovise o vrsti proizvoda. Nakon što su uspješno prošli kontrolu kvalitete, gotovi proizvodi se transportiraju s mjesta proizvodnje u skladište gotovih proizvoda. Ponovna proizvodnja neispravnih proizvoda dok su još u proizvodnji učinkovit je način smanjenja utjecaja na okoliš i troškova proizvodnje.



Slika 4.1Varijanta proizvodnje s kontrolom kvalitete.

Sinopsis

DES može analizirati i optimizirati sljedeće procesne parametre:

- Vrijeme proizvodnog ciklusa i učinak.
- Iskorištenost proizvodnih ćelija i prostora.
- Kapacitet skladišnih prostora kao i vrijeme zadržavanja skladišnih jedinica.
- Korištenje mobilnih resursa (npr. operateri, pokretne trake, viličari).

4.3 Sustavna dinamika

Analiza sustavna dinamika (engl. *system dynamics* – SD) predstavlja pogled menadžera SC-a na proizvodni proces pomoću dosljednog i koherentnog modela. SD se smatra alatom koji je najprikladniji za određivanje strukture, kao i optimalnih količina (za pojedinačnu lokaciju kada i koliko inputa, zaliha i outputa. Stoga, omogućuje učinkovito korištenje proizvodnih i skladišnih objekata.

Konstrukti:

- Zalihe predstavljaju tampone koji mogu pohraniti stavke isporuke u opskrbnom lancu.
- Tokovi predstavljaju opskrbne kanale.
- Petlje povratne sprege predstavljaju parametre finog podešavanja za nadopunjavanje zaliha.

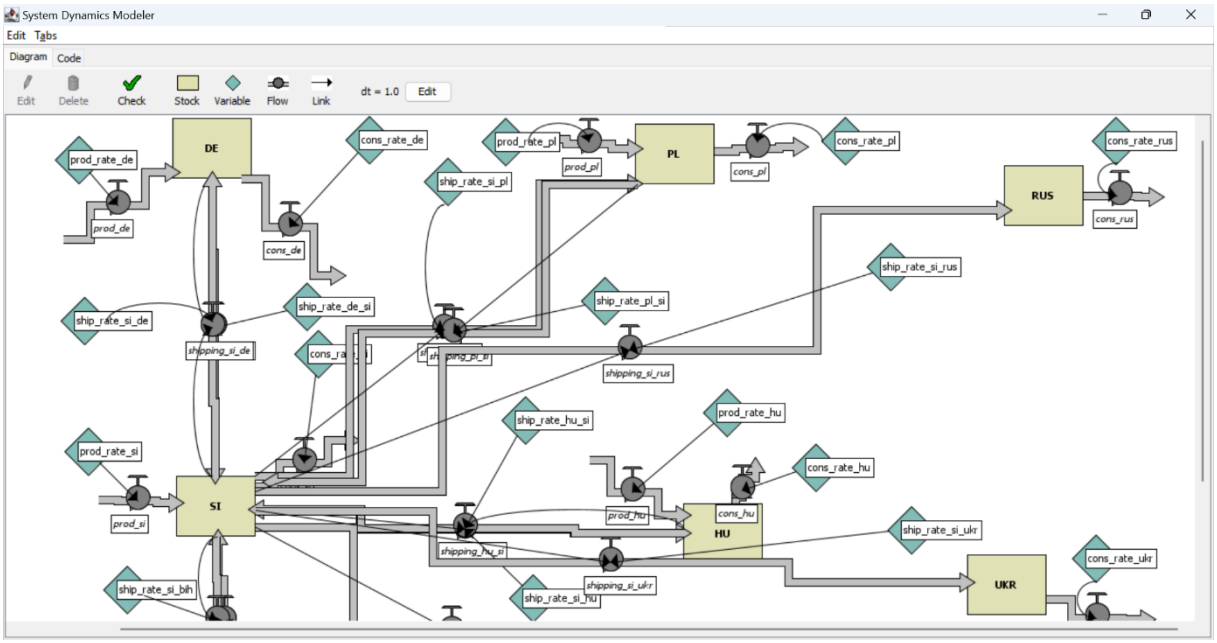
Svojstva:

- Usmjeren na sustav.
- Modeliranje varijabli sustava usmjereno na ključne pokazatelje učinka.
- Homogeni entiteti.
- Entiteti na mikrorazini se zanemaruju.
- Dinamika se uvodi petljom povratne sprege.
- Kontinuirana vremenska progresija; vrijeme napreduje sinkronizirano za sve komponente modela sustava.

- Fleksibilnost se postiže promjenom strukture modela.
- Struktura sustava tijekom simulacije je fiksna.

Primjer

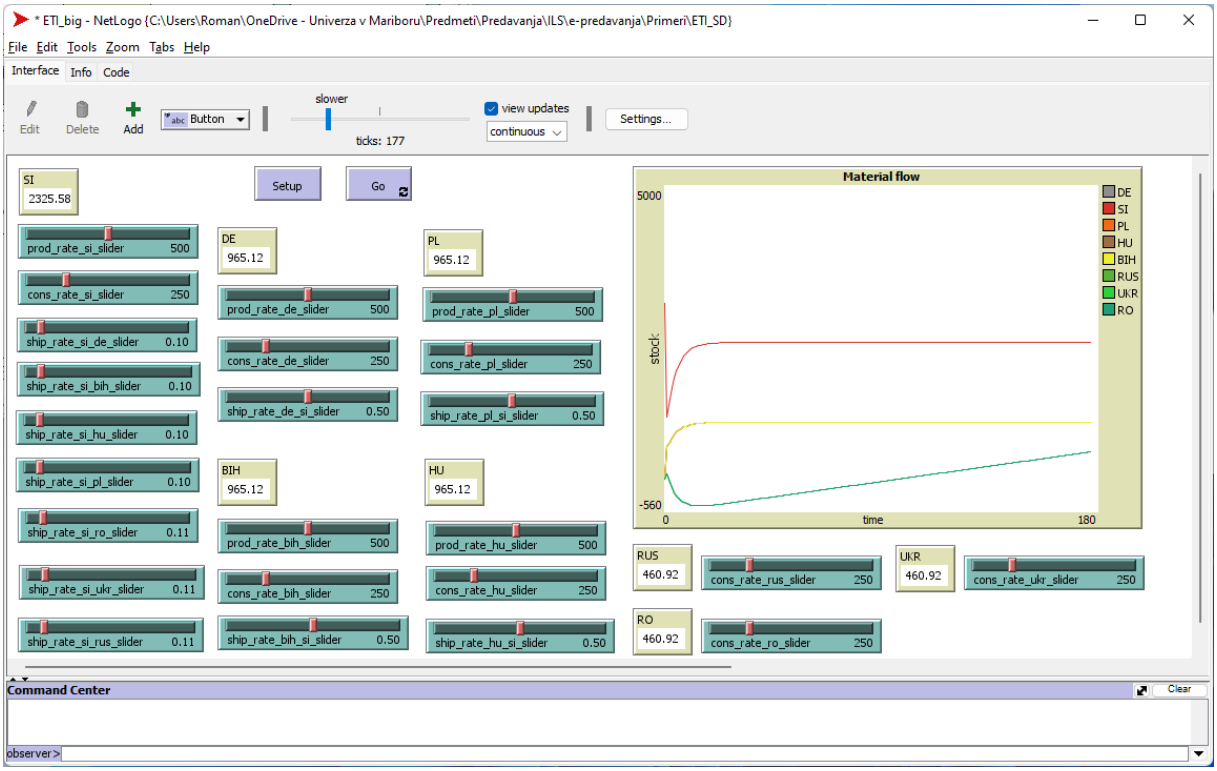
Primjer SD-a (Slika 4.2, iz simulacijskog okruženja NetLogo (Wilensky, 1999)) obuhvaća opskrbni lanac tvrtke kućanskih aparata i opisuje tokove materijala između njezinih podružnica (Gumzej i Rakovska, 2020). Tvrtka ima više proizvodnih lokacija: glavnu lokaciju u Sloveniji (SI) kao i podružnice u Njemačkoj (DE), Poljskoj (PL), Mađarskoj (H) i Bosni i Hercegovini (BIH). Uz proizvodna mjesta, njegova veleprodajna mjesta nalaze se u Rusiji (RUS), Ukrajini (UKR) i Rumunjskoj (RU). Proizvodne lokacije opskrbljuju vlastita tržišta gotovim proizvodima i jedna drugu komponentama proizvoda.



Slika 4.2Layout opskrbnog lanca.

Povezana NetLogo nadzorna ploča (slika 4.3) služi kao alat za podršku odlučivanju (engl. *decision support tool* - DST), za povezivanje proizvodnje i količina zaliha s predispozicijama i njihovom fizičkom distribucijom. Vremenski tijek je kontinuiran kroz svakodnevne transakcije, tj. svaki dan se određeni broj komponenti isporučuje između proizvodnih mjesta i određeni broj gotovih proizvoda se troši na licu mjesta ili se otprema na mjesta distribucije. Na temelju početnih zaliha od 300 jedinica na SI lokaciji, 0 zaliha na drugim lokacijama i modela

distribucije, količine zaliha na pojedinačnim lokacijama predstavljaju prosječne zalihe prema danoj proizvodnji (kom), potrošnji (%) i otpremi (%).



Slika 4.3Nadzorna ploča opskrbnog lanca.

Sinopsis

Simulacija sustavne dinamike omogućuje:

- Planiranje layouta SC-a.
- Optimizacija proizvodnih i distribucijskih kapaciteta.
- Procjena opterećenja kanala distribucije i povezanih troškova.

4.4 Simulacija temeljena na agentima

Simulacija temeljena na agentima (engl. *agent-based simulation* – ABS) analiza nudi pogled na tržište od strane strateškog menadžera ili regulatora tržišta. Stoga se ABS smatra alatom koji je najprikladniji za određivanje optimalne strukture i rasporeda/asortimana nečijeg tržišta i/ili

SC-a uzimajući u obzir njihove globalne karakteristike (npr. demografiju, klimu, BDP, kvalitetu, svijest, itd.).

Konstrukti:

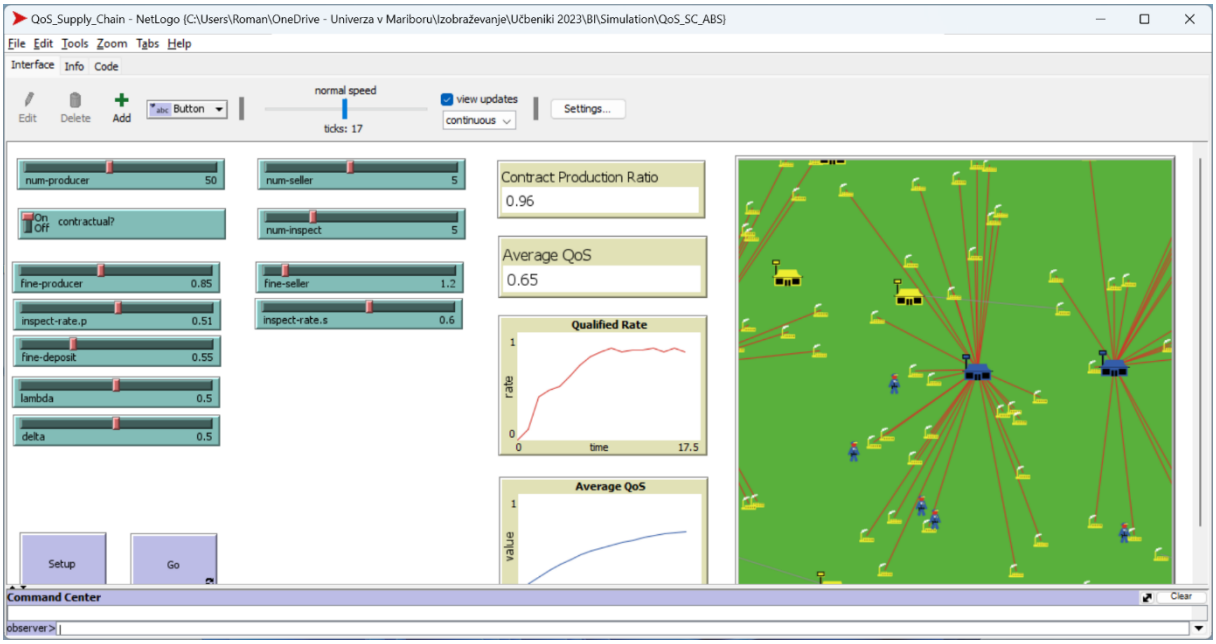
Agenti koji predstavljaju čvorove opskrbnog lanca (npr. dobavljači, trgovci na malo i inspektori) sa svojim svojstvima, odnosima i ponašanjem.

Svojstva:

- Usredotočen na entitet.
- Problemski orijentirano modeliranje entiteta i njihovih interakcija.
- Heterogenost entiteta.
- Mikro-entiteti su aktivni objekti koji djeluju u svom okruženju, međusobno komuniciraju i samostalno donose odluke.
- Odluke i interakcije između agenata unose dinamiku u sustave.
- Agenti i njihovo okruženje čine formalne modele.
- Protok vremena je diskretan i univerzalan na razini modela; vrijeme modela je u skladu s učestalošću SC transakcija i životnim ciklusima SC čvorova.
- Fleksibilnost modela postiže se promjenom strukture sustava i ponašanja agenata.
- Struktura sustava tijekom simulacije je promjenjiva.

Primjer

Primjer ABS-a (slika 4.4, iz simulacijskog okruženja NetLogo (Wilensky, 1999)) korišten je za analizu ponašanje ešalona opskrbnog lanca na otvorenom tržištu (Gumzej i Rakovska, 2020), s obzirom na njihovu kvalitetu usluge (engl. *Quality of Service* - QoS). U primjeru, različite politike koje se tiču ukupnog upravljanja kvalitetom tvrtke istraživane su modelom koji se sastoji od dobavljača, kupaca i regulatora tržišta.



Slika 4.4Regulacija tržišta.

Sinopsis

Simulacija temeljena na agentima omogućuje:

- Planiranje layouta SC-a.
- Modeliranje dinamičkog rasta SC-a.
- Modeliranje ponašanja partnera unutar SC-a.
- Optimizacija globalnih pokazatelja.

4.5 Simulacija mreže

Simulacija mreže (engl. *network simulation* – NS) analiza nudi mrežni regulatorni pogled na mrežu. Stoga se NS smatra alatom koji je najprikladniji za određivanje optimalne strukture, rasporeda i asortimana vlastite mreže uzimajući u obzir njezine globalne karakteristike (npr. propusnost, emisije, QoS pokazatelji, itd.).

Konstrukti:

Agenti koji predstavljaju objekte toka s njihovim svojstvima, odnosima i ponašanjem.

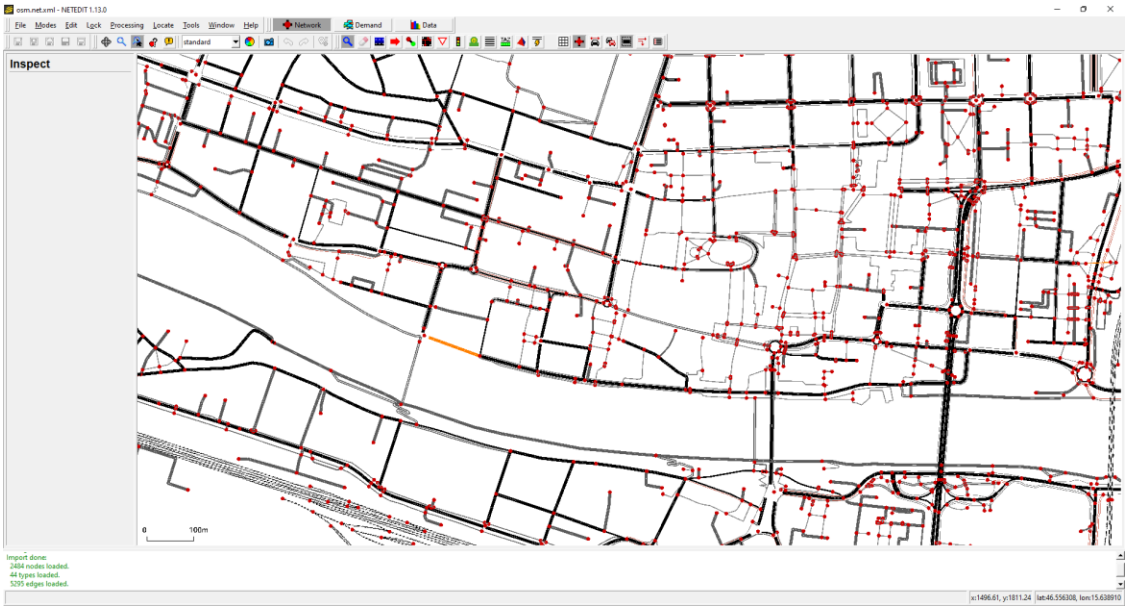
Mreža koja predstavlja mrežu preklapanja (npr. prometnu mrežu) na kojoj se objekti toka kreću.

Svojstva:

- Usmjeren na sustav.
- Problemski orijentirano modeliranje entiteta i njihovih interakcija.
- Heterogenost entiteta.
- Mikro-entiteti su aktivni objekti koji djeluju u svom okruženju, međusobno komuniciraju i samostalno donose odluke.
- Odluke i interakcije između agenata unose dinamiku u sustave.
- Agenti i njihovo okruženje čine formalne modele.
- Protok vremena je diskretan i univerzalan na razini modela; vrijeme je u skladu s relativnim brzinama objekata toka.
- Fleksibilnost modela postiže se promjenom mrežne strukture koja se fiksira tijekom simulacije i ponašanja agenata koji variraju ovisno o stanju (prometa) mreže i njihovim ciljevima.

Primjer

Predstavljeni primjer (slika 4.5, iz SUMO (Pablo et.al., 2018.) simulacijskog okruženja) korišten je za određivanje prometnih tokova i propusnosti ulica u središtu grada pogođenih planiranom blokadom ceste (Šinko i Gumzej, 2021). Osim toga, mjereni su pokazatelji povezani s prometom poput vremena putovanja, potrošnje goriva i emisija.



Slika 4.5Prometna situacija i mreža.

Sinopsis

Mrežna simulacija omogućuje:

- Planiranje layouta mreže.
- Modeliranje dinamičkog ponašanja mreže za određivanje uskih grla i slabih veza.
- Modeliranje toka mrežnih stavki.
- Optimizacija pokazatelja globalne mreže.

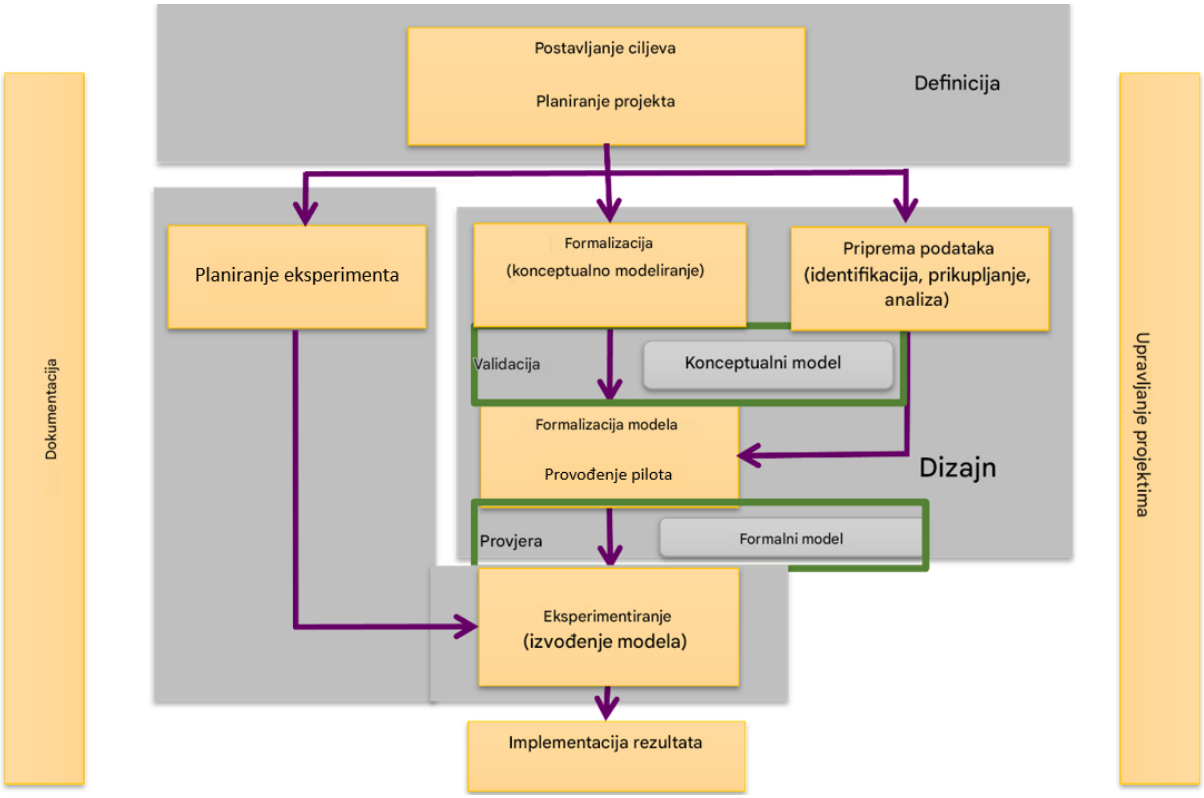
4.6 Projekti logističke simulacije

Projekti logističke simulacije dizajnirani su u skladu s paradigmom Dizajn za šest sigma (engl. *Design for Six Sigma* - DFSS) i temelje se na Demingovom ciklusu poboljšanja:

- Planiranje: definiranje sustava i ciljeva.
- Izvedba: dizajn simulacijskog modela.
- Analiza: eksperimentiranje sa simulacijskim modelom i procjena alternativa.
- Radnja: korištenje rezultata simulacije za implementaciju poboljšanja.

Svaki projekt logističke simulacije (slika 4.6) sastoji se od sedam faza:

1. Strateški plan: analiza postojećih i predloženih resursa i procesa.
2. Konceptualni model: apstraktni model sustava i definicija predispozicija, prikupljanje podataka.
3. Logički model: tok objekta, zalihe i tok ili mrežni dijagram modela sustava.
4. Simulacijski model: izrada odgovarajućeg simulacijskog modela.
5. Verifikacija i validacija simulacijskog modela: provjera konzistentnosti i koherentnosti modela.
6. Analiza na temelju simulacijskog modela: dizajn i izvedba eksperimenta.
7. Korištenje rezultata simulacije za izradu akcijskog plana: projekcija poboljšanja sustava.



Slika 4.6Proces simulacijskog modeliranja i analize (SMA).

4.7 Zaključak

U logistici je SMA važna komponenta operacijskog istraživanja (OR) koja omogućuje optimizaciju procesa.

Sustavna dinamika (SD) je metodologija za analizu složenih, dinamičkih i nelinearnih interakcija u sustavima, što rezultira novim strukturama i politikama za poboljšanje ponašanja sustava. Ovdje se rješavaju fizički i informacijski tokovi s ciljem smanjenja njihovog kašnjenja i konačno zaliha u opskrbnom lancu.

Druga popularna procesno orijentirana metodologija je simulacija diskretnih događaja (DES). To je jedan od najčešće korištenih i najfleksibilnijih analitičkih alata u SMA proizvodnih sustava. Uspješno se nosi s neizvjesnošću i pruža mogućnosti usporedbe alternativnih načina za smanjenje vremena isporuke kao i optimizaciju korištenja strojeva i resursa.

Korisna metodologija za razumijevanje ponašanja organizacija i njihovih interakcija (npr. opskrbnih lanaca i njihovih entiteta) je simulacija temeljena na agentima (ABS). Simulacija mreže (NS), kao posebna vrsta ABS-a, omogućuje modeliranje i optimizaciju (prometnih) preklopnih mreža.

To dovodi do zaključka da holistički pristup primjene SMA u logistici uvelike doprinosi složenim odlukama o dizajnu sustava, gdje postoji mnogo varijabli koje međusobno djeluju. Koristan integrirani pristup, uključujući metodologije SD, DES i ABS, koji može kvantificirati protoke na različitim razinama opskrbnog lanca, predstavljen je u (Gumzej & Rakovska, 2020). U analizi i optimizaciji prometnog toka, NS metodologija pruža potreban okvir u vezi s praćenjem i finim podešavanjem ključnih pokazatelja uspješnosti (Šinko i Gumzej, 2021).

Literatura 4. poglavlja

- Conant R.C. and Ashby W.R. (1970). Every good regulator of a system must be a model of that system, Int. J. Systems Sci., 1(2), pp. 89-97.
- Gumzej, R. and Rakovska, M. (2020). Simulation modeling and analysis for sustainable supply chains. In Ecoproduction – Sustainable logistics and production in industry 4.0 : new opportunities and challenges, Grzybowska, K., Awasthi, A., Sawhney, R. (ed.). Springer Nature, pp. 145-160.
- JaamSim Development Team (2023). JaamSim: Discrete-Event Simulation Software. Version 2023-08. [Available at: https://jaamsim.com, access November 8th, 2023]
- Šinko, S. and Gumzej, R. (2021). Towards smart traffic planning by traffic simulation on microscopic level. International journal of applied logistics, 11(1), pp. 1-17.

- Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. [Available at: http://ccl.northwestern.edu/netlogo/, access November 8th, 2023]
- Pablo A.L., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P. and Wießner, E. (2018). Microscopic Traffic Simulation using SUMO. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE. The 21st IEEE International Conference on Intelligent Transportation Systems, 4.-7. Nov. 2018, Maui, USA, pp. 2575-2582.

5. Linearna regresija s jednom i više regresorskih varijabli

Odluke menadžmenta često se temelje na odnosu između dvije ili više varijabli. Na primjer, voditelj marketinga može pokušati predvidjeti prodaju na određenoj razini troškova oglašavanja nakon ispitivanja odnosa između tih izdataka i prodaje.

U drugom slučaju, javno poduzeće može koristiti omjer između maksimalne dnevne vrijednosti temperature i potrebe za električnom energijom za predviđanje potrošnje električne energije. Ponekad se menadžer oslanja na intuiciju. Intuitivno prosuđuje kako su dvije varijable povezane. Međutim, ako je moguće dobiti podatke, ima smisla koristiti statistički postupak koji se zove regresijska analiza kako bi se pokazalo kako su te dvije varijable povezane jedna s drugom.

U terminologiji regresije, predviđena varijabla se naziva zavisna varijabla.

Varijabla ili varijable koje se koriste za predviđanje vrijednosti zavisne varijable nazivaju se nezavisne varijable.

U analizi učinka izdataka za oglašavanje na prodaju, prodaja bi stoga bila zavisna varijabla. Izdaci za oglašavanje bili bi nezavisna varijabla. U statističkom zapisu y označava zavisnu varijablu, a x označava nezavisnu varijablu.

U ovom ćemo odjeljku pogledati najjednostavniju vrstu regresijske analize koja uključuje jednu nezavisnu varijablu i jednu zavisnu varijablu. Odnos između dviju varijabli bit će aproksimiran ravnom linijom. Naziva se jednostavnom linearnom regresijom. Regresijska analiza koja uključuje dvije ili više nezavisne varijable naziva se višestruka regresijska analiza.

5.1 Jednostavni linearni regresijski model

Best Burger je lanac restorana brze hrane koji se nalazi u području s više država. Najbolje lokacije Burgera nalaze se u blizini sveučilišnih kampusa. Menadžeri vjeruju da je tromjesečna prodaja ovih restorana (označeno s y) u pozitivnoj korelaciji s veličinom studentske populacije (označeno s x). Restorani u blizini kampusa s velikim brojem studenata obično generiraju veću prodaju od onih u blizini kampusa s malim brojem studenata. Pomoću regresijske analize možemo razviti jednadžbu koja pokazuje kako je y zavisna varijabla povezana s nezavisnom varijablom x.

5.2 Regresijski model i regresijska jednadžba

U slučaju Best Burgera, populaciju čine svi restorani Best Burger. Za svaki restoran u populaciji postoji vrijednost x (studentska populacija) i odgovarajuća vrijednost y (tromjesečna prodaja). Jednadžba koja opisuje kako je y povezana s x zove se regresijski model.

$$y = \beta_0 + \beta_1 x + \epsilon$$

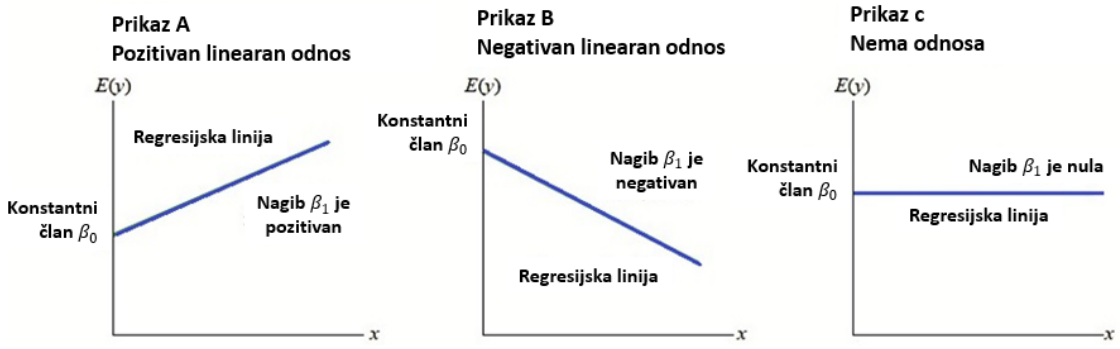
β_0 i β_1 nazivaju se parametri modela, ϵ (grčko slovo epsilon) je slučajna varijabla koja se naziva pogreška modela. Greška predstavlja varijabilnost y što se ne može objasniti linearnim odnosom između x i y.

Populacija svih restorana Best Burger također se može promatrati kao zbirka podpopulacija, jedna za svaku zasebnu vrijednost x. Na primjer, jednu podpopulaciju čine svi restorani Best Burger u blizini sveučilišnih kampusa s 8000 studenata. Drugu podpopulaciju čine svi restorani Best Burger koji se nalaze u blizini sveučilišnih kampusa s 9000 studenata i tako dalje. Svaka subpopulacija ima odgovarajuću raspodjelu vrijednosti y. Svaka raspodjela vrijednosti y ima svoju srednju ili očekivanu vrijednost. Jednadžba koja opisuje što je očekivana vrijednost y, označena s $E(y)$, čime je povezana x naziva se regresijska jednadžba. Regresijska jednadžba za jednostavnu linearnu regresiju je sljedeća

$$E(y) = \beta_0 + \beta_1 x$$

Grafikon jednostavne jednadžbe linearne regresije je ravna linija. β_0 predstavlja početnu vrijednost regresijske linije, β_1 je koeficijent smjera linije i $E(y)$ srednju vrijednost ili očekivanu vrijednost y za danu vrijednost x.

Primjeri mogućih regresijskih linija prikazani su na slici 5.1 u nastavku. Regresijska y linija u slučaju A pokazuje da je vrijednost y u pozitivnoj korelaciji s x. Kako se vrijednosti povećavaju x, vrijednosti se također $E(y)$ povećavaju. Tamo gdje su manje vrijednosti $E(y)$ povezane su s višim vrijednostima x. Regresijska linija na prikazu C prikazuje slučaj u kojem vrijednost y nije povezana s x. To znači da je vrijednost y ista za svaku vrijednost x.



Slika 4.7Primjeri grafikona linearnog odnosa.

5.3 Procijenjena regresijska jednadžba

Kad bi bile poznate vrijednosti parametara populacije β_0 i β_1, mogli bismo koristiti gornju jednadžbu za izračunavanje vrijednosti y za zadanu vrijednost x. U praksi je tim parametrima teško pristupiti, pa se jednostavno procjenjuju korištenjem podataka uzorka. Statistika uzorka (označena s b_0 i b_1) izračunata je kao procjene parametara populacije β_0 i β_1. Zamjenom vrijednosti statistike uzorka b_0 i b_1 umjesto β_0 i β_1 u regresijskoj jednadžbi dobivamo novu, procijenjenu regresijsku jednadžbu. Procijenjena regresijska jednadžba za jednostavnu linearnu regresiju je sljedeća

$$\hat{y} = b_0 + b_1 x$$

Grafikon procijenjene jednostavne linearne regresije naziva se procijenjena regresijska linija. b_0 predstavlja početnu vrijednost regresijske linije, b_1 je koeficijent smjera linije.

U nastavku ćemo pokazati kako koristiti metodu najmanjih kvadrata za izračunavanje vrijednosti b_0 i b_1 u procijenjenoj regresijskoj jednadžbi.

Općenito je $\hat{y}$ (rezultat za $E(y)$) prosječna vrijednost y za danu vrijednost x. Ako sada želimo procijeniti očekivanu vrijednost tromjesečne prodaje za sve restorane Best Burger koji se nalaze u blizini kampusa s 10 000 studenata, vrijednost bi x bila zamijenjena vrijednošću 10

000 u posljednjoj jednadžbi. U nekim slučajevima, međutim, možda ćemo biti više zainteresirani za predviđanje prodaje samo za jedan određeni restoran, na primjer, pretpostavimo da želite predvidjeti kvartalnu prodaju za restoran koji planirate izgraditi u blizini fakulteta s 10 000 studenata, pokazalo se da je čak i u ovom slučaju najbolji prediktor vrijednosti y za danu x vrijednost $\hat{y}$.

5.4 Metoda najmanjih kvadrata

Metoda najmanjih kvadrata je postupak u kojem se pomoću uzoraka podataka nalazi jednadžba procijenjene regresijske linije. Kako bismo ilustrirali metodu najmanjih kvadrata, pretpostavimo da su podaci prikupljeni iz uzorka od 10 restorana s najboljim hamburgerima u blizini sveučilišnih kampusa. Sa x_i će označavati veličinu studentske populacije (u tisućama) i veličinu y_i tromjesečne prodaje (u tisućama EUR). x_i *i* y_i za 10 uzoraka restorana sažeti su u tablici u nastavku. Vidimo da je restoran 1, z x_1= 2 i y_1= 58, blizu kampusa s 2000 studenata i ima kvartalnu prodaju od 58 000 €. Restoran 2, s x_2= 6 i y_2= 105, blizu je kampusa sa 6000 studenata i ima kvartalnu prodaju od 105.000 €. Restoran s najvećom prodajnom vrijednošću je restoran 10, koji se nalazi u blizini kampusa s 26.000 studenata i ima kvartalnu prodaju od 202.000 €.

Slijedi dijagram raspršenosti podataka na slici 5.2 u nastavku. Studentska populacija prikazana je na vodoravnoj osi, a kvartalna prodaja na okomitoj osi. Dijagrami raspršenosti za regresijsku analizu konstruirani su s nezavisnom varijablom x na vodoravnoj osi i ovisnom varijablom y na okomitoj osi. Dijagram raspršenosti nam stoga omogućuje izvođenje preliminarnih zaključaka o mogućem odnosu između varijabli.

Restoran	Studentska populacija (u 1000-ama)	Kvartalna prodaja (u 1000ama eura)
i	x_i	y_i
1	2	58
2	6	105
3	8	88
4	8	118
5	12	117
6	16	137
7	20	157
8	20	169
9	22	149
10	26	202

Slika 4.8 Dijagram raspršenosti podataka.

Koji se preliminarni zaključci mogu izvući iz slike ispod 5.3? Veća tromjesečna prodaja događa se u kampusima s većom populacijom studenata. Osim toga, postoji konstantan odnos između veličine studentske populacije i tromjesečne prodaje, koji se može opisati pravocrtno. Između $x\,i\,y$ se doista podrazumijeva pozitivan linearni odnos. Stoga smo odabrali jednostavan linearni regresijski model za prikaz odnosa između tromjesečne prodaje i studentske populacije. S obzirom na ovaj izbor, naš sljedeći zadatak je koristiti tablicu podataka uzorka za određivanje vrijednosti b_0 i b_1, koji su važni parametri u procjeni jednostavne jednadžbe linearne regresije. Za i-ti restoran procijenjena regresijska jednadžba je

$$\hat{y}_i = b_0 + b_1 x_i$$

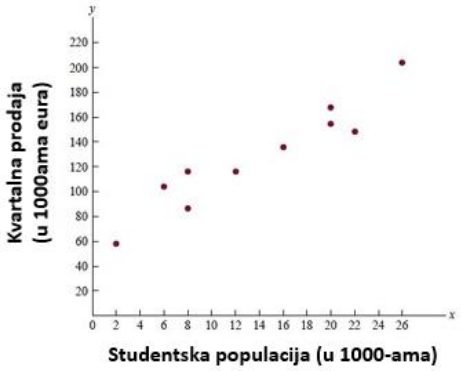
gdje je

$\hat{y}_i$_ procijenjena vrijednost tromjesečne prodaje (1000 €) za i-ti restoran

b_0_ početna vrijednost procijenjene regresijske linije

b_1_ koeficijent smjera procijenjene regresijske linije

x_i _veličina studentske populacije (1000) za i-ti restoran



Slika 4.9 Raspršeni grafikon.

y_i označava opaženu (stvarnu) prodaju za restoran i i $\hat{y}_i$, predstavljajući procijenjenu vrijednost prodaje za restoran i, svaki će restoran u uzorku imati opaženu prodajnu vrijednost od y_i i predviđenu prodajnu vrijednost $\hat{y}_i$. Kako bi procijenjena regresijska linija osigurala dobro

uklapanje u podatke, želimo da razlike između opaženih prodajnih vrijednosti i predviđenih prodajnih vrijednosti budu što manje.

Metoda najmanjih kvadrata koristi uzorke podataka za pružanje vrijednosti b_0 i b_1.

Minimizirajte zbroj kvadrata odstupanja između promatranih vrijednosti zavisne varijable y_i i predviđene vrijednosti zavisne varijable $\hat{y}_i$. Polazna točka za izračunavanje minimalnog zbroja metodom najmanjih kvadrata dana je izrazom

Kriterij minimalnog iznosa: $\min \sum(\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2$

gdje je

$\mathrm{y_i}$=promatrana vrijednost zavisne varijable za i-to opažanje

$\hat{\mathrm{y}}_\mathrm{i}$ =predviđena vrijednost zavisne varijable za i-to opažanje

Koeficijent smjera regresijske linije i početna vrijednost:

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

x_i_ vrijednost nezavisne varijable za i-to opažanje

y_i_ vrijednost zavisne varijable za i-to opažanje

$\bar{x}$ _ prosječna vrijednost za nezavisnu varijablu

$\bar{y}$ _ prosječna vrijednost za zavisnu varijablu

n_ukupan broj opažanja

Neki od izračuna potrebnih za izradu procijenjene linije regresije najmanjih kvadrata prikazani su u nastavku. Na uzorku od 10 restorana imamo n=10 opažanja. Gornje jednadžbe prvo zahtijevaju izračun srednje vrijednosti x i prosječne vrijednosti y.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{140}{10} = 14, \qquad \bar{y} = \frac{\sum y_i}{n} = \frac{1300}{10} = 130$$

Alternativna jednadžba izračuna b_1:

$$b_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2}$$

Koristeći posljednje jednadžbe i informacije na slici 5.4, možemo izračunati usmjereni koeficijent regresijske linije za primjer restorana Best Burger. Izračunavanje nagiba (b_1) je kako slijedi.

Slika 5.5 prikazuje dijagram ove jednadžbe na raspršenom dijagramu.

Nagib procijenjene regresijske jednadžbe ili koeficijent smjera jednadžbe ($b_1 = 5$) je pozitivan.

Restaurant i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	2	58	−12	−72	864	144
2	6	105	−8	−25	200	64
3	8	88	−6	−42	252	36
4	8	118	−6	−12	72	36
5	12	117	−2	−13	26	4
6	16	137	2	7	14	4
7	20	157	6	27	162	36
8	20	169	6	39	234	36
9	22	149	8	19	152	64
10	26	202	12	72	864	144
Totals	140	1300			2840	568
	Σx_i	Σy_i			$\Sigma(x_i - \bar{x})(y_i - \bar{y})$	$\Sigma(x_i - \bar{x})^2$

Slika 4.10 Prikaz jednadžbe na dijagramu raspršenosti.

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{2840}{568} = 5$$

Nakon toga slijedi izračun početne vrijednosti (b_0).

$$b_0 = \bar{y} - b_1\bar{x} = 130 - 5(14) = 60$$

Ovako se procjenjuje regresijska jednadžba:

$$\hat{y} = 60 + 5x$$

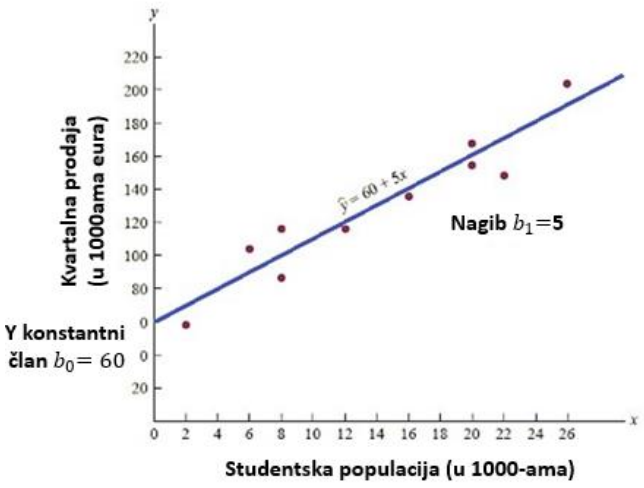
Slika prikazuje dijagram ove jednadžbe na dijagramu raspršenosti.

Nagib procijenjene regresijske jednadžbe (b_1= 5) je pozitivan, što znači da kako se broj studenata povećava, prodaja se povećava. Zapravo, možemo zaključiti (na temelju izmjerene prodaje u 1000-ama i studentske populacije u 1000-ama), što znači da je porast u studentskoj populaciji od 1000 povezan s povećanjem očekivane prodaje od 5000; tj. očekuje se povećanje tromjesečne prodaje za 5 € po studentu.

Ako vjerujemo da regresijska jednadžba, procijenjena najmanjim kvadratima, adekvatno opisuje odnos između x i y, čini se razumnim koristiti procijenjenu regresijsku jednadžbu za predviđanje vrijednosti y za danu vrijednost x. Na primjer, ako želite predvidjeti tromjesečnu prodaju za restoran koji se nalazi u blizini kampusa od 16.000 studenata, izračunali biste

$$\hat{y} = 60 + 5(16) = 140$$

Stoga bismo pretpostavili kvartalnu prodaju od 140.000 za ovaj restoran. U sljedećim odjeljcima raspravljamo o metodama za procjenu prikladnosti korištenja procijenjene regresijske jednadžbe za procjenu i predviđanje.



Slika 4.11Dijagram raspršenosti studentske populacije i tromjesečne prodaje.

5.5 Koeficijent determinacije

Za primjer restorana Best Burger razvili smo procijenjenu regresijsku jednadžbu $y = 60 + 5x$ za približno linearni odnos između veličine studentske populacije x i tromjesečne prodaje y. Sada je pitanje: koliko dobro procijenjena regresijska jednadžba odgovara podacima? U ovom odjeljku pokazujemo da koeficijent determinacije daje mjeru dobrog uklapanja za procijenjenu regresijsku jednadžbu. Za i-to opažanje, razlika između opažene vrijednosti zavisne varijable y_i i predviđene vrijednosti zavisne varijable naziva se i-ti rezidualno odstupanje.

Zbroj kvadrata ovih rezidualnih odstupanja ili pogrešaka je količina koja je minimizirana metodom najmanjih kvadrata. Ova količina, također poznata kao rezidualni zbroj kvadrata, označava se sa SSE.

$$SSE = \sum (y_i - \hat{y}_i)^2$$

SSE vrijednost je mjera pogreške u korištenju procijenjene regresijske jednadžbe za predviđanje vrijednosti zavisne varijable u uzorku. Slika 5.6 prikazuje izračune potrebne za izračunavanje zbroja kvadrata zbog pogreške za slučaj Best Burger.

Restoran i	x_i = Studentska populacija (u 1000-ama)	y_i = Kvartalna prodaja (u 1000ama eura)	Predviđena prodaja $\hat{y}_i = 60 + 5x_i$	Pogreška $y_i - \hat{y}_i$	Standardna pogreška $(y_i - \hat{y}_i)^2$
1	2	58	70	−12	144
2	6	105	90	15	225
3	8	88	100	−12	144
4	8	118	100	18	324
5	12	117	120	−3	9
6	16	137	140	−3	9
7	20	157	160	−3	9
8	20	169	160	9	81
9	22	149	170	−21	441
10	26	202	190	12	144
				SSE =	1530

Slika 4.12 Kvadrati pogrešaka u slučaju Best Burger.

Pretpostavimo da se od nas traži da napravimo procjenu tromjesečne prodaje bez da znamo veličinu studentske populacije. Bez poznavanja bilo koje povezane varijable, koristili bismo prosjek uzorka kao procjenu tromjesečne prodaje u bilo kojem restoranu. Tablica na slici 5.6 pokazala je da je podatke o prodaji y_i=1300. Stoga je prosječna tromjesečna vrijednost prodaje za uzorak od 10 najboljih restorana s hamburgerima y_i/n = 1300/10 = 130. Na slici 5.7 prikazujemo zbroj kvadrata odstupanja dobivenih korištenjem srednje vrijednosti uzorka od 130 za predviđanje vrijednosti kvartalne prodaje za svaki restoran u uzorku. Za i-ti restoran u uzorku razlika y_i daje mjeru pogreške koja je uključena u aplikaciju za predviđanje prodaje. Odgovarajući zbroj kvadrata, koji se naziva ukupan zbroj kvadrata, označava se sa SST.

$$SST = \sum (y_i - \bar{y})^2$$

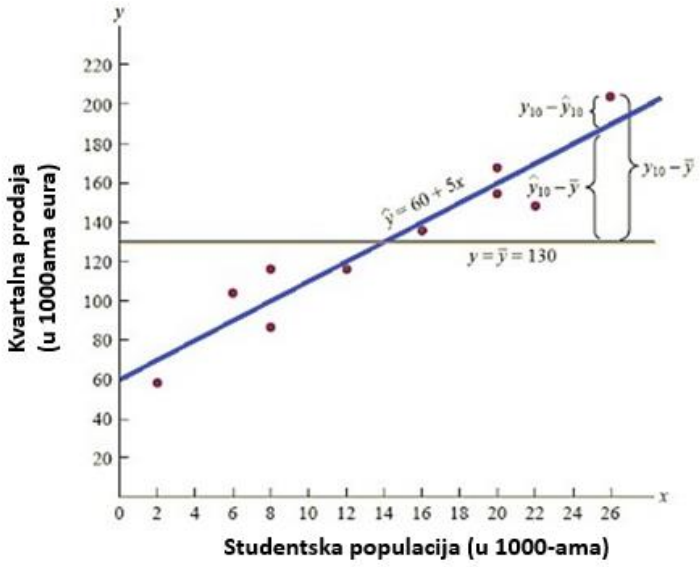
Restoran i	x_i = Studentska populacija (u 1000-ama)	y_i = Kvartalna prodaja (u 1000ama eura)	Devijacija $y_i - \bar{y}$	Standardna devijacija $(y_i - \bar{y})^2$
1	2	58	−72	5184
2	6	105	−25	625
3	8	88	−42	1764
4	8	118	−12	144
5	12	117	−13	169
6	16	137	7	49
7	20	157	27	729
8	20	169	39	1521
9	22	149	19	361
10	26	202	72	5184
			SST =	15,730

Slika 4.13 Zbroj kvadrata odstupanja.

Zbroj na dnu zadnjeg stupca na slici 5.7 je ukupni zbroj kvadrata za BestBurgerove restorane SST = 15,730. Na slici 5.8 prikazujemo procijenjenu regresijsku liniju $y = 60 + 5x$ i liniju koja odgovara y = 130. Imajte na umu da se točke grupiraju bliže oko procijenjene regresijske linije nego oko linije y = 130. Na primjer, za 10. restoran u uzorku, vidi se da je pogreška puno veća kada se 130 koristi za predviđanje y = 10 nego kada se 130 koristi $y = 60 + 5x$ i iznosi 190. Možemo se sjetiti SST kao mjere koliko dobro se opažanja grupiraju oko linije i SSE kao mjere koliko dobro se opažanja grupiraju oko linije.

Da bi se izmjerilo koliko vrijednosti na procijenjenoj regresijskoj liniji odstupaju od sljedećeg, izračunava se još jedan zbroj kvadrata. Ovaj zbroj kvadrata, koji se naziva regresijskim zbrojem kvadrata, označava se kao SSR.

$$SSR = \sum (\hat{y}_i - \bar{y})^2$$



Slika 4.14 Regresijska linija u slučaju Best Burgera.

Iz prethodne rasprave, trebali bismo očekivati da su SST, SSR i SSE povezani. Zapravo, odnos između ta tri zbroja kvadrata jedan je od najvažnijih rezultata u statistici.

5.6 Odnos između SST, SSR i SSE

$$SST = SSR + SSE$$

Gdje je:

SST = ukupan zbroj kvadrata

SSR = regresijski zbroj kvadrata

SSE = rezidualni zbroj kvadrata

Jednadžba ($SST = SSR + SSE$) pokazuje da se ukupni zbroj kvadrata može podijeliti na dvije komponente, regresijski zbroj kvadrata i rezidualni zbroj kvadrata. Dakle, ako su poznate vrijednosti bilo koja dva od ovih zbroja kvadrata, treći zbroj kvadrata može se lako saznati izračunom. Na primjer, u slučaju Best Burger restorana, već znamo da je SSE = 1530 i SST = 15,730; stoga, rješavanjem za SSR u gornjoj jednadžbi, nalazimo da je regresijski zbroj kvadrata

$$SSR = SST - SSE = 15730 - 1530 = 14200$$

Sada ćemo vidjeti kako možemo koristiti tri zbroja kvadrata, SST, SSR i SSE, da bismo dobili kriterij prilagodbe za procijenjenu regresijsku jednadžbu. Procijenjena regresijska jednadžba bi savršeno odgovarala ako bi svaka vrijednost zavisne varijable $yi\_$ ležala nasumično na procijenjenoj regresijskoj liniji. U ovom slučaju to bi bilo nula za svako opažanje, što bi rezultiralo SSE =0. Budući da je SST = SSR + SSE, vidimo da za savršeno pristajanje SSR mora biti jednak SST i omjer (SSR/SST) mora biti jednak jedan. Lošija prilagodba rezultirat će većim vrijednostima za SSE. Rješavajući SSE u jednadžbi, vidimo da je SSE = SST - SSR. Stoga se najveća vrijednost za SSE (a time i najlošije uklapanje) javlja kada je SSR = 0 i SSE = SST.

Za procjenu se koristi omjer SSR/SST, koji ima vrijednosti između nula i jedan koji odgovara procijenjenoj regresijskoj jednadžbi.

Taj se omjer naziva koeficijent determinacije i označava se s r^2.

$$r^2 = \frac{SSR}{SST}$$

Za primjer restorana Best Burger, vrijednost koeficijenta determinacije je

$$r^2 = \frac{SSR}{SST} = \frac{14200}{15730} = 0.9027$$

Kada se koeficijent determinacije izrazi kao postotak, r^2 se može tumačiti kao postotak ukupnog zbroja kvadrata koji se može objasniti pomoću procijenjene regresijske jednadžbe. Za najbolje restorane s hamburgerima možemo zaključiti da se 90,27% ukupnog zbroja kvadrata može objasniti korištenjem procijenjene regresijske jednadžbe $y = 60 + 5x$ za predviđanje kvartalne prodaje. Drugim riječima, 90,27% varijabilnosti u prodaji može se objasniti linearnim odnosom između veličine studentske populacije i prodaje. Trebalo bismo biti zadovoljni što vidimo da se tako dobro uklapa u procijenjenu regresijsku jednadžbu.

5.7 Koeficijent korelacije

Koeficijent korelacije može se smatrati opisnom mjerom snage linearnog odnosa između dviju varijabli, x i y. Vrijednosti koeficijenta korelacije su uvijek između -1 i +1. Vrijednost +1 znači da su $x\ i\ y$ dvije varijable u savršenoj korelaciji u pozitivnom linearnom smislu. To znači da su sve podatkovne točke na a ravna linija s pozitivnim nagibom. Vrijednost -1 znači da su $x\ i\ y$ savršeno povezane u negativnom linearnom smislu, sa svim podatkovnim točkama na ravnoj liniji s negativnim nagibom. Vrijednosti koeficijenta korelacije blizu nule znače da x i y nisu linearno povezani.

Ako je regresijska analiza već provedena i koeficijent determinacije r^2 je izračunat, koeficijent korelacije uzorka može se izračunati na sljedeći način.

$$r_{xy} = (predznak\ b_1)\sqrt{koeficijent\ determinacije}$$

$$r_{xy} = (predznak\ b_1)\sqrt{r^2}$$

PEARSONOV KOEFICIJENT KORELACIJE: UZORCI PODATAKA

$$rxy = \frac{s_{xy}}{s_x s_y} = \frac{\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}\sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}}} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

$$rxy = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

gdje su:

$s_{xy} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1}$, $s_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$, $s_y = \sqrt{\frac{\sum(y_i-\bar{y})^2}{n-1}}$

Predznak koeficijenta korelacije uzorka je pozitivan ako procijenjena regresijska jednadžba ima pozitivan nagib ($b_1 > 0$) i negativan ako procijenjena regresijska jednadžba ima negativan nagib ($b_1 < 0$).

Za slučaj Best Burger, vrijednost koeficijenta determinacije koji odgovara procijenjenoj regresijskoj jednadžbi $y = 60 + 5x$ je 0,9027. Budući da je nagib procijenjene regresijske jednadžbe pozitivan, jednadžba pokazuje da je koeficijent korelacije uzorka prema koeficijentu korelacije uzorka

Rxy = 0,9501, zaključili bismo da postoji jaka pozitivna linearna veza između x i y.

U slučaju linearnog odnosa između dviju varijabli, oba koeficijenta determinacije i koeficijent korelacije uzorka daju mjeru snage odnosa.

Koeficijent determinacije daje mjeru između nula i jedan, dok koeficijent korelacije uzorka daje mjeru između -1 i +1. Iako je koeficijent korelacije uzorka ograničen na linearni odnos između dviju varijabli, koeficijent determinacije može se primijeniti na nelinearne odnose i na odnose koji imaju dvije ili više nezavisnih varijabli. Dakle, koeficijent determinacije pruža širi raspon primjenjivosti.

5.8 Model višestruke regresije

U sljedećim odjeljcima nastavljamo naše proučavanje regresijske analize razmatrajući situacije koje uključuju dvije ili više neovisnih varijabli. Ovo predmetno područje, koje se naziva višestruka regresijska analiza, omogućuje nam da uzmemo u obzir više faktora i tako dobijemo bolja predviđanja nego što je to moguće s jednostavnom linearnom regresijom.

Višestruka regresijska analiza je studija o tome kako je zavisna varijabla y povezana s dvije ili više nezavisnih varijabli. U općem slučaju, s p ćemo označiti broj nezavisnih varijabli.

5.9 Regresijski model i regresijska jednadžba

Koncepti regresijskog modela i regresijske jednadžbe uvedeni u prethodnom odjeljku primjenjuju se u slučaju višestruke regresije. Jednadžba koja opisuje kako je zavisna varijabla

y povezana s nezavisnim varijablama $x_1, x_2, \ldots, x_p$ i pogreškom naziva se modelom višestruke regresije. Počinjemo s pretpostavkom da model višestruke regresije ima sljedeći oblik.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

U modelu višestruke regresije $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ su parametri, a pogreška (ϵ) je slučajna varijabla. Pomno ispitivanje ovog modela otkriva da je y linearna funkcija varijabli $x_1, x_2, \ldots, x_p$ plus pogreška ϵ. Izraz pogreške uzima u obzir varijabilnost y koja se ne može objasniti linearnim učinkom p nezavisnih varijabli.

U odjeljku 5.10 raspravljamo o pretpostavkama za model višestruke regresije i epsilon. Jedna od pretpostavki je da je srednja ili očekivana vrijednost (ϵ) nula. Implikacija ove pretpostavke je da je srednja ili očekivana vrijednost y, označena s $E(y)$, jednaka $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$. Jednadžba koja opisuje kako je srednja vrijednost y povezana s $x_1, x_2, \ldots, x_p$ naziva se jednadžba višestruke regresije.

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

5.10 Procijenjena jednadžba višestruke regresije

Ako su vrijednosti $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ poznate, jednadžba iz 5.9 se može koristiti za izračunavanje prosječne vrijednosti y pri danim vrijednostima $x_1, x_2, \ldots, x_p$. Nažalost, ove vrijednosti parametara općenito neće biti poznate i moraju se procijeniti iz podataka uzorka. Jednostavan slučajni uzorak koristi se za izračun statistike uzorka $b_0, b_1, b_2, \ldots, b_p$ koja se koristi kao točkasti procjenitelj parametara $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$. Ovi uzorci statistike daju sljedeću procjenu jednadžbe višestruke regresije:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

gdje su

$b_0, b_1, b_2, \ldots, b_p$ procjene $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$

$\hat{y}$ = predviđena vrijednost zavisne varijable.

Primjer: Frigo transportna tvrtka

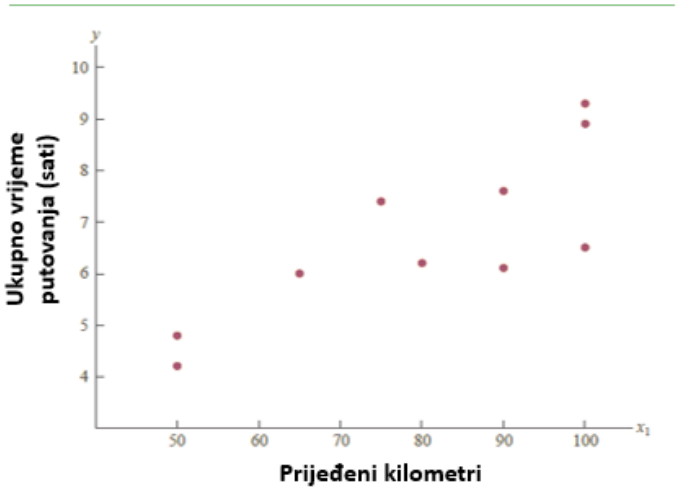
Kao ilustraciju višestruke regresijske analize, razmotrit ćemo problem s kojim se susreće Frigo transportna tvrtka, neovisna autoprijevoznička tvrtka u južnoj Italiji. Najveći dio poslovanja

tvrtke Frigo odnosi se na dostavu na cijelom lokalnom području. Za bolje rasporede rada, menadžeri žele planirati zajedničko dnevno vrijeme putovanja za svoje vozače.

U početku su menadžeri vjerovali da će ukupno dnevno vrijeme putovanja biti usko povezano s brojem prijeđenih kilometara u dnevnim isporukama. Jednostavan nasumični uzorak od 10 dodijeljenih vozača dao je podatke prikazane na slici 5.9 i dijagramu raspršenosti. Nakon pregleda ovog dijagrama raspršenosti, upravitelji su pretpostavili da se jednostavni linearni regresijski model može koristiti za opisivanje odnosa $y = \beta_0 + \beta_1 x_1 + \epsilon$ između ukupnog vremena putovanja (y) i broja prijeđenih km (x_1).

Vozački zadatak	x_1 = prijeđeni kilometri	y = vrijeme putovanja (sati)
1	100	9.3
2	50	4.8
3	100	8.9
4	100	6.5
5	50	4.2
6	80	6.2
7	75	7.4
8	65	6.0
9	90	7.6
10	90	6.1

Slika 4.15 Podaci za primjer Frigo transportne tvrtke.



Slika 4.16 Dijagram raspršenosti za Frigo transportnu tvrtku.

Za procjenu parametara β_0 i β_1 , jednadžba najmanjih kvadrata korištena je za izradu procijenjene regresije.

$$\hat{y} = b_0 + b_1 x_1$$

Gornja slika prikazuje prikaz softverskog programa Minitab-a korištenjem jednostavne linearne regresije na podatke u gornjoj tablici. Procijenjena regresijska jednadžba je

$$\hat{y} = 1{,}27 + 0{,}0678 x_1$$

Na razini značajnosti od 0,05, F-vrijednost od 15,81 i odgovarajuća p-vrijednost od 0,004 pokazuju da je odnos značajan. To znači da možemo odbiti $H_0: \beta_1 = 0$ jer je p-vrijednost manja od $\alpha = 0{,}05$. Imajte na umu da isti zaključak proizlazi iz vrijednosti $t = 3{,}98$ i pridružene p-vrijednosti od 0,004. Stoga možemo zaključiti da je odnos između ukupnog vremena putovanja i broja prijeđenih milja značajan. Dulje vrijeme putovanja povezano je s više prijeđenih kilometara. S koeficijentom determinacije (izraženim u postocima) $R - Sq = 66{,}4\,\%$, vidimo da se 66,4% varijabilnosti vremena putovanja može objasniti linearnim učinkom broja prijeđenih milja.

Ovo otkriće je prilično dobro, ali menadžeri bi mogli razmotriti dodavanje druge nezavisne varijable kako bi objasnili neke od preostalih varijabli u zavisnoj varijabli.

MINITAB OUTPUT FOR FRIGO TRUCKING WITH ONE INDEPENDENT VARIABLE

```
The regression equation is
Time = 1.27 + 0.0678 kM

Predictor      Coef  SE Coef     T      p
Constant      1.274    1.401  0.91  0.390
kM          0.06783  0.01706  3.98  0.004

S = 1.00179   R-Sq = 66.4%   R-Sq(adj) = 62.2%

Analysis of Variance

SOURCE           DF      SS      MS      F      p
Regression        1  15.871  15.871  15.81  0.004
Residual Error    8   8.029   1.004
Total             9  23.900
```

Slika 4.17 Rezultati s jednom nezavisnom varijablom.

Kada su pokušali identificirati drugu nezavisnu varijablu, menadžeri su smatrali da broj dostava također može pridonijeti ukupnom vremenu putovanja. Podaci Frigo transportne tvrtke, s

dodanim brojem dostava, prikazani su na slici 5.12. - (x_1) i broj isporuka (x_2), kao nezavisne varijable. Procijenjena regresijska jednadžba je

$$\hat{y} = -0.869 + 0.0611x_1 + 0.923x_2$$

PODACI ZA FRIGO TRANSPORTNU TVRTKU S PRIJEĐENIM KILOMETRIMA (x_1) I BROJEM DOSTAVA (x_2) KAO NEZAVISNIM VARIJABLAMA

Vozački zadatak	x_1 = prijeđeni kilometri	x_2 = broj dostava	y = vrijeme putovanja (sati)
1	100	4	9.3
2	50	3	4.8
3	100	4	8.9
4	100	2	6.5
5	50	2	4.2
6	80	2	6.2
7	75	3	7.4
8	65	4	6.0
9	90	3	7.6
10	90	2	6.1

Slika 4.18 Podaci Frigo transportne tvrtke i nezavisne varijable.

Pogledajmo pobliže vrijednosti b_1= 0,0611 i b_2= 0,923 u posljednjoj jednadžbi.

Napomena o tumačenju koeficijenata

U ovom trenutku možemo dati jedan komentar o odnosu između procijenjene regresijske jednadžbe sa samo prijeđenim miljama kao nezavisnom varijablom i jednadžbe koja uključuje broj isporuka kao drugu nezavisnu varijablu. Vrijednost b_1 nije ista u oba slučaja. U jednostavnoj linearnoj regresiji tumačimo b_1 kao procjenu promjene y za jednu jediničnu promjenu nezavisne varijable. U višestrukoj regresijskoj analizi ovo tumačenje treba malo modificirati. To jest, u višestrukoj regresijskoj analizi, svaki regresijski koeficijent se tumači na sljedeći način: predstavljao bi procjenu promjene u koja y odgovara promjeni u x_i za jednu jedinicu kada se sve ostale nezavisne varijable drže konstantnima.

U slučaju Frigo transportne tvrtke, to uključuje dvije nezavisne varijable, b_1 = 0,0611 i b_2 = 0,923.

MINITAB OUTPUT FOR FRIGO TRUCKING WITH TWO INDEPENDENT VARIABLES

The regression equation is
Time = - 0.869 + 0.0611 kM + 0.923 Deliveries

Predictor	Coef	SE Coef	T	p
Constant	-0.8687	0.9515	-0.91	0.392
kM	0.061135	0.009888	6.18	0.000
Deliveries	0.9234	0.2211	4.18	0.004

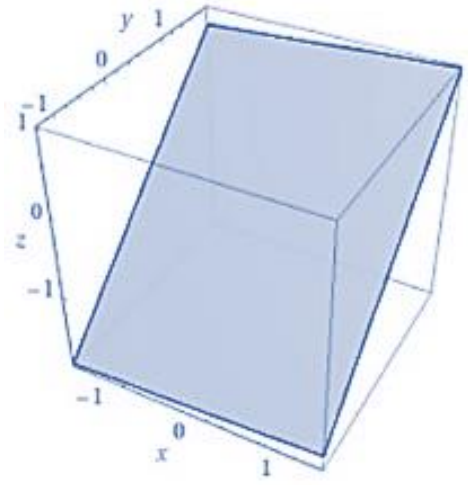
S = 0.573142 R-Sq = 90.4% R-Sq(adj) = 87.6%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	2	21.601	10.800	32.88	0.000
Residual Error	7	2.299	0.328		
Total	9	23.900			

Slika 4.19 Rezultati za Frigo transportnu tvrtku s dvije nezavisne varijable.

Stoga je 0,0611 sati procjena očekivanog povećanja vremena putovanja koja odgovara povećanju od jedne milje po prijeđenoj udaljenosti kada je broj dostava konstantan. Slično, budući da je b_2 = 0,923, procjena očekivanog povećanja vremena putovanja koja odgovara povećanju jedne isporuke kada je broj prijeđenih milja konstantan iznosi 0,923 sata.



Slika 4.20 Vizualni prikaz rezultata za Frigo transportnu tvrtku.

Literatura 5. poglavlja

- *Introductory Statistics.* Bentham Science Publishers, Kahl, A. (Publish 2023). DOI:10.2174/9789815123135123010l
- Introductory Statistics 2e, Openstax, Rice University, Houston, Texas 77005, Jun 23, senior contributing authors: Barbara Illowsky and Susan dean, De anza college, Publish Date: Dec 13, 2023, (https://openstax.org/details/books/introductory-statistics-2e);
- Introductory Statistics 4th Edition, Susan Dean and Barbara Illowsky, Adapted by Riyanti Boyd & Natalia Casper (Published 2013 by OpenStax College) July 2021, (http://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf);
- Journal of the Royal Statistical Society 2024**,** A reputable journal publishing cutting-edge research and articles on various aspects of statistics, including theoretical advancements and practical applications. Recent issues have featured studies on sampling and hypothesis testing.
- Introductory Statistics 7th Edition, Prem S. Mann, eastern Connecticut state university with the help of Christopher Jay Lacke, Rowan university, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, 2011
- Introduction to statistics, made easy second edition, Prof. Dr. Hamid Al-Oklah Dr. Said Titi Mr. Tareq Alodat, March 2014
- Statistics for Business and Economics, Thirteenth Edition, David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran, 2017, 2015 Cengage Learning®
- Statistics for Business, First edition, Derek L Waller, 2008 Copyright © 2008, Derek L Waller, Published by Elsevier Inc. All rights reserved

6. Uvod u operacijska istraživanja

Operacijska istraživanja (engl. *operational research*; *US Air Force Specialty Code: Operations Analysis*), često skraćeno na OR, je disciplina koja se bavi razvojem i primjenom analitičkih metoda za poboljšanje donošenja odluka. Povremeno se kao sinonim koristi pojam znanost o upravljanju.

OR metode se koriste za analizu kapaciteta resursa, uskih grla, vremena dostave i ciklusa, obrazaca potražnje, zaliha, distribucije resursa, održavanja, slanja operatera, miješanja proizvoda, izlaza proizvoda, pouzdanosti, korištenja resursa, pravila i politika, rasporeda i učinkovitosti otpreme, propusnost procesa itd. (Ueda, 2010).

Koristeći tehnike iz drugih matematičkih znanosti, kao što su modeliranje, statistika i optimizacija, operacijska istraživanja dolaze do optimalnih ili gotovo optimalnih rješenja za probleme donošenja odluka. Zbog svog naglaska na praktičnim primjenama, operacijsko istraživanje preklapa se s mnogim drugim disciplinama, posebice industrijskim inženjerstvom i logistikom, što ga čini sastavnim dijelom njihovih sustava upravljanja znanjem (engl. *Knowledge Management Systems* - KMS).

6.1 Strateško logističko planiranje

Prema Robinsonu (2004) OR se uglavnom bavi interakcijom između planiranih tehničkih i ljudskih sustava. Karakterizira ih varijabilnost, međuovisnost među komponentama te strukturna i bihevioralna složenost. Da bi se upravljalo ovim karakteristikama, osmišljen je širok niz metoda za njihovo prikladno rješavanje kao cjeline. Koriste se u strateškom planiranju za:

- omogućavanje složene analize što ako;
- upravljanje složenošću: međuovisnost + varijabilnost + dinamika;
- uključivanje manje troškova i smetnji u procesu nego eksperimentiranje sa stvarnim sustavom;
- fokus na detalje;

- poboljšanje razumijevanja sustava;
- poboljšanje komunikacije između menadžmenta i stručnjaka.

Strateško logističko planiranje (slika 6.1) obuhvaća sve aktivnosti koje je potrebno izvesti na strateškoj, taktičkoj i operativnoj razini kako bi se osiguralo upravljanje potpunom kvalitetom (engl. *Total Quality Management* - TQM) (Ciampa, 1992) i proizvodnja točno na vrijeme (engl. *Just in Time* - JIT) (Britannica, 2023).

Slika 4.21 Strateško logističko planiranje.

6.2 Šest sigma

U strateškom logističkom planiranju referentni model SCOR (AIMS, 2021) pomaže tvrtkama da procijene i usavrše upravljanje opskrbnim lancem za pouzdanost, dosljednost i učinkovitost. Prepoznaje 6 glavnih poslovnih procesa — planiranje, *sourcing*, proizvodnja, isporuka, povrat i omogućavanje.

SCOR proces planiranja obuhvaća sve aktivnosti povezane s razvojem planova za upravljanje i poboljšanje opskrbnog lanca. Kontinuirani napori da se postignu stabilni i predvidljivi rezultati procesa smanjenjem varijacija procesa (6-sigma) od vitalne su važnosti za poslovni uspjeh.

DMAIC i DMADV

Proizvodni procesi (*sourcing*, proizvodnja, isporuka, omogućavanje, kao i rukovanje povratima) imaju karakteristike koje se mogu definirati, mjeriti, analizirati, poboljšati i kontrolirati. Stoga ove faze čine metodologiju upravljanja proizvodnim procesom, skraćeno DMAIC.

Neki praktičari kombinirali su ideje 6-sigma s *lean* proizvodnjom kako bi stvorili metodologiju nazvanu *Lean Six Sigma* (Wheat i dr., 2003). Metodologija *Lean Six Sigma* smatra *lean* proizvodnju (JIT proizvodnja), koja se bavi učinkovitošću procesa, i 6-sigmu, s fokusom na smanjenje varijacija i otpada, kao komplementarne discipline koje promiču poslovnu i operativnu izvrsnost.

Metodologija DMADV (definiraj, mjeri, analiziraj, dizajniraj i provjeri), također poznata kao DFSS (engl. *Design for Six Sigma*, tj. hrv. Dizajn za šest sigma), u skladu je s KBE (engl. *Knowledge Based Engineering*, hrv. Inženjering temeljen na znanju). Faze DFSS metodologije (Chowdhury, 2002) su:

1. Definirajte ciljeve dizajna koji su u skladu sa zahtjevima kupaca i strategijom poduzeća.
2. Mjerite i identificirajte karakteristike koje su ključne za kvalitetu (engl. *Critical to Quality* - CTQ), mjerite mogućnosti proizvoda, kapacitet proizvodnog procesa i mjerite rizike.
3. Analizirajte kako biste razvili i dizajnirali alternative.
4. Dizajnirajte poboljšanu alternativu, najprikladniju za analizu u prethodnom koraku.
5. Provjerite dizajn, postavite probne radove, implementirajte proizvodni proces i predajte ga vlasniku (vlasnicima) procesa.

Projekti poboljšanja poslovanja *Six Sigma* (Tennant, 2001), inspirirani ciklusom „Planiraj–Radi–Proučavaj–Djeluj" (engl. Plan–Do–Study–Act) W. Edwardsa Deminga (Tague, 2005), ovisno o svojoj prirodi, slijede jednu od gore navedenih metodologija, a svaka ima pet faza:

1. DMAIC se koristi za projekte usmjerene na poboljšanje postojećeg poslovnog procesa.
2. DMADV se koristi za projekte usmjerene na stvaranje novih proizvoda ili dizajna procesa.

6.3 Poslovna inteligencija

Poslovna inteligencija (engl. *Business Intelligence* - BI) obuhvaća sve strategije i tehnologije koje poduzeća koriste za analizu podataka o prošlim i trenutnim poslovnim informacijama (Tableau, 2023.). Podržavaju ga sustavi upravljanja znanjem (KMS) koji predstavljaju dio logističkih informacijskih sustava (engl. *Logistics Information Systems* - LIS) koji stručnjacima iz različitih područja omogućavaju savjetovanje i podršku različitim razinama menadžmenta.

Poslovna analitika

Poslovna analitika (engl. *Business Analytics* - BA) je proces temeljen na BI-u koji omogućuje nove uvide u poslovni proces i bolje strateško odlučivanje za budućnost. Potječe iz rudarenja podataka (engl. *Data Mining* - DM) koji je proces pronalaženja anomalija, uzoraka i korelacija u većim skupovima podataka, kako bi se predvidjeli rezultati.

BA proces sadrži sljedeće:

1. Agregacija podataka: prije analize, podaci se prvo moraju prikupiti, organizirati i filtrirati, bilo putem dobrovoljnih podataka ili transakcijskih zapisa.
2. Rudarenje podataka: razvrstava velike skupove podataka koristeći baze podataka, statistiku i strojno učenje za prepoznavanje trendova i uspostavljanje odnosa.
3. Identifikacija pridruživanja i slijeda: identifikacija predvidljivih radnji koje se izvode zajedno s drugim radnjama ili slijede jedna drugu.
4. Rudarenje tekstualnih podataka: istražuje i organizira velike, nestrukturirane tekstualne skupove podataka u svrhu kvalitativne i kvantitativne analize.
5. Predviđanje: analizira povijesne podatke iz određenog razdoblja kako bi se napravile informirane procjene koje su prediktivne u određivanju budućih događaja ili ponašanja.
6. Prediktivna analitika: prediktivna poslovna analitika koristi različite statističke tehnike za stvaranje prediktivnih modela, koji izvlače informacije iz skupova podataka, identificiraju obrasce i daju prediktivnu ocjenu za niz organizacijskih ishoda.

7. Optimizacija: nakon što se identificiraju trendovi i naprave predviđanja, tvrtke mogu koristiti tehnike simulacije za testiranje najboljih scenarija.
8. Vizualizacija podataka: pruža vizualne prikaze kao što su dijagrami i grafikoni za jednostavnu i brzu analizu podataka.

Planiranje prodaje i poslovanja

Planiranje prodaje i operacija (engl. *Sales and operations planning* - SOP) je fleksibilan alat za predviđanje i planiranje proizvodnih aktivnosti. SOP koraci:

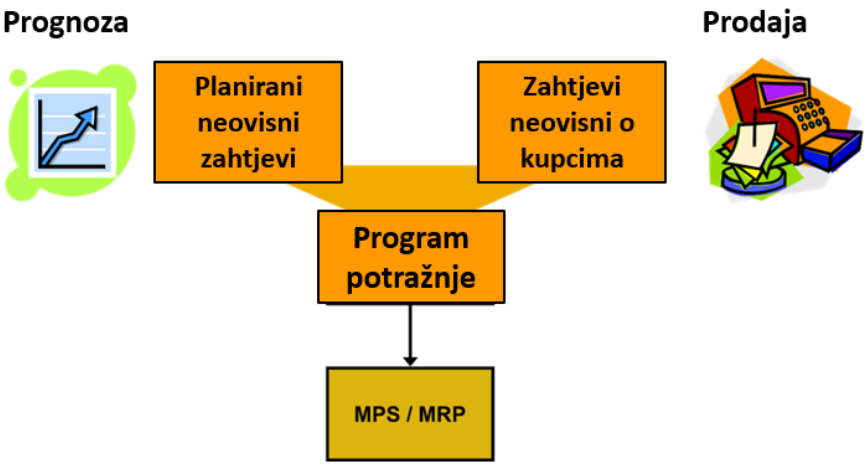
1. plan prodaje,
2. plan proizvodnje i
3. planiranje kapaciteta.

SOP radi na podacima iz različitih izvora informacija u cijeloj tvrtki: prodaja, marketing, proizvodnja, računovodstvo, ljudski resursi i nabava. Obično ih osiguravaju odgovarajući odjeli putem sustava za planiranje resursa poduzeća (engl. *enterprise resource planning* - ERP).

Dok SOP djeluje na strateškoj razini, ERP djeluje na taktičkoj razini logističkog informacijskog sustava tvrtke. Njima se pridružuje i *Demand Management* (upravljanje potražnjom) program koji povezuje strateško planiranje prodaje i operacija (SOP) i detaljno planiranje proizvodnje (*Master Production Scheduling* (hrv. glavno planiranje proizvodnje) / *Material Requirements Planning* (hrv. planiranje potreba za materijalima)) na operativnoj razini. Ovdje prije spomenuto planiranje i simulacija stupaju na scenu kako bi se napravio izvediv i optimalan plan proizvodnje.

Program upravljanja potražnjom (slika 6.2) sastoji se od dvije vrste prognoza:

1. planirani neovisni zahtjevi (engl. *planned independent requirements* - PIR) od projiciranih količina prodaje na temelju marketinga i
2. zahtjevi neovisni o kupcima (engl. *customer independent requirements* - CIR) iz podataka na temelju postojećih i planiranih prodajnih naloga.



Slika 4.22Upravljanje potražnjom.

Primjer SOP-a

Poduzeće trguje s nekoliko vrsta proizvoda u svojoj prodajnoj mreži. Prodajne transakcije pohranjuju se centralno kako bi se moglo pratiti stanje zaliha, kao i izvršiti analizu prodaje za upravljanje potražnjom. Zapisuju se u CSV (engl. *Comma Separated Values*) formatu, koji je lako obraditi u ERP sustavu tvrtke, kao i u odjelu za analitiku, koji koristi proračunske tablice.

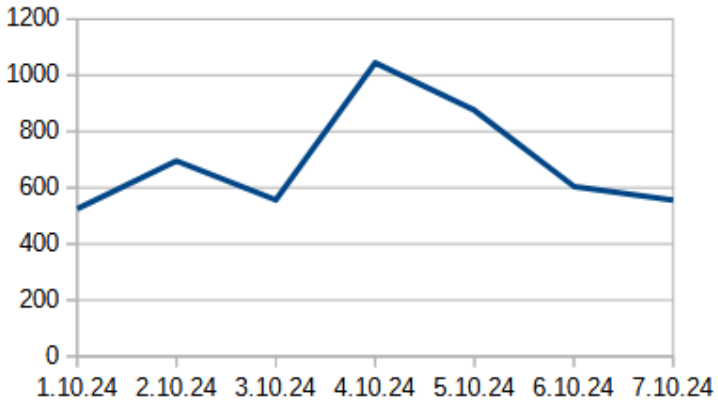
U analizi prodaje, prodajne transakcije se inicijalno filtriraju kako bi se utvrdilo jesu li potpune i ispravno formatirane. Tek tada su spremni za statističku procjenu, budući da bi u protivnom nedostajući ili loše oblikovani podaci mogli rezultirati pogrešnim tumačenjem rezultata.

Datum	ID Prodavača	ID Kupca	ID Transakcije	ID Proizvoda	Cijena proizvoda
1.10.24	1	12	1	101	195,00 €
1.10.24	1	12	1	102	45,00 €
1.10.24	1	12	1	103	35,00 €
1.10.24	2	14	2	104	55,00 €
1.10.24	2	14	3	101	195,00 €
2.10.24	3	15	4	105	85,00 €
2.10.24	3	15	4	101	195,00 €
2.10.24	3	15	4	103	35,00 €
2.10.24	3	16	5	104	55,00 €
2.10.24	1	17	6	101	195,00 €
2.10.24	1	17	6	102	45,00 €
2.10.24	1	17	6	105	85,00 €
3.10.24	2	18	7	106	35,00 €
3.10.24	2	18	7	107	65,00 €
3.10.24	2	18	7	108	86,00 €
3.10.24	4	19	8	105	85,00 €
3.10.24	4	19	8	101	195,00 €
3.10.24	4	19	8	103	35,00 €
3.10.24	4	19	9	104	55,00 €
4.10.24	5	20	10	105	110,00 €
4.10.24	5	20	10	106	125,00 €
4.10.24	5	20	10	104	55,00 €
4.10.24	5	20	10	101	195,00 €
4.10.24	1	21	11	102	45,00 €
4.10.24	1	21	11	105	85,00 €
4.10.24	1	21	12	106	35,00 €
4.10.24	3	12	13	103	35,00 €
4.10.24	3	12	13	104	55,00 €
4.10.24	3	12	13	105	110,00 €
4.10.24	3	12	13	101	195,00 €
5.10.24	1	22	14	107	35,00 €
5.10.24	1	22	14	108	25,00 €

Slika 4.23Podaci o tjednoj prodaji.

Obično se analitikom omogućavaju različiti senzibilni uvidi u prikupljene "sirove podatke" (slika 6.3). To se može postići pivot tablicama koje omogućuju grupiranje podataka prema odabranim atributima i provođenje statističke analize. U principu, bilo koji atribut (stupac) ulaznih podataka može se smatrati pivotom. Stoga često govorimo o "podatkovnoj kocki" više dimenzija. Budući da ne možemo grafički prikazati više od dvije ili tri dimenzije, najjednostavniji, ali obično najkorisniji, prikazi ulaznih podataka formiraju se od dva ili tri pivot atributa. Na temelju zadanog uzorka podataka, u nastavku su navedeni neki primjeri pivot tablica.

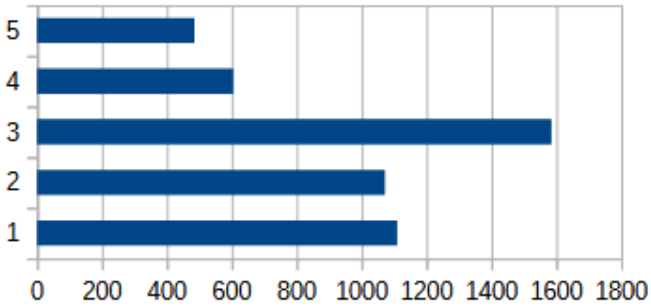
Date	Vsota - Product Price
1.10.24	525,00 €
2.10.24	695,00 €
3.10.24	556,00 €
4.10.24	1.045,00 €
5.10.24	875,00 €
6.10.24	605,00 €
7.10.24	556,00 €
Skupaj Rezul	**4.857,00 €**



Slika 4.24Statistika prodaje po radnim danima.

Statistika prodaje po danima u tjednu ili mjesecu (slika 6.4) omogućuje uvid u sezonska kretanja.

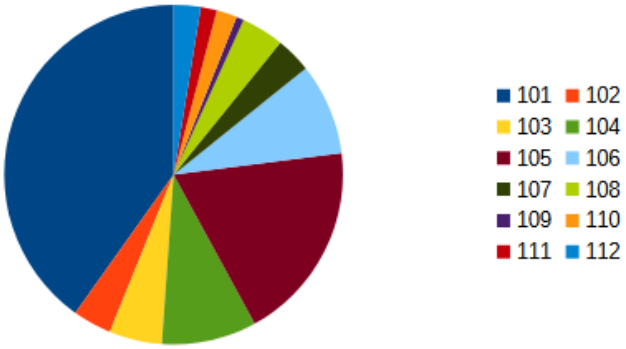
Seller ID	Vsota - Product Price
1	1.110,00 €
2	1.072,00 €
3	1.585,00 €
4	605,00 €
5	485,00 €
Skupaj Rezul	**4.857,00 €**



Slika 4.25Statistika prodaje po prodajnim mjestima.

Statistika prodaje po prodajnom mjestu (slika 6.5) utvrđuje koji su uredi najzaposleniji i/ili ostvaruju najviše prihoda.

Product ID	Vsota - Product Price
101	1.950,00 €
102	180,00 €
103	245,00 €
104	440,00 €
105	925,00 €
106	425,00 €
107	165,00 €
108	197,00 €
109	35,00 €
110	95,00 €
111	75,00 €
112	125,00 €
Skupaj Rezult	**4.857,00 €**



Slika 4.26Statistika prodaje po proizvodima.

Statistika prodaje po proizvodima (slika 6.6) određuje proizvode koji su najtraženiji ili predstavljaju značajan udio u portfelju.

Date	Seller ID	Customer ID	Transaction ID	Product ID	Vsota - Produ
1.10.24	1	12	1	101	195,00 €
				102	45,00 €
				103	35,00 €
	2	14	2	104	55,00 €
			3	101	195,00 €
2.10.24	1	17	6	101	195,00 €
				102	45,00 €
				105	85,00 €
	3	15	4	101	195,00 €
				103	35,00 €
				105	85,00 €
		16	5	104	55,00 €
3.10.24	2	18	7	106	35,00 €
				107	65,00 €
				108	86,00 €
	4	19	8	101	195,00 €
				103	35,00 €
				105	85,00 €
			9	104	55,00 €

Slika 4.27Pregled transakcija.

Pregled transakcija (slika 6.7) po danu, prodajnom uredu, transakciji i kupcu nudi strukturirani uvid u podatke koji je koristan prilikom rješavanja upita za određeni proizvod ili prodajnu transakciju.

6.4 Sustavi za podršku odlučivanju

Sustavi za podršku odlučivanju (engl. *Decision Support System* - DSS) su interaktivni sustavi koji pomaže donositeljima odluka da koriste podatke i modele za rješavanje nestrukturiranih ili djelomično strukturiranih problema. Ekspertni sustav (ES) je aplikacijski program ili okruženje, koje učinkovito podržava rješavanje problema u specijaliziranom problemskom području, zahtijevajući stručno znanje i vještine.

Uz statističke i simulacijske metode koje se koriste u BA, DSS često uključuje modele koji omogućuju donošenje odluka, na temelju skupa razlikovnih kriterija. Ovi modeli mogu biti jednostavni (npr. tablice odlučivanja, stabla itd.) koji vode do jednog ispravnog rješenja. S

druge strane, kada se radi o višestrukim (moguće sukobljenim) kriterijima i višestrukim rješenjima, nužan je razrađeniji model. Prikladan model koji se često koristi u profesionalnom i osobnom životu je višekriterijski model odlučivanja.

Višekriterijsko odlučivanje

Višekriterijsko odlučivanje (engl. *multi criteria decision making* - MCDM) ili analiza višekriterijskog odlučivanja (engl. multiple-criteria decision analysis - MCDA) poddisciplina je OR-a koja eksplicitno procjenjuje više (eventualno) sukobljenih kriterija u donošenju odluka nad skupom mogućih rješenja ili varijanti.

MCDM model odlučivanja sadrži sljedeće:

- Kriteriji – parametri ulaznih varijanti, kritični za naš dizajn.
- Ponderi – relativna važnost odabranih kriterija.
- Funkcija korisnosti – funkcija koja kombinira ponderirane parametre varijanti u vrijednost prikladnosti.
- Podaci – podaci koji predstavljaju varijante; unos podataka u naš MCDM model.

Vrste podataka u MCDM mogu biti:

- Kvantitativni – predstavljaju vrijednosti koje se kao takve mogu usporediti.
- Kvalitativni – predstavljanje relativnih usporednih vrijednosti (npr. visoka, ugodna, niska temperatura, itd.) koje je potrebno kvantificirati kako bi se dobile jedinstvene vrijednosti.
- Binarni – predstavljanje binarnih kriterija; svojstvo ispunjeno (1) ili ne (0).

Postupak višekriterijskog odlučivanja:

1. Predstavljanje varijanti (V) njihovim karakterističnim parametrima (P): $\{V_i (P_{i,1}; P_{i,2}; \dots P_{i,n}); i=1\dots m\}$.
2. Normalizacija parametara izračunavanjem relativne lokalne ocjene $p_{i,j}$ za svaki $P_{i,j}$ $(j=1\dots n)$, u odnosu na maksimalnu vrijednost j^{-tog} parametra $P_{i,j}$ iz svih i uzoraka:
   a) $p_{i,j} = P_{i,j} / \max \{P_{i,j}\}$ ako je veća vrijednost $P_{i,j}$ korisnija.
   b) $p_{i,j} = 1 - P_{i,j} / \max \{P_{i,j}\}$ ako je manja vrijednost Pi,j korisnija.

3. Ocjene se ponderiraju prema preferencijama: $x_{i,j} = p_{i,j} * U_j$ za svaki j=1...n, po ponderima U_j koje treba zbrojiti do 1, tj. 100%.

4. Ponderirane ocjene svih varijanti se zbrajaju: $X_i = \sum x_{i,j}$ za svaki i=1...m kako bi se dobile kompozitne ocjene prema našoj funkciji korisnosti.

5. Izabrana je najbolja varijanta $Y = \max\{X_i\}$.

Primjer MCDM-a

Prilikom odabira nove opreme u poduzeću često moramo provoditi višekriterijsko odlučivanje. Razmotrimo primjer odabira najisplativije (ispod 300 EUR) mobilne platforme s operativnim sustavom Android za našu tvrtku. U tablici 6.1 navedeni su primjerci koji su izabrani na temelju upita među zaposlenicima kako bi suzili naš asortiman. Za svaki od njih navedeni su parametri koji su odabrani kao najrelevantniji. U nastavku se podaci odabranih parametara normaliziraju kako bi se dobile usporedive vrijednosti, ponderiraju kako bi se naglasile značajnije vrijednosti i zbrajaju kako bi se dobile ocjene za odabrane uzorke.

Tablica 4.1MCDM za isplativu Android mobilnu platformu.

PARAMETERS									
	Price (€)	Grade*	Performance			Properties			Camera
Model		(1-10)	proc.speed (GHz)	RAM (GB)	int.mem. (GB)	weight (g)	size (mm3)	bat.cap. (mAh)	(MP)
Honor Magic Lite 5	263 €	2	2,2	6	128	175	94344	5100	64
Honor X7a	210 €	2,5	2,3	4	128	196	106442	6000	50
Samsung A34	297 €	1,8	2,6	8	256	199	103300	5000	48
Redmi Note 12 Pro	240 €	1,9	2,6	6	128	187	99104	5000	50
Redmi Note 12 S	224 €	1,7	2,05	8	256	176	95715	5000	108

*Vir: www.testberichte.de

PARAMETER WEIGHS									
	Price (€)	Grade	Performance			Properties			Camera
			proc.speed (GHz)	RAM (GB)	int.mem. (GB)	weight (g)	size (mm3)	bat.cap. (mAh)	(MP)
			10 %	5 %	5 %	10 %	10 %	10 %	
Utež	20 %	10 %	20 %			30 %			20 %

NORMALIZED PARAMETERS									
	Price (€)	Grade*	Performance			Properties			Camera
Model		(1-10)	proc.speed (GHz)	RAM (GB)	int.mem. (GB)	weight (g)	size (mm3)	bat.cap. (mAh)	(MP)
Honor Magic Lite 5	0,11	0,20	0,85	0,75	0,50	0,12	0,11	0,85	0,59
Honor X7a	0,29	0,00	0,88	0,50	0,50	0,02	0,00	1,00	0,46
Samsung A34	0,00	0,28	1,00	1,00	1,00	0,00	0,03	0,83	0,44
Redmi Note 12 Pro	0,19	0,24	1,00	0,75	0,50	0,06	0,07	0,83	0,46
Redmi Note 12 S	0,25	0,32	0,79	1,00	1,00	0,12	0,10	0,83	1,00

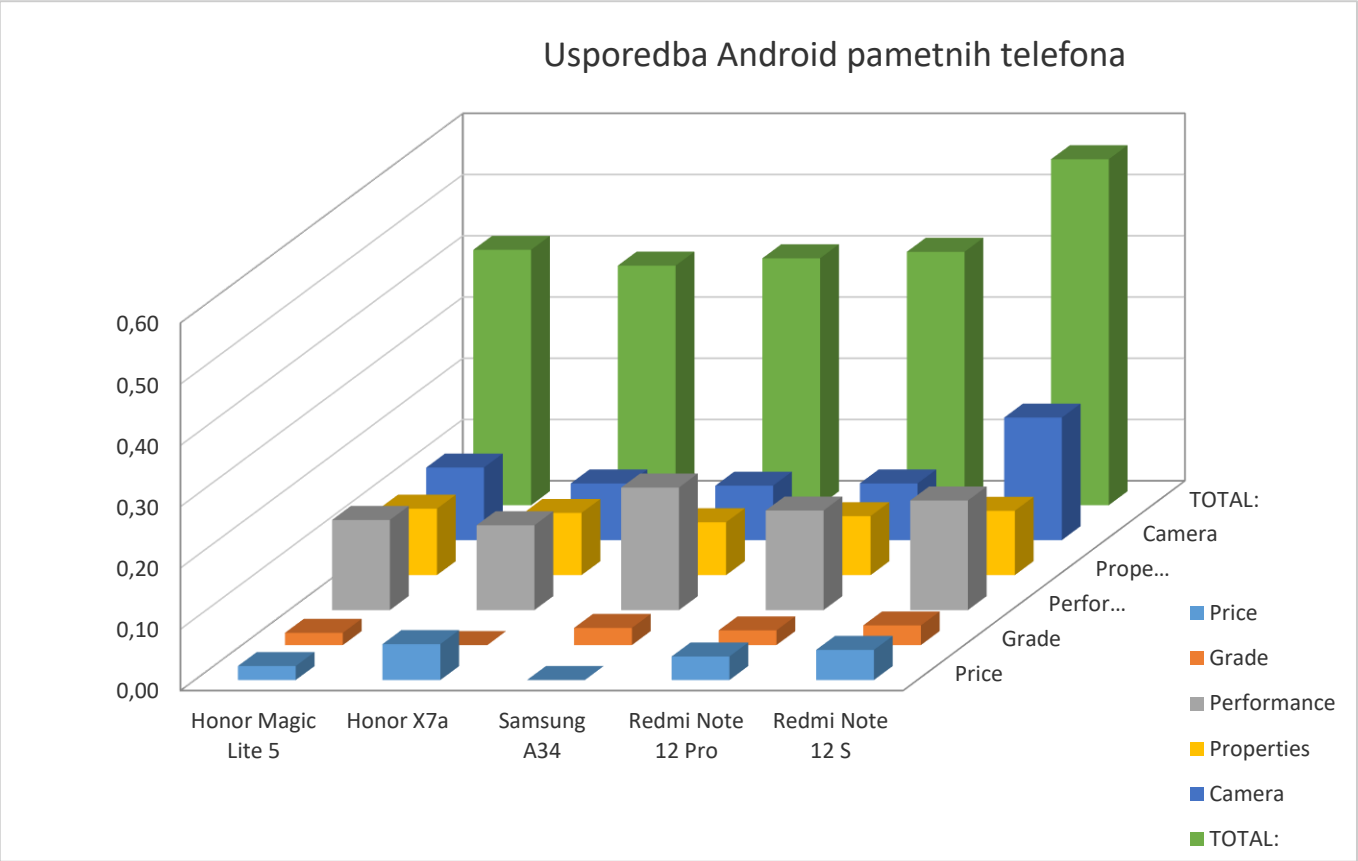
FINAL PARAMETER ASSESSMENT									
	Price (€)	Grade*	Performance			Properties			Camera
Model		(1-10)	proc.speed (GHz)	RAM (GB)	int.mem. (GB)	weight (g)	size (mm3)	bat.cap. (mAh)	(MP)
Honor Magic Lite 5	0,02	0,02	0,08	0,04	0,03	0,01	0,01	0,09	0,12
Honor X7a	0,06	0,00	0,09	0,03	0,03	0,00	0,00	0,10	0,09
Samsung A34	0,00	0,03	0,10	0,05	0,05	0,00	0,00	0,08	0,09
Redmi Note 12 Pro	0,04	0,02	0,10	0,04	0,03	0,01	0,01	0,08	0,09
Redmi Note 12 S	0,05	0,03	0,08	0,05	0,05	0,01	0,01	0,08	0,20

Krajnji rezultat naše analize je sažeta tablica (tablica 6.2) i eventualno grafikon (slika 6.8) koji ukratko prikazuju proces donošenja odluka i predstavlja najbolji izbor kao i prednosti i slabosti pojedinih varijanti. Često najbolje ocijenjeni primjerak nije onaj koji se najbolje pokazao u svim

kategorijama, već onaj koji u prosjeku najbolje odgovara našim kriterijima odabira, kao i vaganju parametara. To je ujedno i glavna snaga metode MCDM, budući da bi nas izbor prema bilo kojem pojedinačnom parametru mogao dovesti u zabludu.

Tablica 4.2MCDM sažetak za naš primjer odabira mobilne platforme.

FINAL PARAMETER ASSESSMENT						
Model	**Price**	**Grade**	**Performance**	**Properties**	**Camera**	**TOTAL:**
Honor Magic Lite 5	0,02	0,02	0,15	0,11	0,12	0,42
Honor X7a	0,06	0,00	0,14	0,10	0,09	0,39
Samsung A34	0,00	0,03	0,20	0,09	0,09	0,40
Redmi Note 12 Pro	0,04	0,02	0,16	0,10	0,09	0,41
Redmi Note 12 S	0,05	0,03	0,18	0,10	0,20	0,56



Slika 4.28 Izbor najbolje mobilne platforme.

Najbolji izbor: 0,56 (Redmi Note 12 S).

6.5 Inženjerstvo temeljeno na znanju

Inženjerstvo temeljeno na znanju (engl. *Knowledge Based Engineering* - KBE) je inženjerska metodologija za sustavnu integraciju inženjerskog znanja u sustav dizajna (Andersson i dr., 2011).

Životni ciklus upravljanja iskustvom

Potreba za prikupljanjem, upravljanjem i korištenjem znanja o dizajnu i automatiziranjem procesa koji su jedinstveni za proizvođačevo iskustvo razvoja proizvoda dovela je do razvoja tehnologije inženjerstva temeljenog na znanju (KBE) (Prasad, 2005). KBE je namijenjen obogaćivanju institucionalnog znanja upravljanjem iskustvom. Faze upravljanja iskustvom, prema (Anderssonu i dr, 2011) su:

1. Identificiraj: odabrana je nesukladnost sa željenim stanjem koja se pojavljuje u procesu proizvodnje zbog loše definiranog proizvoda ili procesa.
2. Uhvati: iskustvo sa svojim svojstvima je uhvaćeno.
3. Analizirajte: napravljena je analiza temeljnog uzroka snimljenog iskustva kako bi se identificirala odgovarajuća strategija lijeka i njezina ponovna uporaba kako bi se spriječile ponovne anomalije.
4. Pohrani: uvidi iz analize arhiviraju se s iskustvom.
5. Traži i dohvati: iskustvo se traži i dohvaća.
6. Upotreba: koristi se element iskustva.
7. Ponovno korištenje: zaključuje se ciklus upravljanja znanjem i započinje novi.

KBE je općenito dopunjeno daljnjim disciplinama, čije bliže razmatranje prelazi opseg ovog poglavlja:

- Računalno potpomognuto upravljanje projektima (PS).
- Računalno potpomognuto projektiranje (CAD), proizvodnja (CAM) i robotika (CIM).
- Računalno simulacijsko modeliranje i analiza (SMA).
- Računalno potpomognuto detaljno planiranje proizvodnje (MPS/MRP).

6.6 Zaključak

Kao što je predstavljeno u ovom poglavlju, glavne primjene BI-a u korporativnom upravljanju odnose se na poslovnu analitiku (BA) i sustave za podršku odlučivanju (DSS). Obično se nazivaju operacijskim istraživanjem (OR). Uz temeljne BI tehnike, opisane u ovom poglavlju, u poglavljima 3 i 4 o upravljanju podacima i simulacijskom modeliranju i analizi (SMA) dana su neka dodatna razmatranja o prikupljanju podataka, manipulaciji i prezentiranju, koja također podržavaju donošenje odluka. Ukratko, BI aplikacije nalaze se u sustavima upravljanja znanjem (KMS), koji se sastoje od:

- sustava za podršku odlučivanju (DSS),
- poslovne analitike (BA) kao nadogradnje Data Mininga (DM) i
- inženjeringa temeljenog na znanju (KBE) kao nadogradnja računalno potpomognutog inženjerstva (CAE).

Na temelju BA, DSS i SMA rezultata osmišljena su iskustva koja unapređuju institucionalno znanje i konstituiraju njihove ekspertne sustave temeljene na znanju (KBS). Kao što je pokazano u Gumzej i dr. (2023), KBE ih može upotrijebiti za uvođenje načela „naučenih lekcija" u upravljanje poboljšanjima poduzeća putem strateškog planiranja logistike.

Literatura 6. poglavlja

- AIMS (2021). SCOR - Supply Chain Operations Model. [available at: https://aims.education/study-online/supply-chain-operations-reference-model-scor/, access June 20, 2023]
- Ciampa, D. (1992). Total Quality: A User's Guide for Implementation, Addison-Wesley.
- Britannica, T. Editors of Encyclopaedia (2023). Just-in-time manufacturing, Encyclopedia Britannica. [available at: https://www.britannica.com/topic/just-in-time-manufacturing, access June 20, 2023]
- Tennant, G. (2001). SIX SIGMA: SPC and TQM in Manufacturing and Services, Gower Publishing, Ltd.
- Wheat B. & Mills C. & Carnell M. (2003). Leaning into Six Sigma: a parable of the journey to Six Sigma and a lean enterprise, McGraw-Hill.

- Tague, N.R. (2005). Plan–Do–Study–Act cycle, The quality toolbox (2nd ed.), ASQ Quality Press, pp. 390–392.
- Chowdhury, S. (2002). Design for Six Sigma: The revolutionary process for achieving extraordinary profits, Prentice Hall,
- Ueda, M. (2010). How to Market OR/MS Decision Support. International Journal of Applied Logistics (IJAL), 1(2), 23-36.
- Tableau (2023). Comparing Business Intelligence, Business Analytics and Data Analytics. [Available from: https://www.tableau.com/learn/articles/business-intelligence/bi-business-analytics, access June 20, 2023]
- Prasad, B. (2005). Knowledge Technology, What Distinguishes KBE From Automation. COE NewsNet – June 2005. [available at: https://web.archive.org/web/20120324223130/http://legacy.coe.org/newsnet/Jun05/knowledge.cfm, access June 20, 2023]
- Robinson, S. (2004). Simulation: The Practice of Model Development and Use, Wiley.
- Gumzej, R., Kramberger, T., Dujak, D. (2023). A knowledge base for strategic logistics planning. In: Dujak, Davor (edt.). Proceedings of the 23rd International Scientific Conference Business Logistics in Modern Management: October 5-6, 2023, Osijek, Croatia. Osijek: Josip Juraj Strossmayer University of Osijek, Faculty of Economics and Business, pp. 317-330, illustr. Business logistics in modern management (Online). ISSN 1849-6148. https://blmm-conference.com/past-issues/.
- Andersson, P. & Larsson, T. & Ola, I. (2011). A case study of how knowledge based engineering tools support experience re-use. In Research into Design – Supporting Sustainable Product Development, Chakrabarti, A. (Edt.), Research Publishing, Indian Institute of Science, Bangalore, India.

7. Statistička obrada podataka SPSS

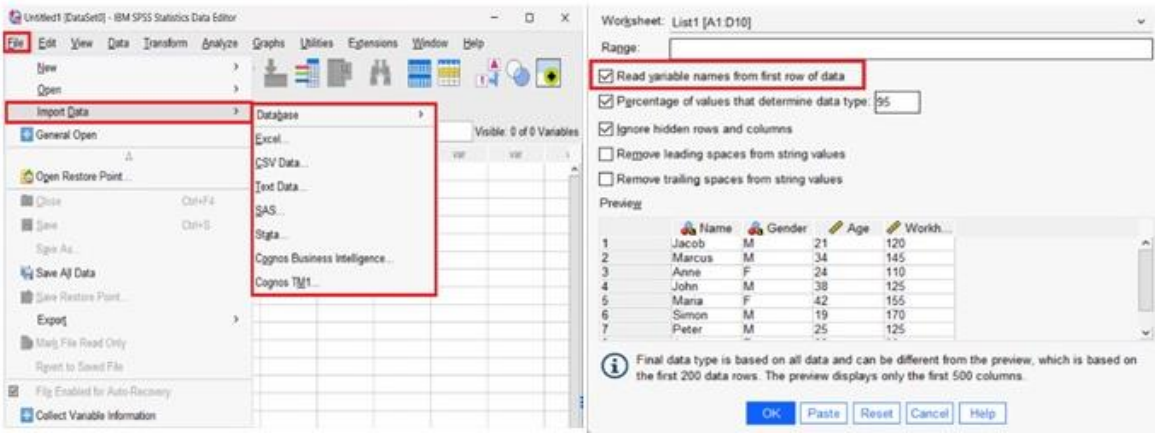
Do sada ste već stekli temeljno razumijevanje statistike, manipulacije podacima, uspostavljanja simulacije, modeliranja i analize unutar logističkih opskrbnih lanaca, zajedno s izravnim metodama linearne regresije. Dok statistika nudi raznolik raspon modela i tehnika za

poboljšanje vaših napora optimizacije, provođenje analize i prepoznavanje potencijalnih poboljšanja, možda ste primijetili da kako složenost analiziranih podataka i izračuna raste, tradicionalni pristupi mogu postati sve zamršeniji i izazovniji za računanje. Kako zamršenost vaših podataka i izračuna raste, konvencionalne metode mogu zastarjeti i, u nekim slučajevima, ugroziti pouzdanost vaših rezultata. Kako bi se to prepoznalo, statistika koristi razne softverske programe koji automatiziraju analizu i interpretaciju prikupljenih podataka, a istovremeno pružaju mnoštvo modela i funkcija za osiguranje pouzdanih rezultata. Jedan takav softver je IBM-ov SPSS, koji će biti ključni alat u ovom poglavlju. U ovom poglavlju pružit ćemo sažeti uvod u primarnu upotrebu softvera SPSS, istražujući njegove funkcionalnosti i praktične primjene. Nakon početnog uvoda slijedi praktična primjena programa kroz četiri temeljna testa za izračun rezultata: T-test, korelacije, Hi-kvadrat i ANOVA. Kako bismo vam olakšali učenje, predstavit ćemo vam jednostavne probleme i njihova rješenja kako biste se lakše upoznali s ovim testovima.

7.1 Osnove IBM-ovog SPSS-a

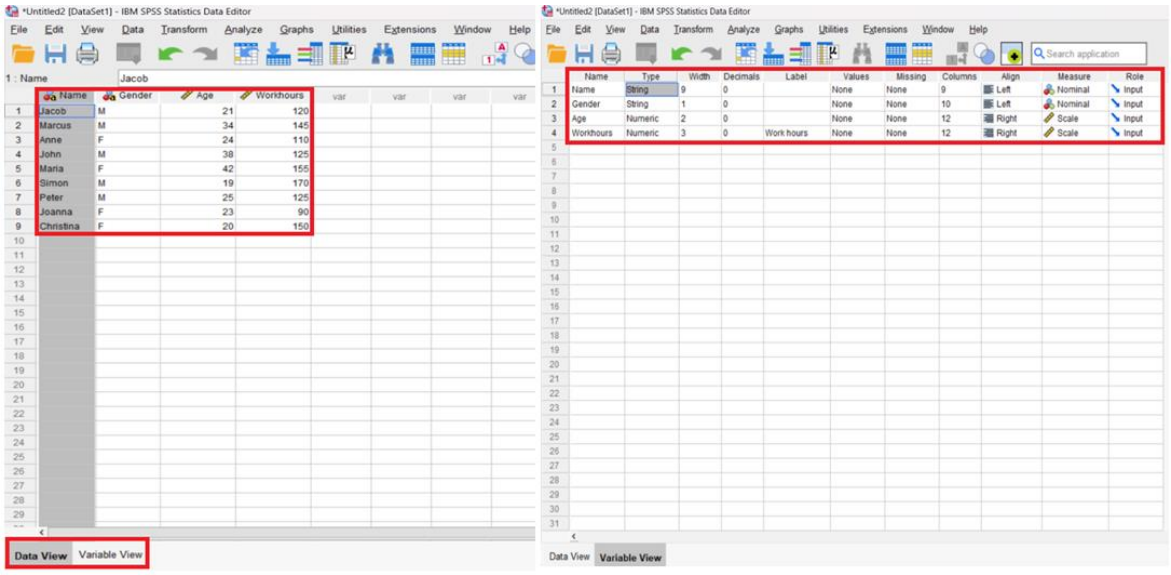
Možda ste već imali iskustva sa SPSS softverom. Međutim, ako ipak trebate, dopustite da vam ponudimo kratki uvod. SPSS, poput svog općepriznatijeg pandana Excela, olakšava manipulaciju podacima, analizu i vizualizaciju. Ipak, za razliku od Excela, koji ponekad može biti naporan i složen u programiranju funkcija, SPSS nudi korisničko sučelje za statističku analizu (IBM, 2021.). Nudi niz funkcija i metodologija za učinkovito rukovanje vašim podacima. Dok se SPSS softver ističe u pružanju opsežnih mogućnosti statističke analize, upravljanje manipulacijom podacima i konfiguriranje početnih postavki za analizu ponekad može biti izazovno (IBM, 2021.). Stoga ćemo se pozabaviti osnovama uvoza podataka i pripreme podataka za naknadne statističke testove.

S obzirom na široku upotrebu Excela za rukovanje numeričkim podacima, vaši se podaci mogu dobiti ili pripremiti u proračunskoj tablici programa Excel. Srećom, SPSS može uvesti podatke iz različitih formata datoteka u svoje proračunske tablice. Nakon što pripremite finalizirane Excel proračunske tablice, otvorite softver SPSS. Na početnom zaslonu idite na karticu "Datoteka" i odaberite "Uvezi podatke". U sljedećem prozoru možete odabrati format podataka koji namjeravate uvesti (pogledajte sliku 7.1). Slijedeći ovaj korak, pronađite pripremljenu datoteku, odaberite je i prijeđite na sljedeći prozor. Ovaj prozor će od vas tražiti da konfigurirate dodatne postavke. Ako ste već uključili nazive stupaca u prvi red vaših podataka, odaberite opciju "Pročitaj nazive varijabli iz prvog retka podataka" (pogledajte sliku 7.1), a zatim kliknite "Završi" da bi se podaci pojavili u proračunskoj tablici (IBM, 2021).



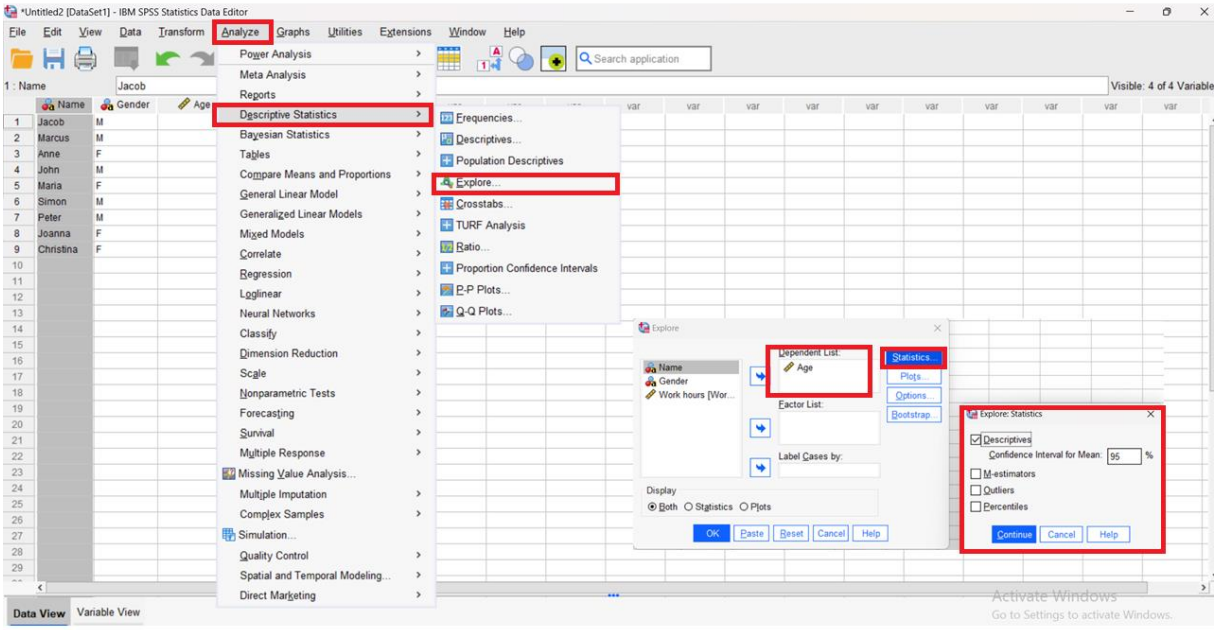
Slika 4.29SPSS postavke uvoza podataka.

Sada kada imamo podatke u našoj proračunskoj tablici, primijetit ćete jasnu razliku u prezentaciji u usporedbi s Excelom. SPSS kategorizira podatke u dvije primarne vrste, svaka s dvije dodatne podvrste. Kao što je prikazano na slici 7.2, podaci se mogu klasificirati kao numerički ili kategorički. Numerički podaci sastoje se od brojeva i mogu se kategorizirati kao diskretni (s ograničenim opcijama) ili kontinuirani (nude beskonačne mogućnosti). S druge strane, kategorički podaci sastoje se od riječi i mogu se dalje razlikovati kao redni (imaju hijerarhiju) ili nominalni (bez hijerarhije). Ovisno o prirodi vaših podataka, možda ćete morati konfigurirati varijable kako bi se uskladile sa željenom analizom. U većini slučajeva, SPSS će automatski prikladno kategorizirati varijable. Pretpostavimo da želite izvršiti daljnju manipulaciju tipovima podataka. U tom slučaju možete pristupiti opciji "Prikaz" i pod "Prikaz varijable" prilagoditi varijabilne informacije kao što su naziv, vrsta, širina, mjera i više (IBM, 2021.).



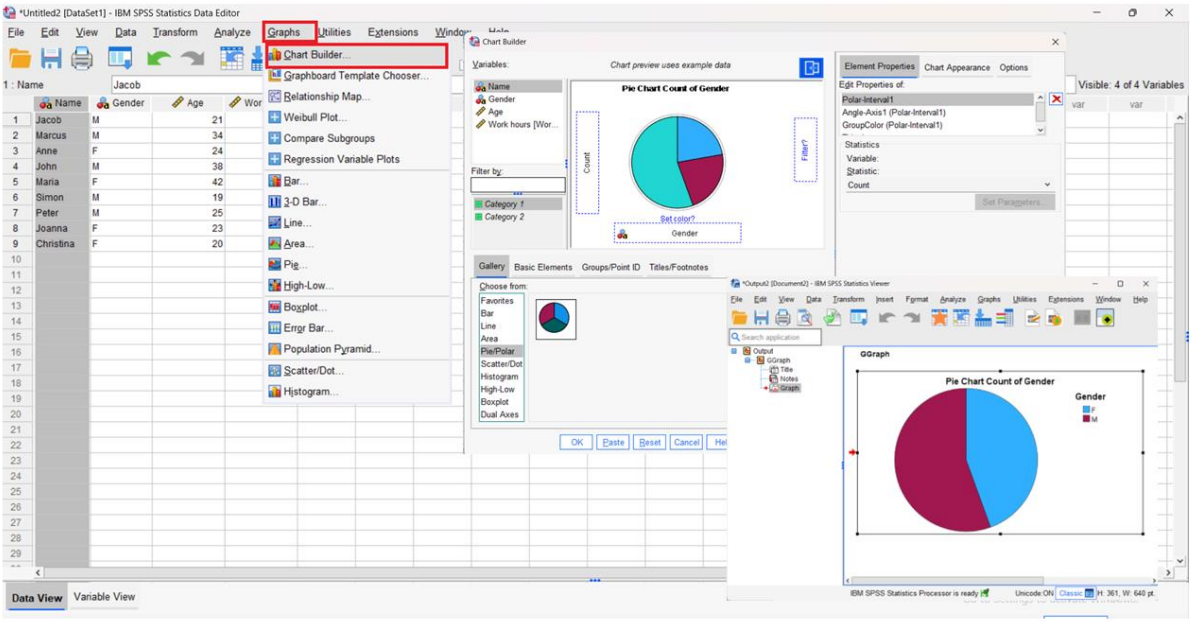
Slika 4.30Prozori za prikaz podataka i varijabli.

Nakon što ispravno postavite svoje podatke, možete ih istraživati unutar SPSS-a. SPSS omogućuje korisnicima izvođenje temeljne statističke analize bez oslanjanja na unaprijed definirane funkcije. Na početnom zaslonu (pogledajte sliku 7.3), idite na "Analiziraj", nakon čega slijedi "Deskriptivna statistika", a zatim odaberite "Istraži". U odjeljku "Istraži" pronaći ćete različite opcije ovisno o karakteristikama podataka koje ste unijeli. U ovom načinu rada SPSS će vam pružiti informacije o "deskriptivnoj statistici" o vašim podacima. Iako je ovo vrijedno za početnu analizu podataka, ono nudi samo temeljne uvide i ne ulazi u detaljniju statističku analizu, koja će biti obrađena u narednim poglavljima. Prije nego što nastavimo dalje, također ćemo istražiti još jednu funkciju u SPSS-u — vizualizaciju grafikona (IBM, 2021).



Slika 4.31Postavke deskriptivne statistike.

SPSS nudi niz opcija za vizualizaciju podataka, uključujući histograme, dijagrame pravokutnika, stupčaste grafikone, raspršene dijagrame, linijske grafikone, tortne grafikone i još mnogo toga. Do ove točke trebali biste imati osnovno razumijevanje o tome što svaka vrsta grafikona predstavlja i kako tumačiti rezultate koje oni pružaju. Stoga ćemo se usredotočiti na to kako izraditi te grafikone unutar softvera SPSS. Da biste izradili grafikone, odaberite karticu "Grafikoni" na početnom ekranu, nakon čega slijedi "Izrada grafikona". U novom prozoru možete odabrati vrstu grafikona koju želite izraditi i odabrati varijable koje želite uključiti. Nakon odabira "Završi", pojavit će se novi prozor s rezultatima vizualiziranim u odabranom formatu grafikona. U ovom novom prozoru možete aktivno komunicirati s grafikonom, što vam omogućuje izmjenu varijabilnih boja i fontova, istraživanje distribucija varijabli na grafikonu, i više (IBM, 2021).



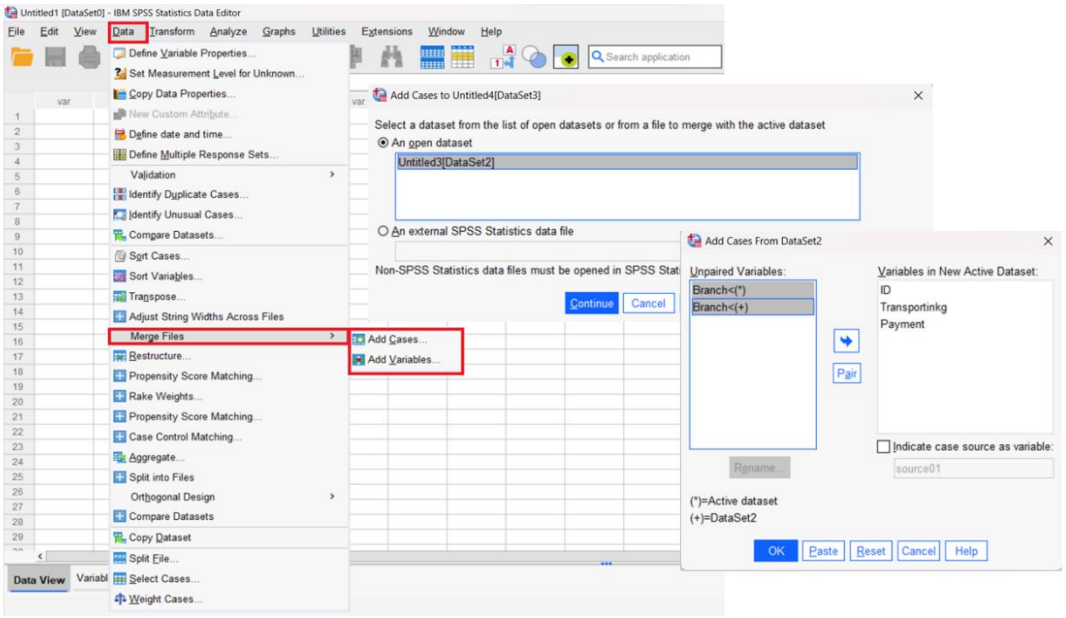
Slika 4.32Postavke izrade grafikona u SPSS-u.

Do ove točke pokrili smo tri od četiri pravila za "Istraživanje podataka", koja uključuju gledanje podataka (istraživanje neobrađenih podataka), identifikaciju podataka (određivanje tipova podataka) i, u određenoj mjeri, grafičko prikazivanje i opisivanje podataka putem deskriptivne statistike i izrade grafikona. Posljednje pravilo je "Formulacija pitanja", gdje se pitamo što želimo postići analizom podataka i u skladu s tim postavljamo grafikone i deskriptivnu statistiku kako bismo dobili odgovore na naša specifična pitanja. Na primjer, u našem trenutnom primjeru, pitanje bi moglo biti: "Je li naša analizirana populacija pretežno ženska?" Korištenjem i grafikona i deskriptivne statistike možemo zaključiti da se naša populacija sastoji uglavnom od muških osoba. Kada formulirate svoja pitanja, uvijek uzmite u obzir dostupne podatke i varijable koje ste identificirali (Garth, 2008). Ovime je završen prvi dio SPSS analize podataka, a sada ćemo nastaviti s pripremom testa.

7.2 Upravljanje podacima

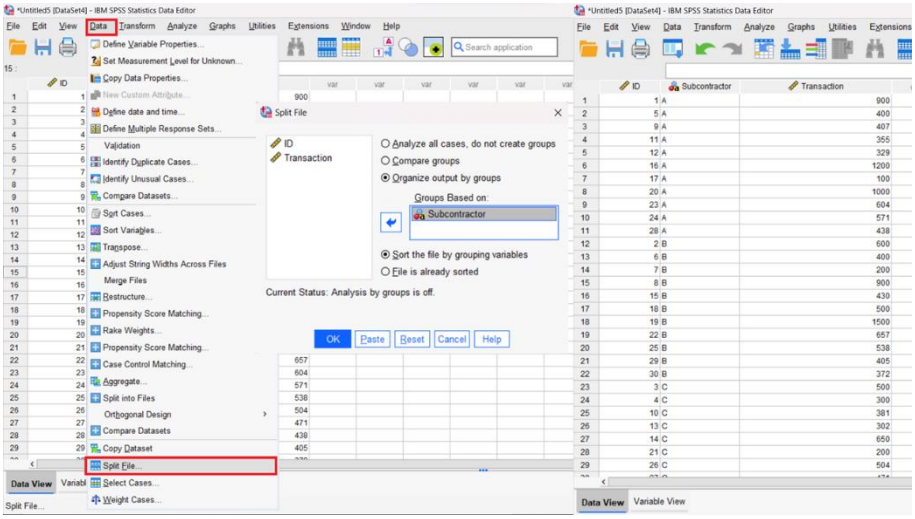
Kada se bavite ključnim podacima u SPSS softveru, postaje ključno razumjeti tehnike za manipuliranje informacijama u pojedinačnim aktivnim skupovima podataka. SPSS pruža funkcionalnosti koje olakšavaju manipulaciju postojećim podacima sadržanim u aktivnim skupovima podataka. Povremeno možete naići na dvije baze podataka odvojeno uvezene u

skupove podataka, no prednost je da ih spojite radi poboljšane analize. Razmotrimo logističku tvrtku s dvije podružnice, od kojih svaka daje podatke o troškovima i prijevozu tereta u kilogramima. Cilj menadžera je analizirati ukupnu učinkovitost poduzeća. U SPSS-u to uključuje navigaciju "Podaci", odabir "Spoji datoteke" i dvije različite opcije. Jedan uključuje odabir "Slučajevi" i određivanje varijable za spajanje, uklanjanje te varijable dok spaja ostale. Alternativno, odabirom opcije "Varijabla" zadržava se varijabla u novom skupu podataka. Praktična primjena očita je u našem scenariju logistike, gdje spajanje skupova podataka pojednostavljuje sveobuhvatnu analizu učinka tvrtke.



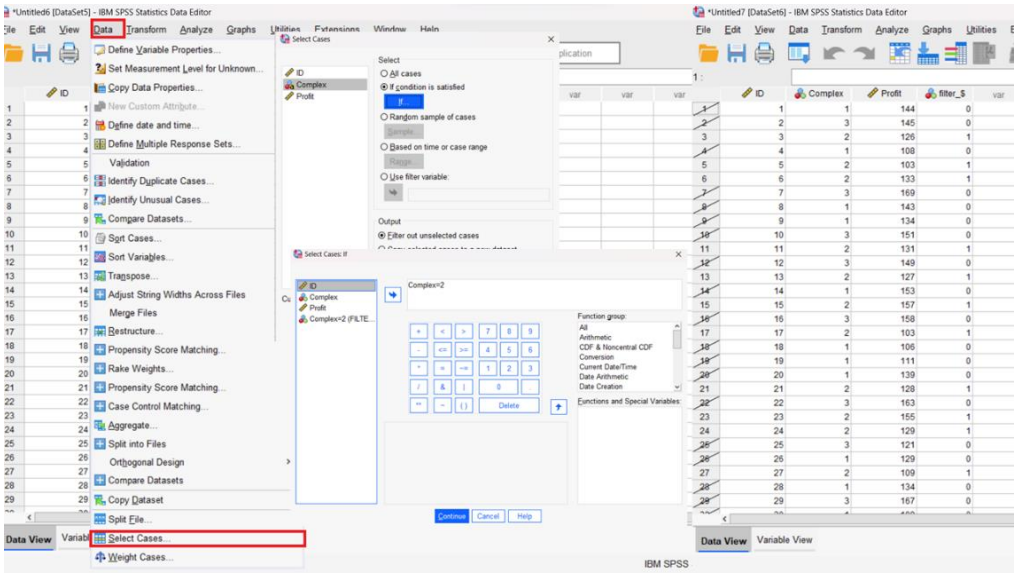
Slika 4.33Prozor za spajanje datoteka.

Dok funkcije spajanja i razdvajanja omogućuju određenu manipulaciju podacima, opcija "Odaberi slučajeve" nudi različite prednosti. Zamislite da imate podatke za prodavaonice B, C i D u jednoj bazi podataka, a fokus je isključivo na usporedbi prodavaonice A i prodavaonice C. Odabirom "Podataka" i "Odaberi slučajeve" možete odrediti varijable od interesa, učinkovito filtrirajući izbaciti neželjene podatke. Na primjer, postavljanje Prodavaonica C kao 2 upućuje softver da se koncentrira isključivo na Prodavaonicu C, generirajući izlaz koji je zatim dostupan za naknadne analize, kao što je deskriptivna statistika, fokusirajući se isključivo na odabrane slučajeve. Takav pristup također omogućuje komparativnu analizu samo između vrijednosti Prodavaonica A i Prodavaonica C.



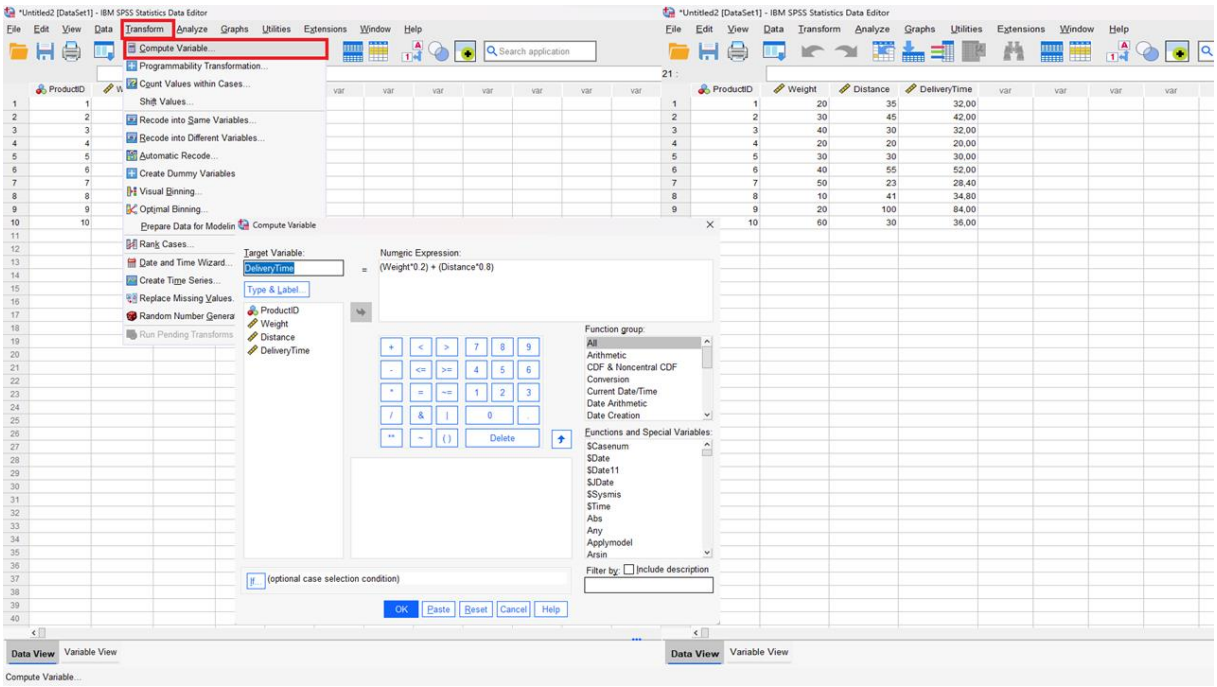
Slika 4.34Prozor za dijeljenje datoteke.

Dok funkcije spajanja i razdvajanja omogućuju određene manipulacije podacima, postoji i opcija "Odaberi slučajeve". Zamislite da pouzdano znamo da prodavaonica A ima u prosjeku 120 € dobiti i želimo to usporediti s prodavaonicom C. Nažalost, u našoj bazi podataka imamo podatke za prodavaonice B, C i D u jednoj bazi podataka i analiza bi uključivala podatke iz sve tri prodavaonice. Klikom na "Podaci" i "Odaberi slučajeve" možemo odabrati koju varijablu želimo fokusirati. U našim smo slučajevima postavili da prodavaonica C treba biti postavljena kao 2, a zatim smo stvorili funkciju za softver da se fokusira samo na prodavaonicu C. Izlaz se zatim može koristiti za naknadnu analizu odabirom ovog novog stupca (npr. deskriptivna statistika).



Slika 4.35Odabir slučaja.

Povremeno skupovi podataka mogu već sadržavati varijable, ali ipak postoji potreba za uvođenjem novih varijabli na temelju postojećih. Uzmimo, na primjer, menadžera logističke tvrtke koji posjeduje podatke o težini i prijeđenoj udaljenosti za razne proizvode, ali zahtijeva vrijeme isporuke za optimizaciju ruta. U SPSS-u, postizanje toga uključuje klik na "Transform", a zatim na "Compute Variables". Nova varijabla, DeliveryTime, stvara se unutar novog prozora postavljanjem numeričkih izraza. U ovom slučaju, dodjeljivanje ljestvice od 0,8 za udaljenost i 0,2 za težinu rezultira novom varijablom koja predstavlja vrijeme isporuke, što je ključni dodatak skupu podataka. Postoji fleksibilnost izračunavanja dodatnih varijabli, kreiranih za potrebe statističkih testova.



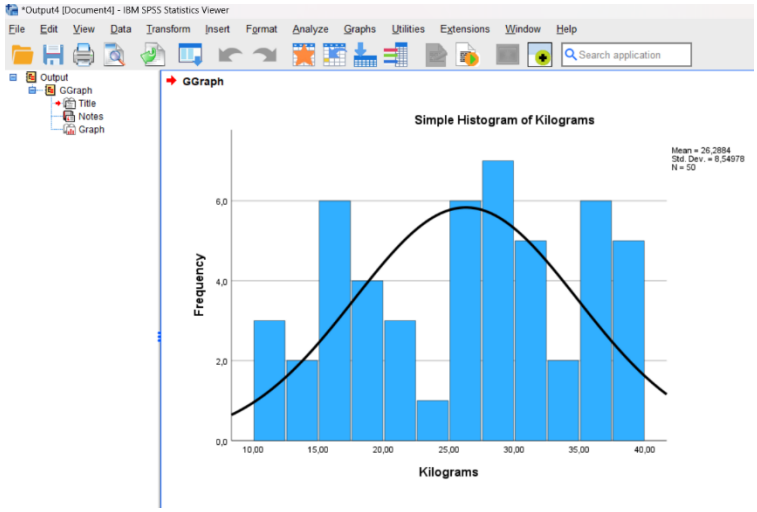
Slika 4.36Postupak izračunavanja varijabli.

Ovo zaključuje mali pregled funkcija upravljanja podacima koje pokriva SPSS, a koje bi mogle biti korisne tijekom sljedećih testova modela koji su obuhvaćeni u ovom poglavlju. Nastavit ćemo s fazama koje su potrebne prije nego što možemo provesti statistički test u softveru SPSS.

7.3 Priprema testa

Prije nego što nastavite sa statističkim testovima, bitno je pridržavati se standardnog tijeka procesa analize podataka, koji uključuje istraživanje podataka (kao što je objašnjeno u

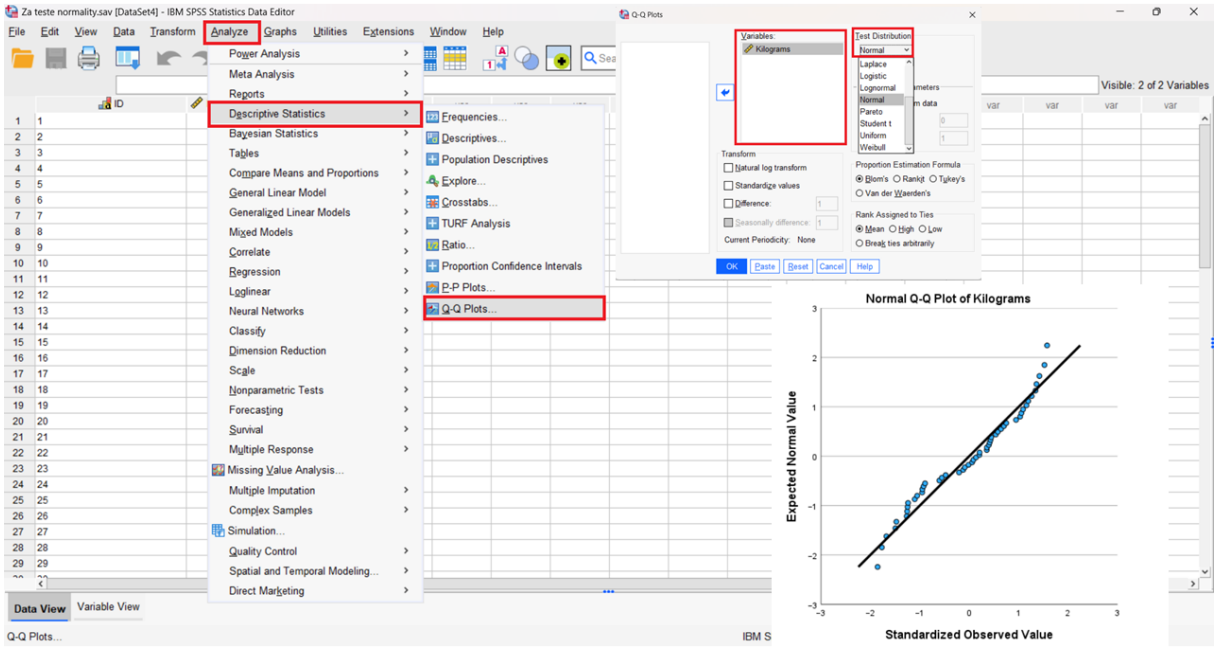
poglavljima 7.1 i 7.2), analizu podataka i interpretaciju rezultata (Garth, 2008; George i Mallery, 2022). U ovom poglavlju naš fokus je na analizi podataka pomoću softvera SPSS. Budući da smo hipoteze već obradili u prethodnim poglavljima, naš primarni fokus bit će na provođenju testova normalnosti unutar SPSS-a. Postoje tri metode za procjenu normalnosti: histogram, QQ-grafikon i test normalnosti. Preporučljivo je upotrijebiti najmanje dvije, ako ne i sve tri ove opcije, budući da svaka pruža različite informacije (Ghasemi i Zadesiasl, 2012.). Za izradu histograma idite na " Graphs ", a zatim na "Chart Builder". U novom prozoru odaberite "Histogram". Ako imate više varijabli, morate ponoviti ovaj postupak za svaku kako biste dobili rezultate. Histogram potvrđuje test za normalnu distribuciju ako stupci koji predstavljaju varijable vrijednosti nalikuju zvonolikoj krivulji. Ako su stupci više nagnuti u lijevu ili desnu stranu, to može značiti eksponencijalnu distribuciju. Na primjer, generirali smo bazu podataka od 100 ID-ova, svaki s varijablom koja predstavlja težinu u kilogramima. Prateći upute izradili smo histogram, kao što je prikazano na slici 7.9. Kao što je vidljivo sa slike, stupci su raspoređeni po grafikonu i iako možda ne odražavaju savršeno krivulju, ipak sugeriraju normalnu distribuciju i pozitivan rezultat testa (George i Mallery, 2022; Goeman i Solari, 2021).



Slika 4.37Histogram rezultata testa normalnosti.

Još jedna opcija za provođenje testova normalnosti je QQ-grafikon, koji se može pokrenuti klikom na "Analyze" (hrv. Analiziraj), nakon čega slijedi "Descriptive Statistics", a zatim odabirom "Q-Q Plots". Prednost ovog pristupa je što omogućuje procjenu više varijabli istovremeno (Williamson, bd). Test se smatra uspješnim kada se točke na dijagramu grupiraju usko oko ravne crte, što predstavlja normalnu distribuciju. Ako točke formiraju "repove", to ukazuje na neuspješan test normalnosti (Andersen i Dennison, 2018). Koristeći istu bazu

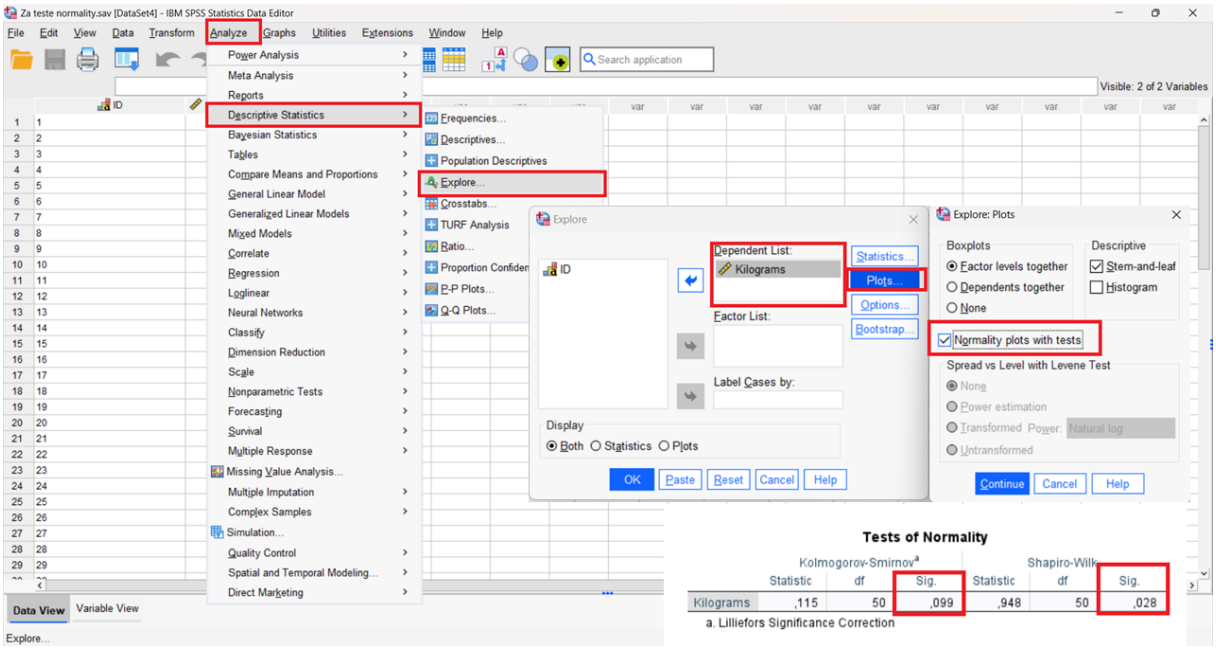
podataka iz testa histogramskog grafikona, proveli smo QQ grafikon test. Na slici 7.10 u nastavku možete primijetiti da je većina točaka klastera za našu varijablu poravnata s ravnom linijom, što ukazuje na normalnu distribuciju naših podataka. Iako smo već u ovoj fazi mogli zaključiti da je test normalnosti pozitivan, odlučili smo tražiti potvrdu iz sva tri testa.



Slika 4.38QQ grafikon test normalnosti - postavke i rezultati.

Konačna opcija za provođenje testa normalnosti je takozvani Test normalnosti, koji se smatra statističkim testom. Obično se koristi Kolmogorov-Smirnov test, ali za male veličine uzorka može se koristiti Shapiro-Wilkov test (Goeaman i Solari, 2021.). U SPSS-u možete izvršiti ovaj test klikom na "Analyze", nakon čega slijedi "Descriptive Statistics", a zatim "Explore". Morate postaviti varijable koje želite provjeriti ispod okvira "Dependent List" (hrv. Zavisna lista). Zatim pod "Plots" odaberite "Normality Plots with Tests" (hrv. Grafike normalnosti s testovima). Test se smatra uspješnim ako je stupac Sig (p -vrijednost) u rezultatima veći od 0,05, što ukazuje na normalnu distribuciju. Ako je p -vrijednost manja od 0,05, to ukazuje da distribucija nije normalna i test se smatra neuspješnim. Ovaj smo test još jednom proveli koristeći istu bazu podataka kao i u prethodnim testovima. Iz rezultata možemo zaključiti da je prema standardu Kolmogorov-Smirnov test pozitivan jer je p -vrijednost veća od 0,05. Međutim, za Shapiro-Wilkov test, p -vrijednost je niža, što ukazuje na negativan rezultat testa. Do ovih različitih rezultata dolazi jer

oba pristupa imaju različite postavke osjetljivosti i snagu u otkrivanju odstupanja (Ghasemi i Zahediasl, 2012.). Budući da smo već proveli testove QQ dijagrama i histogramskog grafikona, Test normalnosti može se općenito smatrati pozitivnim. Uz potvrđene testove normalnosti, možemo provesti glavne testove, kao što je test jednog uzorka.



Slika 4.39Postavke i rezultati testa normalnosti.

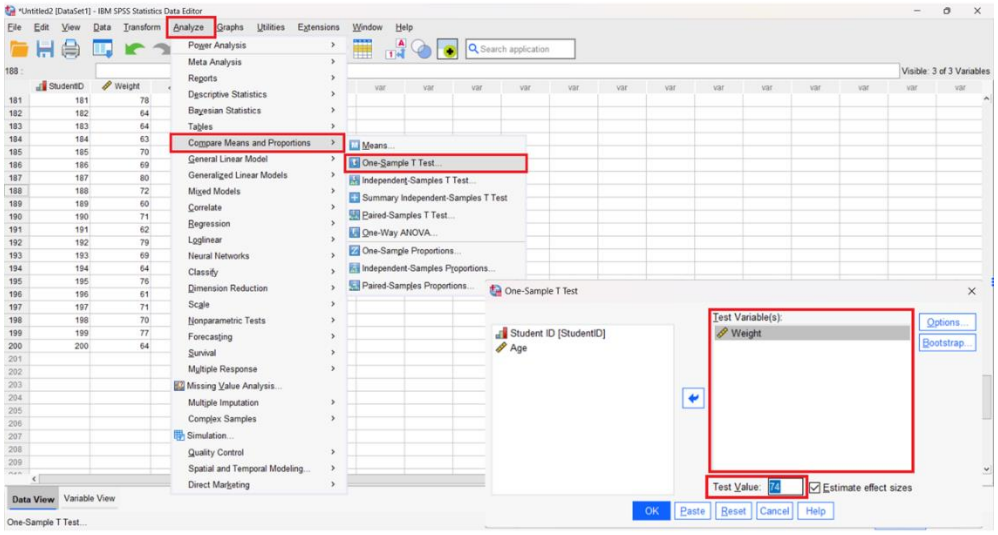
7.4 T-test jednog uzorka

Već ste pokrili teoriju iza T-testa jednog uzorka u prethodnim poglavljima; stoga ćemo se prvenstveno usredotočiti na provođenje testa sa softverom SPSS. Za naš T-test jednog uzorka pripremili smo bazu podataka s uzorkom od 200 ispitanika, koji uključuje 1 kategoričku varijablu (ID studenta) i 2 numeričke varijable (težinu i dob) (Kim, 2015.). Prateći upute iz prethodnih potpoglavlja provodimo sljedeće korake:

- Istražite podatke, točnije naše **varijable** i **deskriptivnu statistiku** i postavite naše **pitanje**.
- Provjerite **normalnost**, budući da bi samo jedan varijabilni histogram i QQ grafikon trebali biti dovoljni.
- Postavite hipotezu, gdje se za **nultu** varijabla ne razlikuje od određene vrijednosti i **alternativu** gdje je drugačija.

- Provedite **Studentov T-test** .
- Tumačite rezultate, fokusirajući se na to je lit **nulta hipoteza odbijena** ili **ne** , odgovorite na pitanje i napišite izvješće o našem testu.

U našem slučaju odlučili smo da naše pitanje bude: Je li prosječna težina učenika veća od 74 kilograma? Nakon pitanja postavljamo našu hipotezu za pitanje, a to je "Nulta = nema razlike" i "Alternativa = postoji razlika". Proveli smo histogram i QQ grafikone kako bismo provjerili testove normalnosti, a nakon njihovog završetka, slijedio je T-test. Da bismo pokrenuli T-test, kliknemo "Analyze" i nastavimo s "Compare Means" (hrv. Usporedi srednje vrijednosti) i "One-Sample T-test" (hrv. T-test jednog uzorka). U okvir s varijablama testa stavljamo studentski ID, postavljamo vrijednost testa na 74 i započinjemo test (pogledajte sliku 7.12).



Slika 4.40Postavke T-testa jednog uzorka.

Nakon potvrde testa, pojavit će se drugi prozor s rezultatima naše analize (pogledajte sliku 7.13). Ovaj prozor pruža nekoliko informacija u vezi s našom analizom. U ovom su slučaju obje *p* -vrijednosti niže od 0,05, što ukazuje na značajnost testa. Dodatno, provjeravamo vrijednosti *t* i *df* , koje su u našem slučaju -9,806 odnosno 199. Iz ovih rezultata možemo zaključiti da je naša nulta hipoteza odbačena. Stoga je cjelovito izvješće o rezultatima sljedeće: "Prosječna težina studenta značajno je niža (srednja vrijednost = 69,63) od vrijednosti od 74 kg (t-test jednog uzorka, t = -9,806, df = 199, p -vrijednost < 0,001)".

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Weight	200	69,63	6,303	,446

One-Sample Test

Test Value = 74

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Weight	-9,806	199	<,001	<,001	-4,370	-5,25	-3,49

One-Sample Effect Sizes

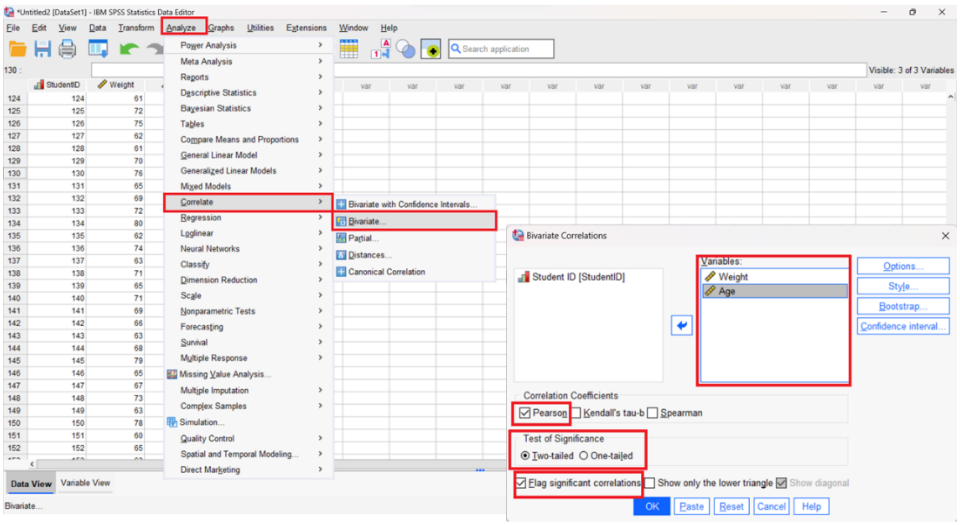
		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Weight	Cohen's d	6,303	-,693	-,847	-,538
	Hedges' correction	6,326	-,691	-,844	-,536

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Slika 4.41Rezultati T-testa jednog uzorka.

7.5 Korelacija

Prijeđimo sada na drugi test, a to je test korelacije. Provest ćemo ga koristeći istu bazu podataka kao u primjeru t-testa s jednim uzorkom. Slično t-testu s jednim uzorkom, slijedit ćemo postupak uz nekoliko izmjena. Kada se vrši korelacija između dvije varijable, važno je odrediti koja je zavisna, a koja nezavisna varijabla (Janse i dr., 2021.; Mishra i dr., 2019). Ovaj odabir možete napraviti na temelju vašeg istraživačkog pitanja. U našem slučaju želimo istražiti "Postoji li korelacija između dobi studenta i njegove težine?". Nakon pitanja, težinu smatramo zavisnom varijablom, a dob nezavisnom varijablom, jer želimo istražiti jesu li varijacije u dobi povezane s varijacijama u težini. Definiramo naše nulte i alternativne hipoteze (vidi 7.3 i 7.4), a zatim pokrećemo test klikom na "Analyze", nakon čega slijede "Correlate" (hrv. Koreliraj) i "Bivariate" (hrv. Bivarijantno). Obje varijable treba staviti u polje "Variable". Provjerite jesu li odabrani ili postavljeni "Pearson", "Two-Tailed" i "Flag Significant" (pogledajte sliku 7.14). U ovom smo slučaju odabrali "Pearson" jer naši podaci pokazuju normalnu distribuciju i mogu se analizirati pomoću parametarskih metoda. Ako normalna distribucija nije naznačena, treba koristiti neparametarske metode (u ovom slučaju, odabrali biste Spearmana umjesto Pearsona) (George i Mallery, 2022; McClure, 2005).



Slika 4.42Postavke testa korelacije.

Još jednom dobivamo rezultate u novom prozoru (pogledajte sliku 7.15). Iz rezultata možemo vidjeti da je naša Pearsonova korelacija -0,038, a *p* -vrijednost 0,596. U korelacijskoj analizi, što je vrijednost korelacije bliža nuli, to je korelacija između varijabli slabija. U našem slučaju, korelacija je vrlo blizu nule, što ukazuje da nema značajne korelacije između dvije varijable (McClure, 2005). Dodatno, visoka *p* -vrijednost (0,596) sugerira da nema značajnih dokaza za zaključak da postoji značajna korelacija između dviju odabranih varijabli (Williamson, bd). Kao rezultat toga, naša nulta hipoteza nije odbačena. Na temelju toga možemo izvijestiti da "nije bilo korelacije između dobi i težine studenta".

Correlations

Correlations

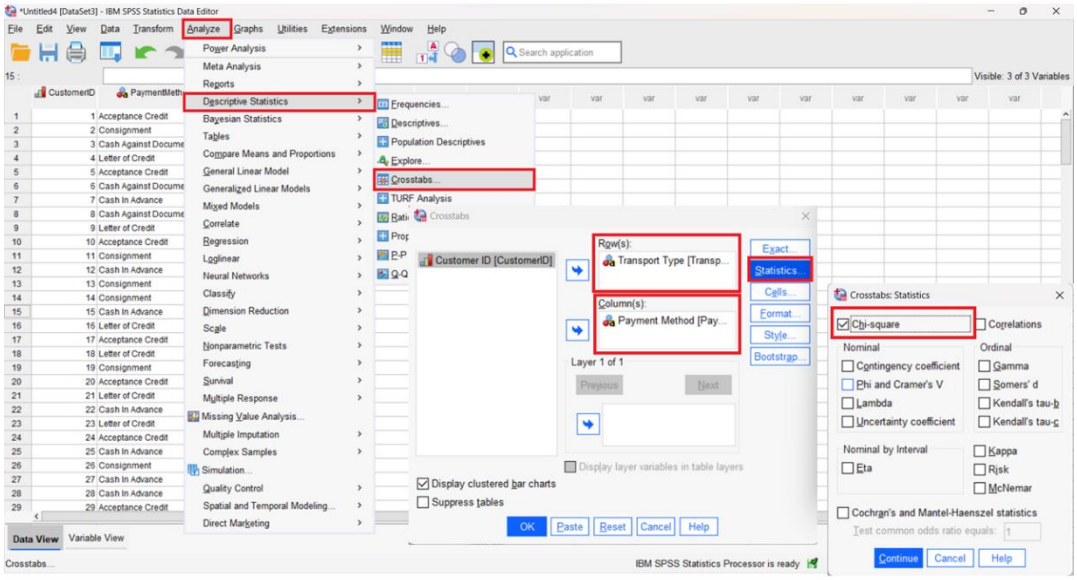
		Weight	Age
Weight	Pearson Correlation	1	-,038
	Sig. (2-tailed)		,596
	N	200	200
Age	Pearson Correlation	-,038	1
	Sig. (2-tailed)	,596	
	N	200	200

Slika 4.43Rezultati testa korelacije.

7.6 Hi-kvadrat

Treći test koji ćemo izvesti u SPSS softveru je Hi-kvadrat test. Za razliku od prethodna dva testa, Hi-kvadrat test uspoređuje dvije kategoričke varijable, a ne numeričke varijable (Turhan, 2020). Kao i postupak u odjeljcima 7.4 i 7.5, počinjemo istraživanjem podataka i formuliranjem

istraživačkog pitanja. U našem primjeru imamo logističku tvrtku s 200 kupaca, te imamo podatke o vrsti plaćanja i vrsti prijevoza koju je svaki kupac odabrao. Pitanje na koje želimo odgovoriti je: "Pokazuju li različite vrste plaćanja različite preferencije za vrste prijevoza?" Budući da se radi samo o kategoričkim varijablama, nema potrebe za testom normalnosti. Postavljamo našu nultu hipotezu (preferencije za vrste prijevoza iste su za sve vrste plaćanja) i alternativnu hipotezu. Za provođenje hi-kvadrat analize kliknite na "Analyze", nakon čega slijedi "Descriptive Statistics", i odaberite "Crosstabs" (hrv. Unakrsne analize). Ključno je smjestiti varijable na temelju vašeg istraživačkog pitanja u okvir stupca ili retka (pogledajte sliku 7.16) (Garth, 2008.).



Slika 4.44Postavke Hi-kvadrat testa.

Nakon analize, novi prozor prikazuje rezultate (pogledajte sliku 7.17). U ovom prozoru možete primijetiti da Pearsonova hi-kvadrat vrijednost iznosi 11,614, *df* vrijednost 12, a *p* -vrijednost (asimptotska značajnost) 0,477. Na temelju ovih rezultata možemo zaključiti da ne postoji značajna povezanost između dviju varijabli, a nulta hipoteza nije odbačena. Stoga izvješće slijedi: "Nema otkrivenih značajnih preferencija između različitih vrsta plaćanja za različite vrste prijevoza (dvostrani Chi-Square test, chi-sq = 11,614, *df* = 12, *p* -vrijednost = 0,477)."

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Transport Type * Payment Method	200	100,0%	0	0,0%	200	100,0%

Transport Type * Payment Method Crosstabulation

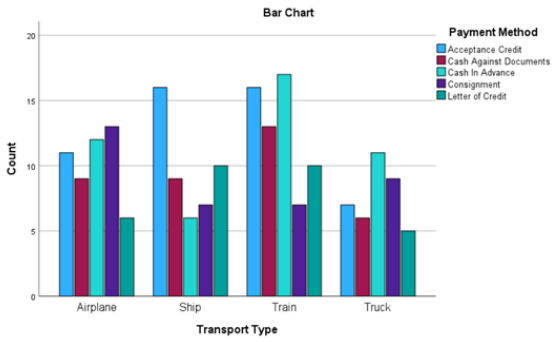
Count

		Payment Method					
		Acceptance Credit	Cash Against Documents	Cash In Advance	Consignment	Letter of Credit	Total
Transport Type	Airplane	11	9	12	13	6	51
	Ship	16	9	6	7	10	48
	Train	16	13	17	7	10	63
	Truck	7	6	11	9	5	38
Total		50	37	46	36	31	200

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	11,614[a]	12	,477
Likelihood Ratio	11,965	12	,448
N of Valid Cases	200		

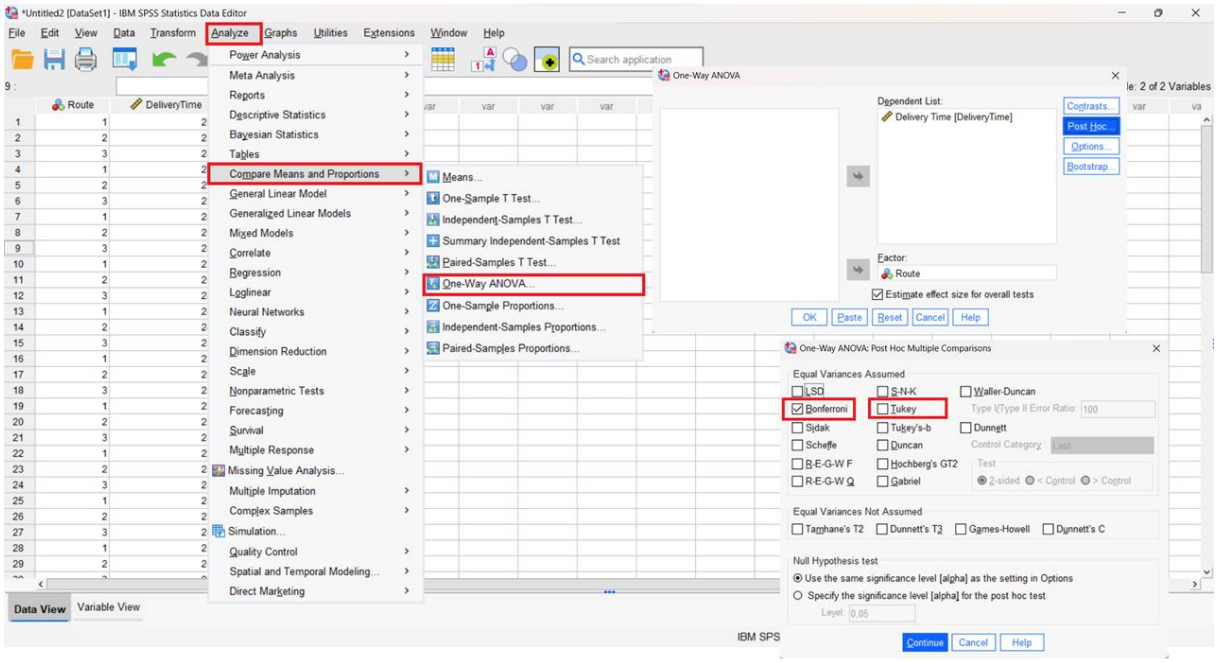
a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,89.



Slika 4.45Rezultati Hi-kvadrat testa.

7.7 ANOVA

Posljednji test koji ćemo pokriti je ANOVA test, koji se posebno fokusira na jednostavniji model poznat kao jednosmjerna ANOVA, koji uključuje kategoričku varijablu i numeričku varijablu (Goeman i Solari, 2021). Kao i kod T-testa, slijedit ćemo isti postupak: istražiti podatke, formulirati istraživačko pitanje, provesti test normalnosti i postaviti hipoteze. Razmotrimo studiju slučaja transportnog dispečera koji radi za logističku tvrtku. Dispečer blisko surađuje s partnerskom tvrtkom i redovito planira tri različite rute kamionima za isporuku robe. Zbog politike "Just-in-time" koja naglašava brže isporuke, postavlja se pitanje: "Utječe li izbor rute dostave na vrijeme isporuke za tvrtku?" Da biste pokrenuli ANOVA test u SPSS-u, idite na "Analyze" nakon čega slijedi "Compare Means..." i zatim "One-way ANOVA". Postavite zavisnu varijablu u okvir "Dependent List" (hrv. Popis zavisnih), a varijablu Faktor u okvir "Factor" (pogledajte sliku 7.18). Za temeljitu analizu uključili smo i Post Hoc postavku. Važno je napomenuti da se Post Hoc analiza treba provesti samo ako je početni ANOVA test pozitivan. Primjenom Post Hoc analize možemo identificirati optimalan izbor (u našem slučaju rutu). Najpouzdanije metode koje se koriste za Post Hoc analizu su ili Bonferronijeva korekcija ili Tukeyjeva HSD metoda (Goeman i Solari, 2021.).



Slika 4.46Postavke za ANOVA analizu.

Rezultati naše analize pokazuju da je naša *F*-statistička vrijednost 11,173 (više vrijednosti ukazuju na više varijacija između skupina) i *p*-vrijednost <0,001, što znači da je naša nulta hipoteza odbačena (vidi sliku 7.19). Budući da postoji značajna razlika između tri rute (<0,001), post hoc test je također valjan u našem slučaju (George & Mallery, 2022). Nakon provođenja Bonferronijevog testa korekcije, možemo vidjeti da su najbolje *p*-vrijednosti zabilježene u slučaju rute 2 (pogledajte sliku 7.19). U izvješću možemo zaključiti da je „postojala značajna razlika u odabiru rute isporuke u korelaciji s vremenom isporuke (1-way ANOVA, F =11,173, df = 47, p-vrijednost = <0,001). Ruta 2 imala je najbolje rezultate u vremenu isporuke."

ANOVA

Delivery Time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16,527	2	8,263	11,173	<,001
Within Groups	33,280	45	,740		
Total	49,807	47			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Delivery Time

Bonferroni

(I) Route	(J) Route	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-1,4250*	,3040	<,001	-2,181	-,669
	3	-,5500	,3040	,231	-1,306	,206
2	1	1,4250*	,3040	<,001	,669	2,181
	3	,8750*	,3040	,018	,119	1,631
3	1	,5500	,3040	,231	-,206	1,306
	2	-,8750*	,3040	,018	-1,631	-,119

*. The mean difference is significant at the 0.05 level.

Slika 4.47Početni rezultati ANOVA analize i rezultati post hoc testa.

Zaključujemo ovo poglavlje knjige uz razumijevanje da smo u ovom poglavlju pokrili neke od uobičajenih testova. Postoje i drugi testovi, kao što je ANOVA ponovljenih mjerenja, testovi pouzdanosti i testovi osjetljivosti, koji se također mogu modelirati i analizirati pomoću softvera SPSS. Ovi dodatni testovi pružaju širi raspon alata za analizu podataka i daju smislene uvide u različita istraživanja i praktične primjene.

Literatura 7. poglavlja

- Andersen, A.J. & Dennison, J.R. (2018). An Introduction to Quantile-Quantile Plots for the Experimental Physicist. Journal Articles, 51.
- Garth, A. (2008). Analysing data using SPSS [available at: https://students.shu.ac.uk/lits/it/documents/pdf/analysing_data_using_spss.pdf, access October 26, 2023]
- George, D. & Mallery, P. (2022). IBM SPSS Statistics 27 Stet by Step: A Simple Guide and Reference, 17TH edition, Abingdon: Routledge
- Ghasemi, A. & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. International Journal of Endocrinology and Metabolism, 10(2), pp. 486-489.
- Goeman, J.J. & Solari, A. (2021). Comparing Three Groups. The American Statistician, 76(2), pp. 168-176
- IBM (2021). IBM SPSS Statistics 28 Brief [available at: https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Statistics_Brief_Guide.pdf, access October 26, 2023]
- Janse, R.J., Hoekstra, T., Jager, K.J., Zoccali, C., Tripepi, G., Dekker, F.W. & van Diepen, M. (2021). Conducting correlation analysis: important limitations and pitfalls, 14(11), pp. 2332-2337.
- Kim, T.K. (2015). T test as a parametric statistic. Korean Journal of Anesthesiology, 68(6), pp. 540-546
- Landau, S. & Everitt, B.S. (2004). A Handbook of Statistical Analyses using SPSS, 1st edition, London: Chapman & Hall/CRC
- McClure, P. (2005). Correlation Statistics Review of the Basics and Some Common Pitfalls. Journal of Hand Therapy, 18(3), pp. 378-380

- Mishra, P., Singh, U., Pandey, C.M., Mishra, P. & Pandey, G. (2019). Application of Student's t-test, Analysis of Variance, and Covariance. Annals of Cardiac Anesthesia, 22(4), pp. 407-411
- Turhan, N.S. (2020). Karl Pearson's chi-square tests. Educational Research and Reviews, 15(9), pp. 575-580
- Williamson, M. (b.d.). Data Analysis using SPSS [available at: https://med.und.edu/research/daccota/_files/pdfs/berdc_resource_pdfs/data_analysis_using_spss.pdf, access October 26, 2023]

8. Temelji poslovne analitike uključujući R i SQL

Što je poslovna analitika (engl. *business analytics* - BA)? Koje probleme rješava i koje alate koristi? Što su R i SQL? Kako je BA povezan s R i SQL? Postoje li primjeri dobre prakse u kojima se ti softveri koriste za rješavanje logističkih poslovnih problema?

Na ova i slična pitanja pokušat ćemo dati odgovore u sljedećem poglavlju.

8.1 Što je poslovna analitika?

BA predstavlja holistički pristup analizi podataka i poslovnom odlučivanju. To je okruženje vođeno podacima s ciljem poboljšanja poslovnih performansi tvrtke pružanjem temelja za informiranije donošenje odluka. To je sustavni proces razmišljanja koji primjenjuje kvalitativne, kvantitativne i statističke računalne alate i metode za analizu podataka, stjecanje uvida, informiranje i podršku donošenju odluka. Svaka određena analiza može koristiti različite tehnike uključujući dijagnostičke, prediktivne, preskriptivne i optimizacijske modele (Power i dr., 2018). Mikalef i dr., (2019) daju plan za akademsko istraživanje i praktičnu primjenu, ističući transformativni potencijal analitike kada se pravilno integrira u organizacijske procese. U skladu s tim, autori navode da BA zahtijeva od organizacija da radikalno redizajniraju način na koji se takvim inicijativama pristupa, kako se dizajniraju i usavršavaju, kako se planiranje resursa i orkestracija izvršava i strateški usklađuje, kao i da ponovno vrednuju svoje očekivane rezultate izvedbe, njihovu povezanost sa strateškim ciljevima i, kao rezultat toga, razviju odgovarajuće KPI-eve (Mikalef et al., 2019).

Glavni zadaci BA su osigurati kanal znanja kako bi se osigurala koherentna veza između sirovih podataka i poslovnih odluka. Opći cilj je poslovna učinkovitost kroz 'vertikalizaciju', upotrebljivost i integraciju s operativnim sustavima (Kohavi i dr., 2002). BA ima mnogo područja primjene i povezanih izvedenica: financijska analitika, analitika opskrbnog lanca, analitika krize, analitika znanja, marketinška analitika, analitika kupaca, analitika usluga, analitika ljudskih resursa, analitika talenata, analitika procesa, analitika rizika (Holsapple i dr., 2014) .

Postoje tri vrste platformi poslovne analitike:

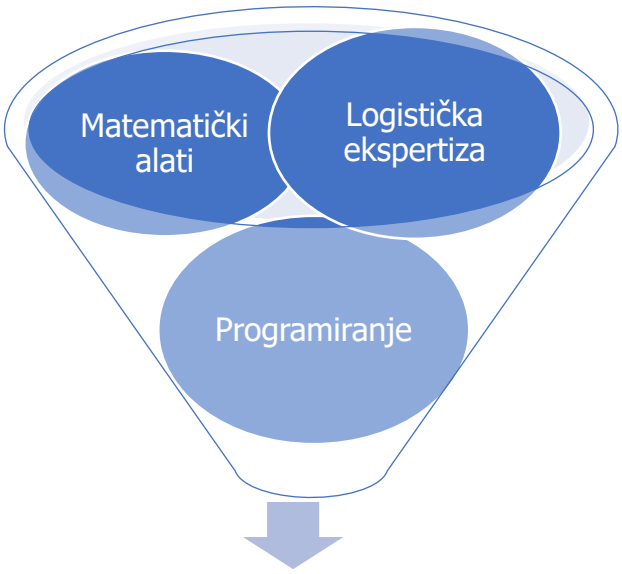
- **Opisne** – gledaju postojeće podatke i daju sažetak statistike i osnovnu vizualizaciju.
- **Prediktivne** – koriste postojeće podatke za procjenu najvjerojatnijih budućih scenarija.
- **Preskriptivne** – automatski obrađuju veliku količinu podataka (engl. *big data*), poslovna pravila, tržišne uvjete itd. Ove platforme koriste metode strojnog učenja i umjetne inteligencije. Cilj je potpuno automatizirano donošenje odluka o tome koje akcije tvrtka treba poduzeti s obzirom na trenutnu situaciju kako bi postigla željene poslovne ciljeve.

Zanimanje za *big data* i poslovnu analitiku eksponencijalno je poraslo tijekom proteklog desetljeća (Mikalef i dr., 2019). Suvremeni BA ukorijenjen je u stalnom napretku sustava za podršku odlučivanju. Ovaj napredak uključuje sve snažnije mehanizme za stjecanje, generiranje, asimilaciju, odabir i emitiranje znanja relevantnog za donošenje odluka. S obzirom na nasljeđe podrške odlučivanju, poslovna analitika nužno sudjeluje u tim mehanizmima i iskorištava ih. Znanje koje se mora obraditi kreće se od kvalitativnog do kvantitativnog, a BA se bavi radom na obje vrste znanja, kako je prikladno za donošenje odluke (Kohavi i dr., 2002). Razlog zašto bi neka organizacija trebala primijeniti BA je u problemima koje rješava. Problemi koje BA ističe su problemi smanjenja učinkovitog upravljanja poduzećem. Sukladno tome, postoji nekoliko razloga za primjenu BA (Holsapple i dr., 2014):

- Ostvarivanje konkurentske prednosti
- Podrška strateškim i taktičkim ciljevima organizacije
- Bolji organizacijski učinak
- Bolji ishodi odluka
- Bolji ili informiraniji procesi odlučivanja
- Proizvodnja znanja
- Dobivanje vrijednosti iz podataka

Bez obzira na vrstu platforme koja se koristi u BA za rješavanje i podršku procesu donošenja odluka u svakoj tvrtki, postoje tri ključna stupa svakog BA rješenja (slika 8.1). Općenito, BA zauzima mjesto u spektru između računalnih znanosti/matematike/podatkovnih znanosti (s

jedne strane) i poslovanja i menadžmenta (s druge strane). Poslovna analitika zahtijeva i tehničko i poslovno znanje. Glavni problem u dizajniranju BA je to što granice nisu jasne (Power i dr., 2018). Sukladno tome, za izvođenje BA potrebni su matematički alati za identifikaciju, izdvajanje i predstavljanje uvida na pravi način putem tablica, grafika, formula itd. Dodatno, alati za programiranje služe kao potpora ovoj vrsti aktivnosti i omogućuju brze izračuna bez grešaka, u usporedbi s tradicionalnim pristupom papira i olovke. Posljednje, ali ne manje važno, potrebna je logistička ekspertiza u određenom području ili poslovnom problemu kako bi se odredili ključni utjecajni čimbenici i povezani ekosustav.



Slika 4.48Ključni stupovi BA u kontekstu opskrbnog lanca i logistike.

Ključni potrošač je poslovni korisnik, čiji posao, vjerojatno u *merchandisingu*, marketingu ili prodaji, nije izravno povezan s analitikom *per se*, ali koji obično koristi analitičke alate za poboljšanje rezultata nekog poslovnog procesa duž jedne ili više dimenzija (kao što su profit i vrijeme do tržišta). Poslovni korisnici ne žele imati posla s naprednim statističkim konceptima; žele jednostavne vizualizacije i rezultate relevantne za zadatak (Kohavi i dr., 2002).

8.2 Što je R?

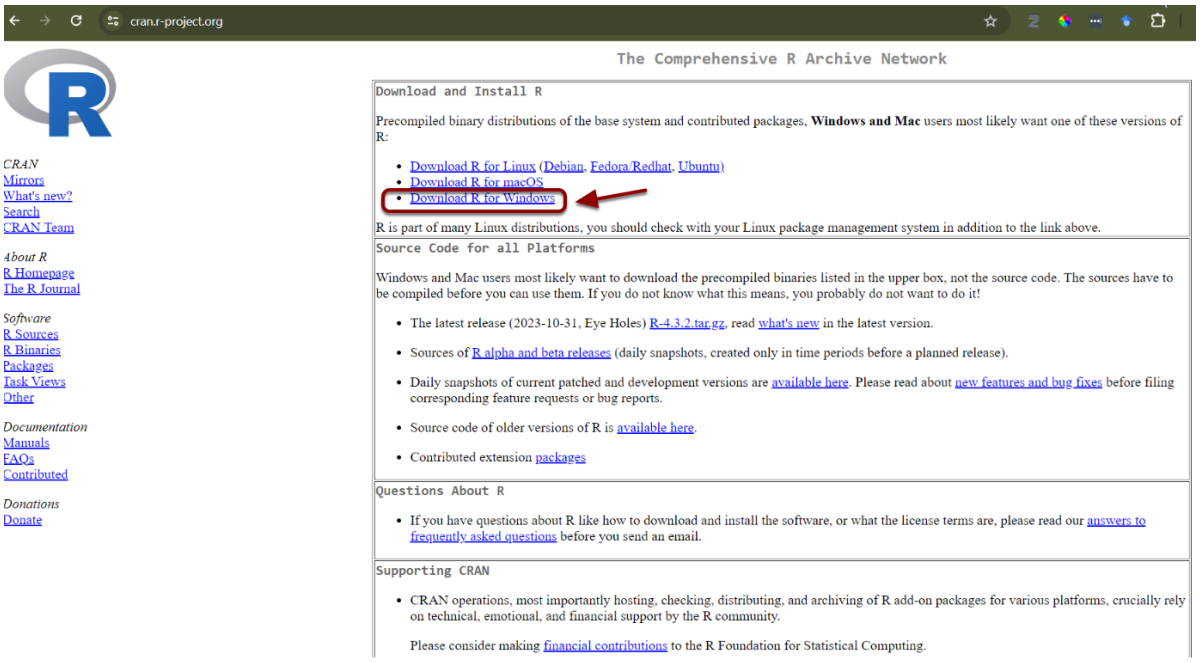
R je integrirani paket softverskih mogućnosti za manipulaciju podacima, izračun i grafički prikaz (R Core Team, 2019). Između ostalog, ima i sljedeće:

- učinkovito rukovanje i skladištenje podataka,
- skup operatora za izračune na nizovima, posebno matricama,
- velika, koherentna, integrirana zbirka posrednih alata za analizu podataka,
- grafičke mogućnosti za analizu i prikaz podataka bilo izravno na računalu ili u tiskanoj kopiji, te dobro razvijen, jednostavan i učinkovit programski jezik (nazvan 'S') koji uključuje uvjete, petlje, korisnički definirane rekurzivne funkcije i mogućnosti unosa i izlaza (većina funkcija koje pruža sustav same su napisane u S jeziku).

Glavne prednosti R-a su činjenica da je R besplatan i da postoji puno dostupne pomoći online. Prilično je sličan drugim programskim paketima kao što je MatLab (nije besplatan), ali je lakši za korištenje od programskih jezika kao što su C++ ili Fortran (Torfs i Brauer, 2014). R je u velikoj mjeri sredstvo za nove metode interaktivne analize podataka. Brzo se razvijao i proširio velikom kolekcijom paketa. Međutim, većina programa napisanih u R-u u biti su prolazni, napisani za jednu analizu podataka (R Core Team, 2019).

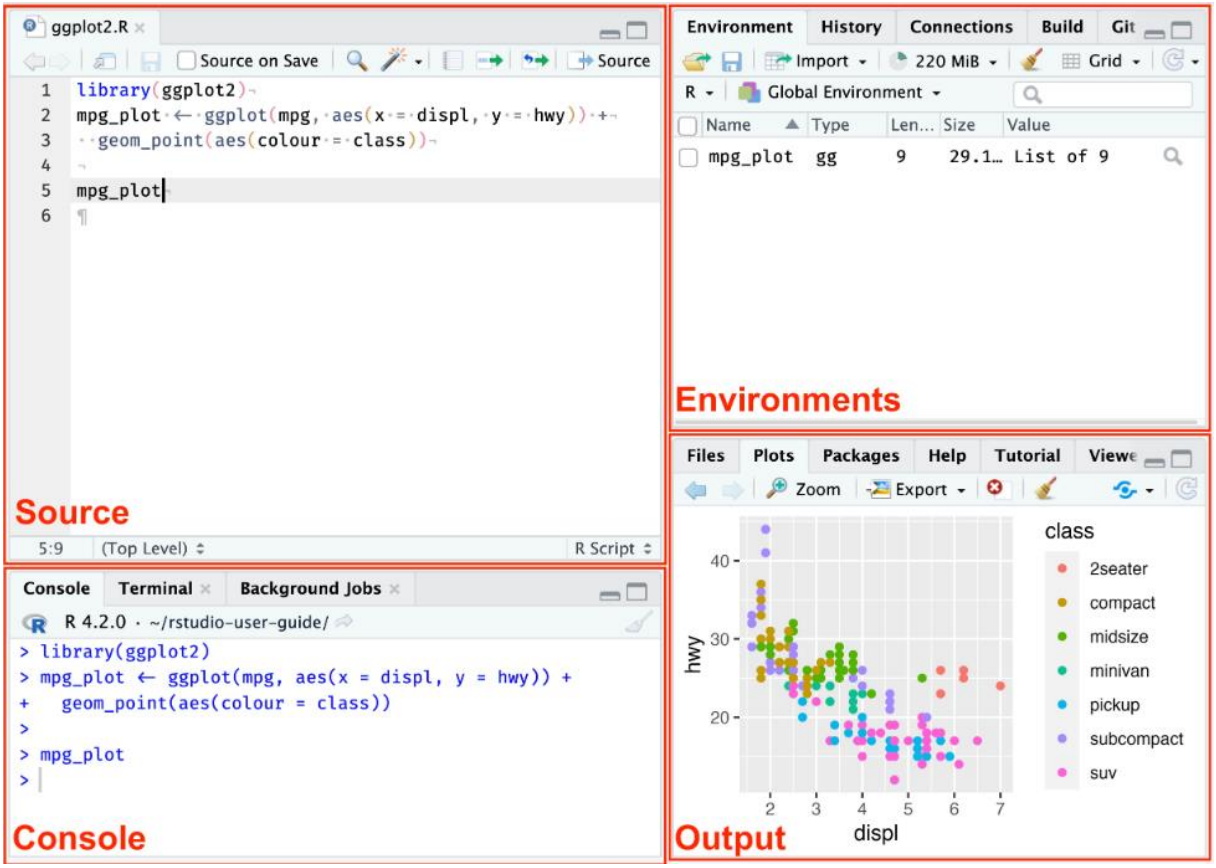
Instalacija R-a i R Studija

Da biste instalirali R, idite na cran.r-project.org i kliknite na preuzimanje R za određeni operativni sustav na vašem računalu (obično Windows) (slika 8.2).



Slika 4.49Stranica za preuzimanje softvera R.

Ovo će preuzeti softver R, a postupak instalacije je isti kao i za bilo koji drugi softver. Kada se R softver instalira, bit će bez naprednog integriranog razvojnog okruženja (engl. *integrated development environment* - IDE) koje pomaže korisnicima u izradi različitih analiza. Iako je moguće napraviti bilo kakvu analizu samo s instaliranim R-om, poželjno ga je upariti s nekim modernim IDE-om, poput RStudija, koji je jedan od najpopularnijih IDE-a. Postupak instaliranja RStudija sličan je osnovnom R softveru. Idite na https://posit.co/download/rstudio-desktop/, potražite RStudio Desktop licencu otvorenog koda, preuzmite je i instalirajte. Nakon instaliranja programa R i RStudio korisnik će imati sljedeći zaslon korisničkog sučelja (slika 8.3).



Slika 4.50Korisničko sučelje i R i Rstudio (RStudio, 2024).

RStudio korisničko sučelje ima 4 primarna prozora (RStudio, 2024.):

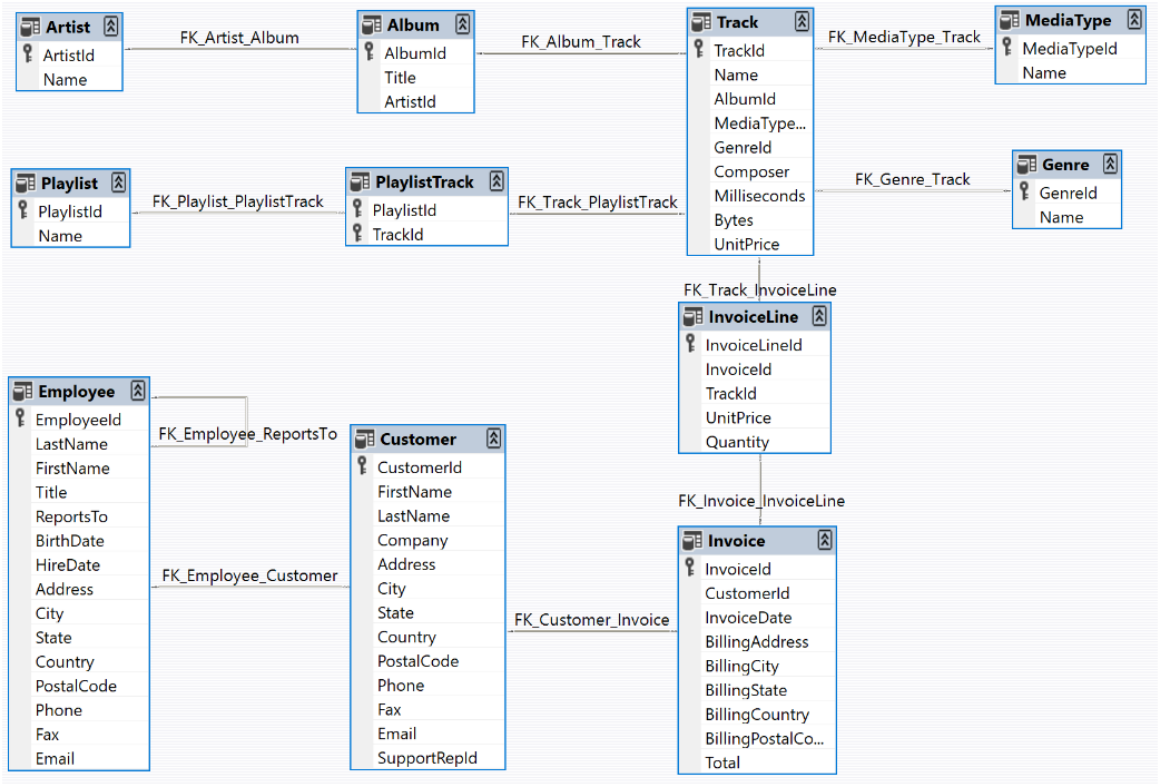
- „Source" (hrv. Izvor);
- „Console" (hrv. Konzola);
- „Environments" (hrv. Okruženje), koji sadrži kartice Okruženje, Povijest, Veze, Izrada, VCS i Vodič;

- „Ouput" (hrv. Rezultat), koji sadrži kartice Datoteke, Ploče, Paketi, Pomoć, Preglednik i Prezentacija.

Svaki od prozora te pododjeljak i opcije u njemu omogućuju korisnicima izvođenje različitih operacija, kontrolu nad nekim analizama podataka ili strukturiraniji i jasniji pogled na proces analitike podataka koji je u tijeku.

8.3 Što je SQL i kako je povezan s BA i R?

Chinook baza podataka je ogledna baza podataka koja se koristi za učenje i demonstraciju sustava za upravljanje bazom podataka (engl. *database management systems* - DBMS) i SQL upita. Dizajniran je kao digitalna multimedijska prodavaonica koja koristi stvarne podatke iz iTunes knjižnice; izmišljena imena/adrese kupaca i zaposlenika; i slučajni podaci za informacije o prodaji. Baza podataka sadrži različite tablice koje predstavljaju podatke glazbene prodavaonice, uključujući informacije o izvođačima, albumima, pjesmama, kupcima, fakturama i više (slika 8.4).



Slika 4.51Model podataka baze podataka Chinook.

Slika 8.4 prikazuje model podataka koji okružuje Chinook bazu podataka s različitim podatkovnim tablicama i njihovim ključevima (jedinstveni identifikatori) i zajedničkim tablicama poput tablice PlaylistTrack. Tablice prenose različite informacije o određenoj digitalnoj prodavaonici (tablica 8.1).

Tablica 4.3Informacije sadržane u svakoj od tablica baze podataka Chinook.

Naziv tablice	Opis
Umjetnik	Sadrži informacije o glazbenim umjetnicima.
Album	Sadrži informacije o glazbenim albumima, od kojih je svaki povezan s izvođačem.
Zapis	Sadrži informacije o pojedinačnim glazbenim zapisima, uključujući reference na albume, vrste medija i žanrove.
Žanr	Sadrži informacije o glazbenim žanrovima.
Vrsta medija	Sadrži informacije o različitim vrstama medija (npr. audio, video).
Kupac	Sadrži podatke o kupcima, uključujući podatke za kontakt i podatke o predstavniku podrške.
Zaposlenik	Sadrži informacije o zaposlenicima, uključujući njihove uloge, odnose s izvješćima i podatke za kontakt.
Fakture	Sadrži informacije o fakturama, uključujući pojedinosti o kupcima, podatke o naplati i ukupne iznose.
Linija fakture	Sadrži detaljne informacije o svakoj stavci na fakturi, uključujući reference na tragove i količine.
Popis pjesama	Sadrži informacije o popisima za reprodukciju.
Zapis popisa pjesama	Povezuje pjesme s popisima za reprodukciju, pokazujući koje su pjesme uključene u koje popise za reprodukciju.

8.4 Kako su poslovna analitika, SQL i R povezani?

Veza između BA, SQL i R je prirodna jer bi svi poslovni podaci trebali biti pohranjeni u SQL bazama podataka. Ovo je još uvijek idealistički cilj budući da još uvijek postoji loše upravljanje podacima u nekim dijelovima malih i srednjih poduzeća koja još uvijek ne razumiju u potpunosti snagu podataka. U velikim tvrtkama to je davno prepoznato i podaci su pravilno strukturirani u bazama podataka (SQL ili neki drugi, ali najčešće u SQL-u). S druge strane, analizu podataka moguće je izvršiti u SQL-u, ali je u tu svrhu bolje koristiti statistički orijentirani softver, gdje je u fokusu R, kao jedna od najpopularnijih statističkih platformi za analizu podataka.

Sukladno tome, SQL i R mogu se smatrati savršenim alatom za suradnju kada je problem u pitanju iz područja BA. Postoji nekoliko glavnih razloga, a jedan od njih je taj što se BA podaci svakodnevno mijenjaju i ažuriraju u skladu s realnim tržištem i aktivnostima tvrtke: prodaja, zaposlenici, prihodi itd. SQL baze podataka savršene su za bilježenje tih promjena i ažuriranje postojećih podataka, dok su R skripte vrlo dobre u automatizaciji zadataka kao i u dizajniranju novih paketa za analizu danih podataka. Razlog za to je što je R više izgrađen oko koncepta analize podataka, nego na općem programiranju kao što je Python, na primjer.

Upit SQL bazi podataka s R

R i SQL baze podataka imaju prirodnu vezu budući da je R uglavnom izgrađen za statističku analizu podataka, a većina transakcijskih podataka nalazi se u bazama podataka. "Način" na koji R radi za upravljanje manipulacijom podataka iz SQL baza podataka je preko DBI i RSQLite programskih paketa. DBI paket pruža standardizirano sučelje za interakciju s različitim DBMS-ovima, dopuštajući korisnicima povezivanje, postavljanje upita i dosljedno upravljanje transakcijama u različitim bazama podataka. Paket RSQLite, koji se pridržava DBI sučelja, posebno olakšava interakciju s bazama podataka SQLite, omogućujući korisnicima izvršavanje SQL upita, dohvaćanje podataka i izvođenje operacija baze podataka izravno iz R-a. Zajedno, ovi paketi pojednostavljuju proces rada s bazama podataka u R-u, nudeći kohezivan i učinkovit tijek rada.

Kako bismo učinkovito demonstrirali izvođenje SQL operacije iz R-a i generiranje željenih uvida iz podataka u vezi s postojećim problemom, dali smo nekoliko isječaka koda na slikama 8.5 i 8.6.

```
---
title: "BUSINES ANALYTICS FOUNDATINS INCLUDING THE R AND SQL"
format: html
editor: visual
---

# R & SQL

## Loading libraries

```{r setup, warning=FALSE, message=FALSE}
library(DBI)
library(RSQLite)
```

Connect to the Chinook SQLite database

```{r}
con <- dbConnect(RSQLite::SQLite(), dbname = "Chinook_Sqlite.sqlite")
```

List all tables in the database

```{r}
tables <- dbListTables(con)
print(tables)
```
```

Slika 4.52Isječak koda za uspostavljanje veze između SQL-a i R-a i istraživanje podatkovnih tablica sadržanih u SQL-u.

Prvi korak u postavljanju upita SQL bazi podataka putem R-a je uspostavljanje veze (slika 8.5). Slika prikazuje korištenje DBI i RSQlite paketa koji omogućuju uspostavljanje veze putem dbConect() funkcije. Rezultat veze i podatkovne tablice koje se otkrivaju putem gore navedene veze zatim se izvoze putem funkcija dbListTables() koje ispisuju popis svih podataka pronađenih putem veze: Album, Izvođač, Kupac, Zaposlenik, Žanr, Fakture, Linija fakture, Vrsta medija, Popis pjesama, Zapis popisa pjesama, Zapis.

Nakon što se veza uspostavi, postoji niz mogućih analiza koje se mogu provesti, ovisno o poslovnom cilju i budućoj upotrebi danih rezultata. Ovdje ćemo, zbog ograničenja prostora, pokazati samo djelić moguće analize podataka, s malim isječkom koda i skupom pravila koda potrebnih za izdvajanje informacija iz SQL-a. Kod vrši upite bazi podataka putem R-a i prikazuje najprodavanije albume, njihove autore i prodani broj (slika 8.6). Tablica 8.2 predstavlja rezultate upita podataka putem isječka koda na slici 8.6.

```
## Choose a Album table from the database
```{r}
query_album <- "SELECT * FROM Album LIMIT 10"
data_album <- dbGetQuery(con, query_album)
print(data_album)
```

Query to get album details along with artist names
```{r}
query_album_artist <- "
SELECT Album.AlbumId, Album.Title AS AlbumTitle, Artist.Name AS ArtistName
FROM Album
JOIN Artist ON Album.ArtistId = Artist.ArtistId
LIMIT 10"

data_album_artist <- dbGetQuery(con, query_album_artist)
print(data_album_artist)
```

Query to get the top-selling albums along with artist names
```{r}
query_top_selling_albums <- "
SELECT
    Album.Title AS AlbumTitle,
    Artist.Name AS ArtistName,
    SUM(InvoiceLine.Quantity) AS TotalQuantitySold
FROM
    InvoiceLine
JOIN
    Track ON InvoiceLine.TrackId = Track.TrackId
JOIN
    Album ON Track.AlbumId = Album.AlbumId
JOIN
    Artist ON Album.ArtistId = Artist.ArtistId
GROUP BY
    Album.AlbumId, Album.Title, Artist.Name
ORDER BY
    TotalQuantitySold DESC
LIMIT 10"

# Execute the query
top_selling_albums <- dbGetQuery(con, query_top_selling_albums)
knitr::kable(top_selling_albums)
```
```

Slika 4.53Isječak koda za postavljanje upita SQL-u putem R-a i određivanje 10 najprodavanijih albuma.

Tablica 4.410 najprodavanijih albuma u digitalnoj prodavaonici Chinook.

Naslov albuma	Ime umjetnika	Prodana količina
Minha Historia	Chico Buarque	27
Greatest Hits	Lenny Kravitz	26
Unplugged	Eric Clapton	25
Acústico	Titãs	22
Greatest Kiss	Kiss	20
Prenda Minha	Caetano Veloso	19
Chronicle, Vol. 2	Creedence Clearwater Revival	19

Naslov albuma	Ime umjetnika	Prodana količina
My Generation - The Very Best Of The Who	The Who	19
International Superhits	Green Day	18
Chronicle, Vol. 1	Creedence Clearwater Revival	18

Literatura 8. poglavlja

- Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. Decision Support Systems, 64, 130-141.
- Kohavi, R., Rothleder, N. J., & Simoudis, E. (2002). Emerging trends in business analytics. Communications of the ACM, 45(8), 45-48.
- Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data and business analytics: A research agenda for realizing business value. *Information & Management*. https://doi.org/10.1016/j.im.2019.103237
- Power, D. J., Heavin, C., McDermott, J., & Daly, M. (2018). Defining business analytics: An empirical approach. Journal of Decision Systems, 27(1), 40–53. https://doi.org/10.1080/2573234X.2018.1507605
- R Core Team. (2019). An introduction to R: Notes on R, a programming environment for data analysis and graphics. The R Foundation.
- RStudio. (2024). RStudio IDE cheatsheet: UI panes. Posit. https://docs.posit.co/ide/user/ide/guide/ui/ui-panes.html
- Torfs, P., & Brauer, C. (2014). A (very) short introduction to R. Hydrology and Quantitative Water Management Group, Wageningen University, The Netherlands, 1-12.

9. Predviđanje potražnje, vizualizacija i inženjering značajki vremenskih serija u opskrbnim lancima

Što je predviđanje potražnje? Kako možemo učinkovito vizualizirati podatke o kupcima i donijeti zaključke o njima? Kako provesti inženjering značajki vremenskih serija?

Na ova i slična pitanja pokušat ćemo dati odgovore u sljedećem poglavlju.

9.1 Što je potražnja kupaca i predviđanje potražnje?

Zahtjev krajnjeg kupca pokreće cijeli opskrbni lanac (Syntetos i dr., 2016). Sukladno tome, potražnja kupaca je ključna komponenta za planiranje svih logističkih procesa u opskrbnom lancu pa je određivanje razine potražnje kupaca od velikog interesa za menadžere opskrbnog lanca. Komplementarno, predviđanje potražnje bitna je aktivnost za planiranje i raspoređivanje logističkih aktivnosti unutar promatranog opskrbnog lanca (Mircetic i dr., 2017). Precizni modeli predviđanja potražnje izravno utječu na smanjenje logističkih troškova budući da daju procjenu potražnje kupaca (Mircetic i dr., 2016). Predviđanje u opskrbnim lancima nadilazi operativni zadatak ekstrapolacije zahtjeva potražnje na jednom nivou. Uključuje složena pitanja kao što su koordinacija opskrbnog lanca i dijeljenje informacija između višestrukih dionika (Syntetos i dr., 2016).

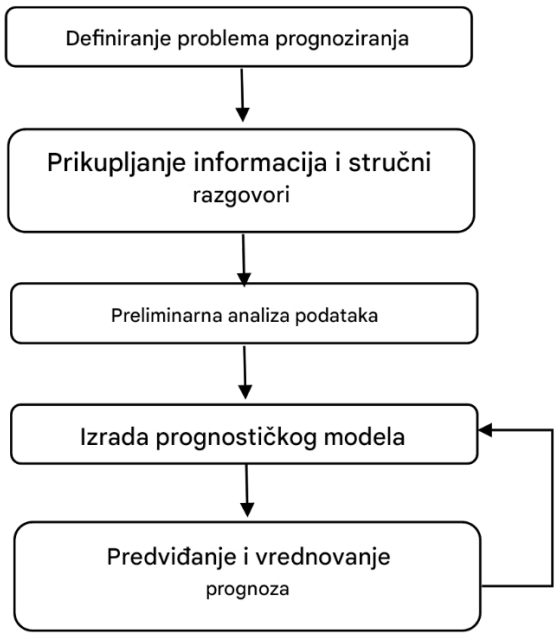
Potražnja kupaca i popratne prognoze od vitalnog su značaja za opskrbne lance, budući da pružaju osnovne ulazne podatke za planiranje i kontrolu svih funkcionalnih područja, uključujući logistiku, marketing, proizvodnju itd. (Mircetic, 2018). Kad bi potražnja krajnjih kupaca bila stalna, ili poznata sa sigurnošću puno unaprijed, tada bi rad opskrbnog lanca bio izravna (unatrag) vježba planiranja. Međutim, potražnja nije poznata i stoga je treba predvidjeti. Nesigurnost povezana s ovom potražnjom čini upravljanje opskrbnim lancem vrlo teškim (Syntetos i dr., 2016). Na učinkovitost predviđanja potražnje utječu inherentne neizvjesnosti u vremenskoj seriji potražnje koje opskrbni lanci imaju (Rostami-Tabar, 2013.).

Posljedično, rješavanje i razumijevanje ovih neizvjesnosti veliki je izazov za menadžere prilikom koordinacije i planiranja operacija unutar opskrbnih lanaca (Mircetic, 2018).

Nesigurnost potražnje jedan je od najznačajnijih izazova za suvremene opskrbne lance. Nedavna pandemija COVID-19 dodatno je naglasila ovaj problem, uzrokujući široko rasprostranjene poremećaje koji su komplicirali planiranje i kontrolu opskrbnog lanca (Nikolopoulos i dr., 2020). Predviđanje potražnje u opskrbnim lancima često uključuje predviđanje potražnje za brojnim artiklima. Prognostičari u opskrbnim lancima obično ekstrapoliraju vremenske serije podataka za svaku jedinicu za čuvanje zaliha (engl. *stock-keeping unit* – SKU) pojedinačno. Na primjer, trgovac na malo može koristiti podatke o prodajnom mjestu za generiranje predviđanja na razini pojedinačne prodavaonice (Mircetic i dr., 2022).

9.2 Koraci predviđanja potražnje u opskrbnom lancu?

U skladu s gore navedenim izjavama i zaključcima, u pogledu važnosti potražnje i predviđanja potražnje za opskrbne lance, važno je slijediti specifične korake pri izradi prognoza u opskrbnim lancima (slika 9.1).



Slika 4.54Osnovni koraci za pravilnu implementaciju predviđanja unutar poduzeća (Makridakis i dr., 1998; Makridakis i dr., 1983).

Svaki od koraka na slici 9.1 ima svoje prednosti i doprinos stvaranju pouzdanih i korisnih (poslovno orijentiranih) prognoza. Sukladno tome, definiranje problema često je najteži dio predviđanja i zahtijeva razumijevanje načina na koji će se predviđanja koristiti, kao i uloge funkcija predviđanja unutar promatranog poduzeća. Prognostičar bi trebao potrošiti dosta vremena na komunikaciju sa svima koji su uključeni u prikupljanje podataka, održavanje baze podataka i korištenje prognoza za buduće planiranje. Jedan od glavnih otežavajućih čimbenika u definiranju problema je kako će se konačna prognoza koristiti u svakodnevnim logističkim operacijama (koja platforma, dizajn softvera, korisničko sučelje itd.).

Za korak prikupljanja informacija uvijek su potrebne najmanje dvije vrste informacija: statistički podaci i akumulirana stručnost ljudi koji prikupljaju podatke i koriste predviđanja. U praksi je često teško dobiti povijesne podatke za stvaranje dobrog statističkog modela. Također, postoji veliki nesporazum o tome što su podaci o potražnji i što se može koristiti kao njihova zamjena. Postoji loša praksa korištenja podataka o otpremi i isporuci kao zamjene za podatke o potražnji, što će samo pogoršati proces donošenja odluka na temelju predviđanja napravljenih na jednostavnoj vrsti podataka. Podaci o prodaji jedina su pouzdana zamjena za podatke o potražnji (Syntetos i dr., 2016), iako ovo pojednostavljenje nije savršeno, osobito u opskrbnim lancima s puno situacija kada nema zaliha (engl. *out-of-stock* – OOS).

Za korak preliminarne analize podataka, preporučuje se uvijek započeti analizu podataka s grafičkim prikazima kako bi se odgovorilo na sljedeća pitanja. Postoje li dosljedni obrasci? Postoji li značajan trend? Postoji li primjetna sezonalnost? Postoje li dokazi o poslovnim ciklusima? Koliko su jaki odnosi između varijabli? Ovo su pitanja na koja jednostavna grafika može dati odgovore i omogućiti daljnju analizu podataka sužavanjem fokusa koje modele primijeniti na otkrivena obilježja potražnje. Obično, jednostavni model, određen na ovaj način, može pobijediti one sofisticiranije i kompliciranije (Rostami-Tabar & Mircetic, 2023).

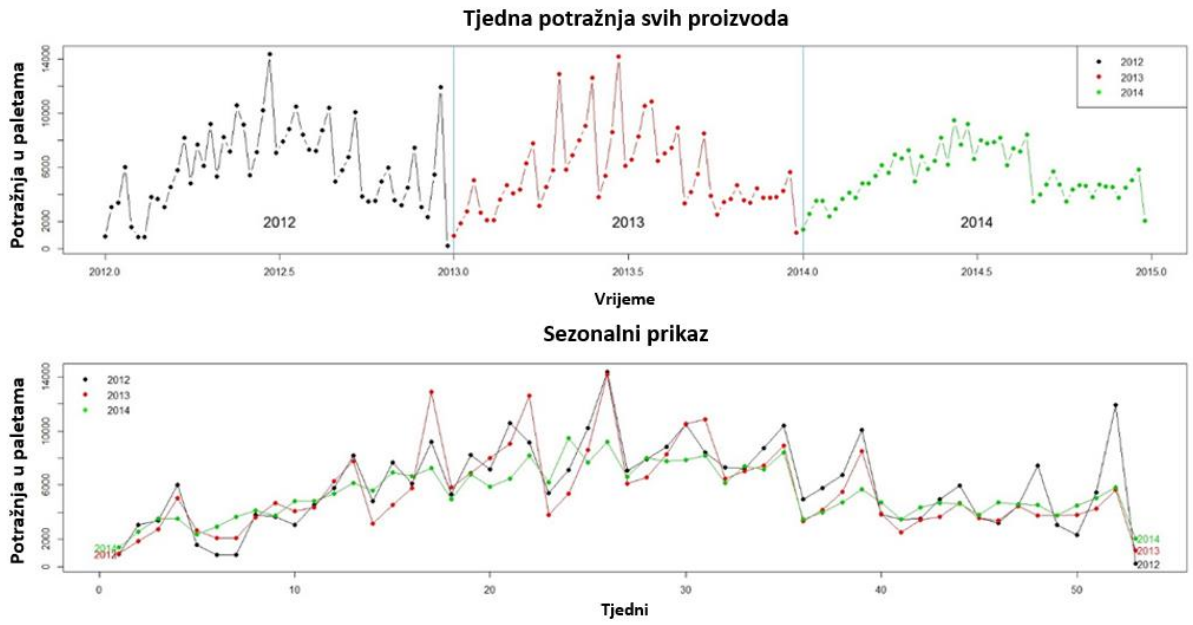
Odabir i izrada modela predviđanja najvažniji je korak pri izradi modela predviđanja. Koji model koristiti ovisi o nekoliko čimbenika, od kojih su najvažniji dostupnost povijesnih podataka i korelacija između zavisnih i nezavisnih varijabli. Uobičajeno je da se prilikom odabira modela uspoređuju dva ili tri potencijalna modela. Svaki je model umjetna konstrukcija temeljena na skupu pretpostavki (eksplicitnih i implicitnih) i općenito uključuje jedan ili više parametara koji se moraju izraditi korištenjem poznatih povijesnih podataka. U sljedećem poglavlju bit će

prikazan proces razvoja i primjene modela ARIMA, kao jednog od najboljih i najpopularnijih modela.

Ocjenjivanje modela predviđanja je korak kojim se mjeri upotrebljivost izrađenog modela. Nakon odabira prognostičkog modela i procjene njegovih parametara, model se koristi za izradu prognoza. Točnost modela procjenjuje se korištenjem različitih statistika, ali također je važno testirati predviđanja putem mjerila poslovnih implikacija (tj. metrike korisnosti).

9.3 Predviđanje potražnje u prehrambenoj industriji

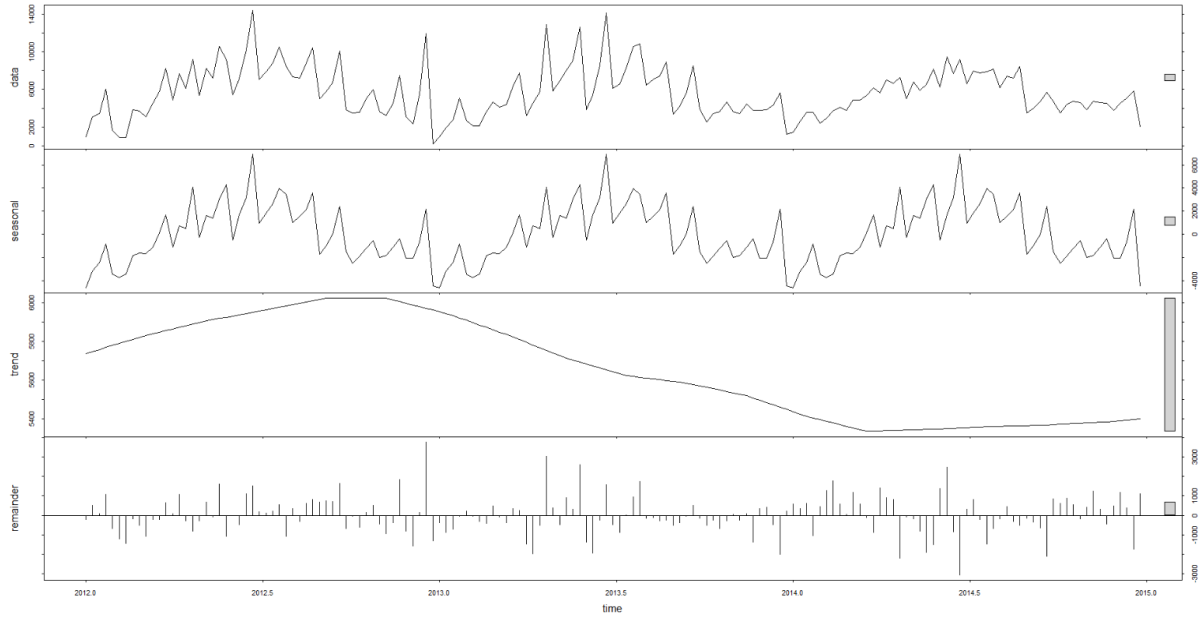
Podaci o potrošnji za sve proizvode promatranog poduzeća u prehrambenoj industriji prikazani su kao tjedna potražnja u rasponu od siječnja 2012. do prosinca 2014. (slika 9.2). X os predstavlja vrijeme, dok su vrijednosti potražnje prikazane na y osi. Ovi podaci prikazani su u tjednim intervalima, jer to odgovara razdoblju u kojem se vrši opskrba krajnjih prodajnih mjesta. Slijedom toga, menadžment tvrtke fokusiran je na predviđanje tjedne potrošnje tržišta.



Slika 4.55Podaci o potražnji za promatranu prehrambenu tvrtku.

Slika 9.2 pokazuje nekoliko važnih obilježja i karakteristika danih podataka. Prvo, podaci imaju snažan sezonski karakter s vršnom prodajom koja se događa sredinom godine (ljetni mjeseci). Drugo, podsezonski grafikon (donji grafikon na slici 9.2) prikazuje značajnu promjenu uzorka trenda pada u 2014.! Ovo je vrlo značajna karakteristika za odabir pravog modela predviđanja

i za menadžere u poduzeću budući da otkriva značajan pad potrošnje i gubitak tržišta. Za daljnje istraživanje promatranih karakteristika koristi se metodologija dekompozicije sezonskog trenda (engl. *seasonal trend decomposition* - STL) (slika 9.3). STL dekompozicija dijeli izvorne obrasce potražnje na tri komponente: sezonsku, trend i ostatak.



Slika 4.56STL dekompozicija podataka o potražnji.

STL je otkrio da promatrane vremenske serije pokazuju aditivnu prirodu, što znači da se fluktuacije oko krivulje trend-ciklus ne povećavaju značajno tijekom vremena. Kao rezultat toga, Box-Coxove transformacije nisu bile potrebne za neobrađene vremenske serije. Dekompozicija je otkrila da je sezonska komponenta dominantna u promatranom nizu, pokazujući visoke fluktuacije unutar jedne godine. S obzirom na ograničeni broj godina promatranja, teško je identificirati poslovni ciklus. Dekompozicija je također pokazala da je trend u seriji minimalan, s opadajućim uzorkom počevši od sredine 2013.

Ove identificirane karakteristike predstavljale su važan input tijekom procesa dizajna odgovarajućeg prognostičkog modela. U tu svrhu odabran je model S-ARIMA (engl. *Seasonal Autoregressive Integrated Moving Average*). Model S-ARIMA vrlo je učinkovit za predviđanje jer kombinira i autoregresivnu komponentu i komponentu pomičnog prosjeka, zajedno s razlikom kako bi podaci bili stacionarni. Ovaj je model posebno vješt u hvatanju i modeliranju sezonskih obrazaca u podacima o vremenskoj seriji, što ga čini idealnim za industrije s cikličkim obrascima potražnje, kao što je prehrambena industrija. Dodatno, sposobnost S-ARIMA-e da

se nosi sa složenim sezonskim strukturama i trendovima omogućuje preciznije i pouzdanije prognoze, koje su ključne za učinkovito planiranje opskrbnog lanca i upravljanje zalihama. S-ARIMA strukturni oblik predstavljen je u jednadžbi (1).

$$\Phi(B^m)\phi(B)(1-B^m)^D(1-B)^d y_t = c + \Theta(B^m)\theta(B)\varepsilon_t, \quad (1)$$

gdje su $\Phi(z)$ i $\Theta(z)$ polinomi reda *P* odnosno *Q*, od kojih svaki ne sadrži korijene unutar jedinične kružnice. B je operator pomaka unazad koji se koristi za opisivanje procesa diferenciranja, tj. $By_t = y_{t-1}$. Ako je $c \neq 0$, postoji implicirani polinom reda $d + D$ u funkciji prognoze. Budući da je S-ARIMA visoko parametrizirani model, ključno pitanje pri korištenju S-ARIMA modela je odabir odgovarajućeg redoslijeda modela, što uključuje određivanje vrijednosti *p, q, P, Q, D* i *d*. Ako su *d* i *D* poznati, poredak *p, q, P* i *Q* može se odabrati korištenjem informacijskog kriterija kao što je Akaikeov informacijski kriterij (AIC) ili Bayesov informacijski kriterij (BIC). Formule za AIC i BIC dane su prema:

$$AIC = -2\log(L) + 2(k);$$

$$BIC = N\log\left(\frac{SSE}{N}\right) + (k+2)\log(N) \quad (2)$$

gdje *je* $k=p+q+P+Q+1$ ako je uključen konstantni izraz i 0, inače, *L* je maksimizirana vjerojatnost modela prilagođenog različitim podacima, *SSE* je zbroj kvadrata pogrešaka, *N* je broj opažanja koji se koristi za procjenu, a *k* je broj prediktora u modelu.

Za određivanje optimalnog skupa parametara Hyndman i Khandakar (2007) predložili su Canova-Hansen i KPSS test jediničnog korijena kroz sljedeće korake:

- Upotrijebite Canova-Hansen test za određivanje *D* u okviru ARIMA.
- Odaberite *d* primjenom uzastopnog KPSS testa jediničnog korijena na sezonski diferencirane podatke (ako je $D = 1$) ili na izvorne podatke (ako je $D = 0$).
- Odaberite optimalne vrijednosti za *p, q, P* i *Q* minimiziranjem AIC-a.

9.4 Razvoj modela predviđanja S-ARIMA

Razvoj S-ARIMA prognostičkog modela

U skladu s gore navedenim postupkom ispitano je nekoliko postavki parametara, a njihova izvedba prikazana je u tablici 9.1. Za testiranje performansi različitih modela koristi se nekoliko mjera: srednja apsolutna postotna pogreška (engl. *mean absolute percentage error* - MAPE), korijen srednje kvadratne pogreške (engl. *root mean square error* - RMSE), srednja apsolutna skalirana pogreška (engl. *mean absolute scaled error* - MASE), AIC i BIC (jednadžbe 2 i 3)

$$MAPE = \frac{1}{N}\sum_{1}^{N}|e_i|;$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{1}^{N}e_i};$$

$$MASE = \frac{1}{N}\sum_{1}^{N}|q_j|, \qquad (3)$$

gdje su e_i reziduali i $q_j = \frac{e_j}{\frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|}$.

Najpopularnije mjere za menadžere u opskrbnim lancima su RMSE i MAPE jer daju "osjećaj" koliko je model dobar u realnim brojevima (RMSE) i postocima (MAPE).

Tablica 4.5Izvedba S-ARIMA modela s različitim postavkama parametara.

Modeli [a]	RMSE	MAPE	MASE	AIC	BIC
S- ARIMA (5,0,1)(1,0,0) $_{52}$ [b]	1023	14,18 %	0,60 %	86.62	110.5 0
S- ARIMA (4,0,0)(1,0,0) $_{52}$ [c]	1041	14,53 %	0,62 %	86.73	105.31
S- ARIMA (4,0,0)(0,1,1) $_{52}$ [d]	1882. godine	22,12 %	1,05 %	41.74	53,56 0
S- ARIMA (4,0,1)(1,0,0) $_{52}$ [e]	1050	14,18 %	0,60 %	88.23	109.47
S- ARIMA (0,0,1)(0,1,0) $_{52}$ [f]	1797. godine	21,42 %	1,02 %	36.52	40,46 0

[a] Pogreške modela izračunavaju se na skupu testnih podataka.

[b] Detalji o modelu S- ARIMA (5,0,1)(1,0,0) $_{52}$ navedeni su u nastavku .

[c] $(1-0.39B+0.06B+0.04B-0.46B)(1-0.65B^{52})y_t = 8.52$.

[d] $(1-0.29B+0.19B-0.05B-0.19B)(1-B^{52})y_t = (1+0.12B^{52})e_t$.

[e] $(1-0.29B+0.01B+0.05B-0.48B)(1-0.65B^{52})y_t = 8.52+(1+0.13B)e_t$.

[f] $(1-B^{52})y_t = (1+0.3B)e_t$.

Tablica 9.1 pokazuje da je model S-ARIMA (5,0,1)(1,0,0) $_{52}$ nadmašio konkurentske modele postižući najniže pogreške RMSE, MAPE i MASE. Oblik modela S-ARIMA (5,0,1)(1,0,0) $_{52}$ predstavljen je u jednadžbi (4), gdje je $\phi_1 = -0{,}5421$, $\phi_2 = 0{,}2962$, $\phi_3 = -0{,}099$, $\phi_4 = 0{,}3974$, $\phi_5 = 0{,}4994$, $\Phi_1 = 0{,}9558$, $c = 8{,}523$, i $\Theta_1 = 0{,}6345$.

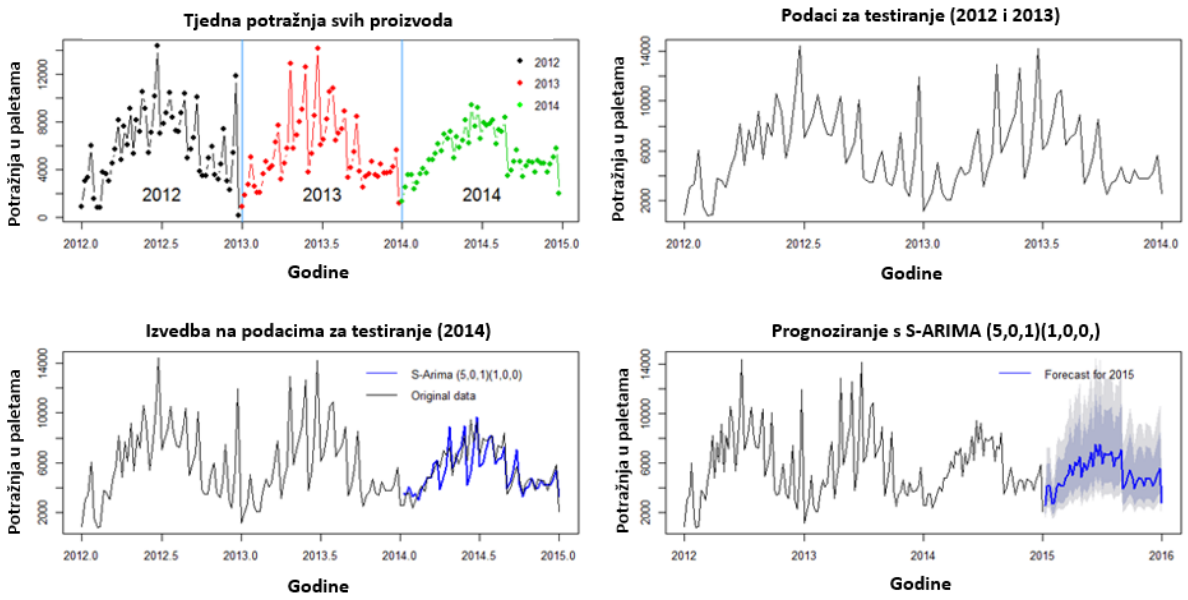
$$(1-\phi_1 B-\phi_2 B-\phi_3 B-\phi_4 B-\phi_5 B)(1-\Phi_1 B^{52})y_t = c+(1+\theta_1 B)e_t, \quad (4)$$

Slična izvedba primijećena je s modelima S-ARIMA (4,0,0)(1,0,0) $_{52}$ i S-ARIMA (4,0,1)(1,0,0) $_{52}$. Prilikom izvođenja predviđanja, model S-ARIMA (5,0,1)(1,0,0) $_{52}$ u prosjeku proizvodi pogrešku od 1023 proizvoda, što znači 14,18 %. Ovi rezultati naglašavaju važnost pažljivog odabira uvjeta za uključivanje u S-ARIMA model budući da nema značajnih razlika između performansi modela s različitim parametrima. Procjena je pokazala da su modeli koji uključuju autoregresivne (*p, P*) i komponente pomičnog prosjeka (*q, Q*) učinkovitiji u predviđanju potrošnje pića od onih koji uključuju sezonske ili nesezonske razlike. Osim toga, usporedni pregled otkrio je neke neočekivane nalaze, kao što su modeli koji uključuju sezonsku razliku S-ARIMA (4,0,0)(0,1,1) $_{52}$ i S-ARIMA (0,0,1)(0,1 ,0) $_{52}$ imaju lošiju izvedbu od jednostavnog prosječnog modela naivne prognoze, što dokazuje njihova MASE pogreška veća od jedinice.

9.5 Predviđanja buduće potražnje

Slika 9.4 prikazuje izvedbu S-ARIMA (5,0,1)(1,0,0) $_{52.}$ Gornji lijevi prikaz predstavlja početne podatke o potražnji, obojene radi lakšeg razlikovanja različitih marketinških godina. Gornji desni prikaz su ulazni podaci za S-ARIMA (5,0,1)(1,0,0) $_{52}$, koji predstavljaju podatke o obuci,

tj. testiranje za koje se parametri u S-ARIMA određuju prema postupku opisanom u potpoglavlju 9.3. Ovo je vrlo važno za razumijevanje koliko je težak posao modela predviđanja, budući da u ovom slučaju ima dvije godine podataka kao ulaz i treba predvidjeti buduću potražnju godinu dana unaprijed! Ovo je prilično čest scenarij u opskrbnim lancima i logistici jer postoji nepisano pravilo da tvrtke čuvaju povijest svojih podataka tri godine nakon čega odbacuju podatke.



Slika 4.57Obuka, testiranje i predviđena potražnja S-ARIMA (5,0,1)(1,0,0) $_{52}$ modela.

Donji lijevi prikaz na slici 9.4 predstavlja performanse promatranog modela. Učinkovitost modela već je prikazana u tablici 9.1 putem različitih statističkih mjera, ali menadžerima je obično teško steći dojam koliko je model dobar ili loš. U tu svrhu prikaz grafički prikazuje performanse modela. Moglo bi se tvrditi da model prilično dobro prati testne podatke i u većini razdoblja pokazuje izvrsne performanse. Kako bi se generirale buduće prognoze, S-ARIMA (5,0,1)(1,0,0) $_{52}$ je preuređen dodavanjem podataka iz 2014. godine. Nakon toga, model je proizveo prognoze za 52 tjedna unaprijed za 2015. godinu. Prognoze za 2015. godinu prikazane su u donjem desnom prikazu. Prognoze su popraćene intervalima predviđanja od 80% i 95%, koji pokazuju da se moguće buduće prognoze razlikuju od srednjih predviđenih vrijednosti. Model predviđa kontinuirani pad potražnje koji je započeo 2014. Uzroci ovog pada mogu biti različiti i trebali bi ih dodatno istražiti menadžeri na strateškoj razini poduzeća.

Literatura 9. poglavlja

- Hyndman, R., & Khandakar, Y. (2007). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3). http://www.jstatsoft.org/v27/i03
- Makridakis, S., Wheelwright, S., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). New York: John Wiley & Sons.
- Makridakis, S., Wheelwright, S., & McGee, V. (1983). *Forecasting: Methods and applications* (2nd ed.). New York: John Wiley & Sons.
- Mircetic, D. (2018). *Boosting the performance of top-down methodology for forecasting in supply chains via a new approach for determining disaggregating proportions*. University of Novi Sad.
- Mircetic, D., Nikolicic, S., Maslaric, M., Ralevic, N., & Debelic, B. (2016). Development of S-ARIMA model for forecasting demand in a beverage supply chain. *Open Engineering, 6*(1).
- Mircetic, D., Rostami-Tabar, B., Nikolicic, S., & Maslaric, M. (2022). Forecasting hierarchical time series in supply chains: An empirical investigation. *International Journal of Production Research, 60*(8), 2514-2533.
- Mircetic, D., Nikolicic, S., Stojanovic, D., & Maslaric, M. (2017). Modified top-down approach for hierarchical forecasting in a beverage supply chain. *Transportation Research Procedia, 22*, 193–202.
- Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., & Vasilakis, C. (2020). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. *European Journal of Operational Research, 290*(1), 99–115.
- Rostami-Tabar, B. (2013). *ARIMA demand forecasting by aggregation*. Université Sciences et Technologies Bordeaux I.
- Rostami-Tabar, B., & Mircetic, D. (2023). Exploring the association between time series features and forecasting by temporal aggregation using machine learning. *Neurocomputing, 548*, 126376.

- Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research, 252*(1), 1-26.

10. Umjetna inteligencija i strojno učenje u opskrbnim lancima

Što je umjetna inteligencija (AI)? Je li to stvarno "živo" stvorenje sposobno razmišljati i donositi vlastite odluke na temelju svog uma, prošlih iskustava, etike, uvjerenja itd.? Kako je to povezano sa strojnim učenjem (engl. *machine learning* - ML)? Jesu li AI i ML ista stvar?

Koji se alati koriste u AI & ML? Koju ulogu imaju AI i ML u poslovnom kontekstu i kako se mogu koristiti u svakodnevnim poslovnim operacijama i optimizacijama? Postoje li specifične arhitekture i primjeri primjene AI-a i ML-a na opskrbne lance?

Na ova i slična pitanja pokušat ćemo dati odgovore u sljedećem poglavlju, završavajući stvarnim primjerom studije slučaja primjene AI & ML algoritama u distribucijskom skladištu.

10.1 Što je umjetna inteligencija?

Područje umjetne inteligencije započelo je 1950-ih kada su se računalni znanstvenici zapitali: "Mogu li računala razmišljati kao ljudi"? Istraživači su u to vrijeme bili oduševljeni mogućnošću učenja računala da izvršavaju složene zadatke i u skladu s tim razvili su skup različitih algoritama za tu svrhu. Definicija ovog polja mogla bi se navesti kao napor da se automatiziraju intelektualni zadaci koje obično obavljaju ljudi (Chollet, 2021). Algoritmi u području umjetne inteligencije danas dolaze iz ML-a i dubokog učenja, koji su podskupovi AI-a. Osim ML-a i dubokog učenja, AI uključuje i mnogo algoritama koji nisu učeći. Štoviše, u ranoj fazi razvoja umjetne inteligencije, ovi su algoritmi bili dominantniji. Prema tome, taj dio umjetne inteligencije poznat je kao simbolička umjetna inteligencija koja se temelji na ideji da se može postići ljudska razina performansi i inteligencije programiranjem računala s velikim skupom eksplicitnih pravila za rješavanje promatranog problema. Ovaj pristup dao je izvrsne rezultate u logičkom problemu koji je bio dobro definiran, poput računala koje igra šah, ali se pokazao kompliciranim za složenije probleme. Stvarni svijet se pokazao puno kompliciranijim nego što bi se sva eksplicitna pravila programiranja mogla unijeti u računalo. Ovaj pristup je usredotočen na ideju za danu situaciju - učini ovo ili ono (ako-onda pravila). Ovo je lako razumljiv pristup, ali s druge strane vrlo dugotrajan i ponekad je jako teško odrediti sve moguće

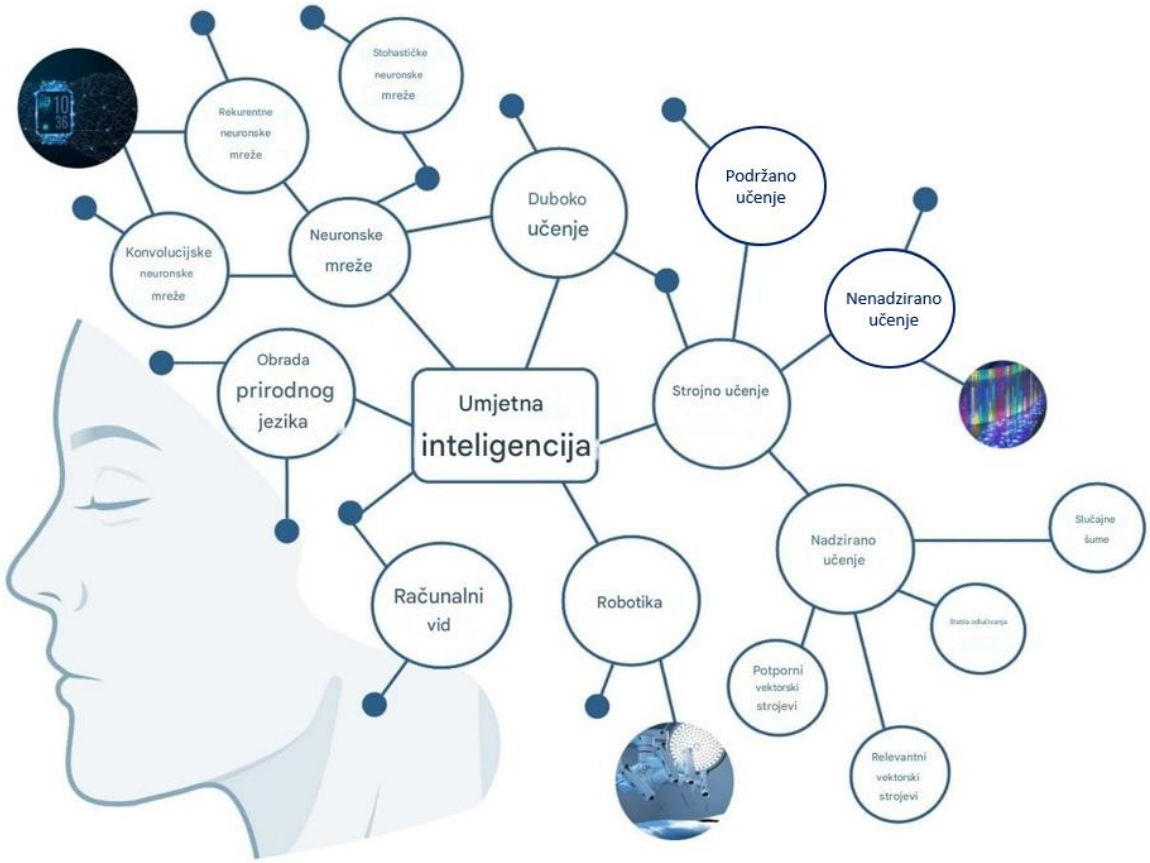
scenarije koje je potrebno unijeti u program. Kao smiješan primjer, ali dobra ilustracija ove teme, pogledajte sliku 10.1 koja prikazuje nekoliko scenarija za određivanje prognoze na temelju statusa kamena.

Slika 4.58Ako – Onda pravila programiranja (Gibbs, 2019).

Područje simboličke umjetne inteligencije najveću je popularnost steklo 1980-ih s pojavom ekspertnih sustava (ES). ES predstavljaju podskup sustava za podršku odlučivanju (DSS) (Turban, 1998), usmjerenih na pružanje računalnih sposobnosti donošenja odluka sličnih onima stručnjaka unutar određenog područja. Ovi su sustavi izrađeni za rješavanje zamršenih problema upotrebom niza pravila ili algoritama koji simuliraju procese ljudskog zaključivanja. Olson i Courtney (1992) opisuju ekspertne sustave (ES) kao računalne programe koji simuliraju ljudske misaone procese za donošenje odluka unutar određene domene, uključujući stupanj umjetne inteligencije koji odgovara zaključcima do kojih bi ljudski stručnjak došao. ES komponenta je posebno korisna za podršku donositeljima odluka u područjima koja zahtijevaju specijalizirana znanja (Turban i dr., 2005). U osnovi, ES prikuplja ekspertizu ljudskog stručnjaka (ili drugog izvora) i prenosi je na računalo. Ova tehnologija može ili pomoći donositeljima odluka ili ih u potpunosti zamijeniti, što je čini jednim od najšire primjenjivanih i komercijalno najuspješnijih oblika umjetne inteligencije (Turban i dr., 2005). Jedan od ključnih razloga za razvoj ES-a je distribucija stručnog znanja široj publici (Jackson, 1999). U sljedećim potpoglavljima, demonstrirat ćemo primjenu ES-a temeljenog na AI i ML algoritmima, u centralnom skladištu kao dijelu cjelokupnog DSS-a menadžerima.

Danas se područje umjetne inteligencije sastoji od različitih pristupa i algoritama, no oni koji se najviše koriste opisani su na slici 10.2. Razišlo se od ES sustava, a za ponovni razvoj ovog

područja najviše su zaslužni algoritmi dubokog učenja koji su u posljednjih 12 godina imali značajan uspjeh u problemima prepoznavanja slike, prepoznavanja govora, segmentacije slike, prepoznavanja lica itd. Duboko učenje koristi više slojeva apstrakcije za prepoznavanje složenih uzoraka u visokodimenzionalnim podacima. Ovaj pristup je postigao značajan napredak u područjima kao što su prepoznavanje govora i slike, otkrivanje lijekova i obrada prirodnog jezika. Duboko učenje ima sposobnost automatskog otkrivanja relevantnih obilježja smanjuje potrebu za ljudskom intervencijom u dizajnu obilježja, što ga čini vrlo učinkovitim u iskorištavanju velikih skupova podataka i računalne snage (LeCun i dr., 2015).



Slika 4.59Glavna polja i potpolja AI (Athanasopoulou i dr., 2022).

Veliki korak prema današnjoj situaciji bilo je istraživanje klasifikacije znamenki koje su proveli Hinton i dr. (2006), koje je uspjelo postići više od 98% točnosti na klasifikaciji baze podataka Modificiranog nacionalnog instituta za standarde i tehnologiju (engl. *Modified National Institute of Standards and Technology* - MNIST). Jedan od načina razmišljanja o tome kako je umjetna inteligencija prešla iz simboličke umjetne inteligencije u ML i što je bit ML-a jest zamisliti ML

algoritme kao amorfnu masu koja se sama oblikuje prema željenim ishodima. Sustav pravila koji uzima input u output mijenja se od problema do problema i prilagođava postojećem stanju. Cilj mu je pronaći pravila koja će automatizirati zadatak traženjem statističkih obrazaca unutar podataka. Ovakav pristup rješavanju različitih problema značajno smanjuje vrijeme postavljanja sustava (u usporedbi s ako – onda pravilima) i čini ga univerzalnijim pristupom za rješavanje različitih problema.

Bengio i dr. (2021) naglasili su da je budućnost umjetne inteligencije u dubokom učenju i revolucionarnom utjecaju meke pažnje i transformatorskih arhitektura u umjetnoj inteligenciji. Ove inovacije omogućuju neuronskim mrežama da se dinamički fokusiraju na važne ulaze i pohranjuju informacije u diferencijalne memorije, značajno poboljšavajući sekvencijalnu obradu.

10.2 Što je ekosustav AI-a i ML-a?

Ekosustav AI i ML algoritama sastoji se od tri ključna stupa:

- Ulazni podaci;
- Izlazni podaci;
- Funkcija troška.

Ulazni podaci predstavljaju zapise podataka određene značajke ili značajki (ovisno o uočenom problemu). Točnost ulaznih podataka ključna je za izgradnju točnih algoritama. To često nije slučaj u stvarnim aplikacijama i obično se značajno vrijeme i trud posvećuju prikupljanju podataka, čišćenju, argumentiranju, objedinjavanju, provjeri lažnih unosa itd. Osim točnih podataka, drugi važan aspekt karakteristika ulaznih podataka je njihova reprezentacija i kodiranje. Različiti načini kodiranja podataka mogu otkriti različite značajke podataka i značajno "pomoći" ML modelima u otkrivanju skrivenih obrazaca u podacima. Ovdje vidimo Ahilovu petu ML algoritama. Često se previše pozornosti posvećuje kreiranju metoda za izvlačenje informacija i inteligencije iz podataka (tj. samih algoritama), dok se premalo pozornosti posvećuje ulaznim podacima i njihovom odnosu s izlaznim podacima. Općenito se podrazumijeva da su ulazni podaci u uzročno-posljedičnoj vezi s izlaznim podacima, što ponekad uopće nije slučaj. Stoga bi sljedeći veliki korak u razvoju ML-a trebao biti pronalaženje boljih načina za prikupljanje, predstavljanje i kodiranje podataka.

Izlazni podaci predstavljaju mjerenja određenog problema koji pokušavamo riješiti. U problemu klasifikacije, to bi bila oznaka klasa. U regresiji, to će biti stvarni broj koji pokušavamo predvidjeti. U kontekstu specifičnog problema, za probleme s prepoznavanjem govora, izlazni podaci mogu biti ljudski generirani prijepisi zvučnih datoteka. U problemu prepoznavanja slike, izlaz može biti oznaka klase slike, itd.

Troškovna funkcija predstavlja način mjerenja izvedbe AI i ML. U osnovi, idealno bismo željeli da odgovori algoritama odgovaraju izlaznim podacima, za dane ulazne podatke. Troškovna funkcija također je povratni signal skupu parametara koji usmjeravaju rad algoritma, tj. omogućuje optimizaciju ukupne izvedbe algoritma kroz proces učenja (pronalaženje optimalnog skupa parametara). Proces učenja obično uključuje nadzirano učenje, gdje se model trenira na označenim podacima kako bi se smanjile pogreške predviđanja pomoću tehnika kao što su stohastički gradijentni spuštanje i širenje unazad. To omogućuje modelu da učinkovito prilagodi svoje unutarnje parametre, što dovodi do poboljšane izvedbe zadataka kao što su otkrivanje i klasifikacija objekata (LeCun i dr. 2015).

10.3 Koji se alati koriste u ML-u?

Općenito, ML algoritmi mogu se klasificirati u dvije glavne kategorije: nadzirano i nenadzirano učenje.

Nadzirano učenje uključuje uvježbavanje algoritama na označenom skupu podataka, gdje je svaka ulazna podatkovna točka uparena s točnim izlazom. Ova jasna "slika" o tome kakav bi trebao biti točan odgovor za dati ulaz omogućuje algoritmu da nauči funkciju mapiranja od ulaza do izlaza. Sukladno tome, poznati su i ulazni i izlazni podaci (Athanasopoulou i dr., 2022). Uobičajene primjene nadziranog učenja uključuju zadatke klasifikacije (npr. određivanje je li e-pošta spam ili ne) i zadatke regresije (npr. predviđanje cijena kuća na temelju različitih značajki). Neki od najpopularnijih algoritama koji su dokazani brojnim primjenama su generalizirani aditivni modeli, slučajne šume, boosting, klasifikacijska i regresijska stabla, potporni vektorski strojevi, proširena linearna regresija, logistička regresija, k-najbliži susjedi, linearna diskriminantna analiza, laso, neuronske mreže, adaptivni neuro-fuzzy sustav zaključivanja itd. (Rostami-Tabar i Mircetic, 2023). Nadzirano učenje je moćno jer iskorištava podatke koje su komentirali ljudi za postizanje visoke točnosti u predviđanjima. Međutim, njegova učinkovitost uvelike ovisi o kvaliteti i količini označenih podataka.

Nasuprot tome, nenadzirano učenje bavi se skupovima podataka koji nemaju označene odgovore. Sukladno tome, algoritmi nenadziranog strojnog učenja koriste neoznačene skupove podataka koji uključuju samo ulaze (Athanasopoulou i dr., 2022). Ovdje se algoritam daje samo s ulaznim podacima, a cilj mu je pronaći temeljne obrasce, strukture ili odnose unutar podataka. Uobičajene tehnike učenja bez nadzora uključuju grupiranje (npr. grupiranje kupaca prema kupovnom ponašanju) i smanjenje dimenzionalnosti (npr. smanjenje broja varijabli u skupu podataka uz zadržavanje važnih informacija). Nenadzirano učenje dragocjeno je za istraživačku analizu podataka i otkrivanje skrivenih struktura u podacima. Često se koristi kada su označeni podaci rijetki ili nedostupni.

Još jedna važna kategorija, iako se razlikuje od nadziranog i nenadziranog učenja, jest podržano učenje (engl. *reinforcement learning*). Ovdje algoritam uči interakcijom s okolinom i primanjem povratnih informacija u obliku nagrada ili kazni. Ovaj pristup pokušaja i pogreške pomaže algoritmu da nauči optimalne radnje kako bi maksimizirao kumulativne nagrade. Podržano učenje široko se koristi u područjima kao što su robotika, igranje igrica i autonomni sustavi.

Jedan od najpopularnijih i najuspješnijih algoritama za ML dolazi iz grane neuronskih mreža. Neuronske mreže postoje od 1950-ih, ali su svoju popularnost stekle 1980-ih i u posljednjih 12 godina. Izgrađeni su na aproksimaciji bioloških neurona i načina na koji dijele dijelove informacija u mozgu, ali osim toga, nema značajnih veza između njih dvoje. Danas se najčešće koristi oblik neuronskih mreža u obliku dubokog učenja, koji predstavlja nekoliko skrivenih slojeva između ulaznih i izlaznih značajki, koji izvode nekoliko nelinearnih transformacija ulaznih značajki. Budući da se ovo pokazalo vrlo uspješnim, duboko učenje danas je jedno od najistaknutijih potpodručja strojnog učenja (Chollet, 2021).

10.4 Studija slučaja?

Nalazi Wenzela i dr. (2019) o ML-u u upravljanju opskrbnim lancem ukazuju na rastuću integraciju ML aplikacija u različitim zadacima opskrbnog lanca. Sukladno tome, promatrana studija slučaja predstavlja primjenu AI-a i ML-a, izvedenu u centralnom skladištu tvornice hrane (Mirčetić i dr., 2016; Mirčetić i dr., 2014). U krugu tvornice nalazi se 30 viličara. Viličari su angažirani na različitim poslovima unutar kompleksa koji su ključni za logističke poslove u proizvodnji, skladištenju i otpremi proizvoda. Centralno skladište ima kapacitet od 11.100

paletnih mjesta i godišnju proizvodnju od 300.000 do 350.000 paleta. Trenutno tvornica izravnom dostavom opskrbljuje oko 20.000 prodavaonica.

Problem angažmana viličara povezan je s činjenicom da preveliki ili premali angažman viličara u različitim tvorničkim procesima dovodi do značajnih financijskih i tržišnih gubitaka. Trenutno se proces donošenja odluka gdje će i što će koji viličar raditi temelji na odlukama stručnjaka (menadžera). Stručne odluke temelje se na njihovom iskustvu, bez pomoći bilo kakvog sustava za podršku odlučivanju (DSS). Brojni empirijski dokazi sugeriraju da ljudska intuitivna prosudba i donošenje odluka često nisu optimalni, osobito u uvjetima složenosti i stresa (Druzdzel i Flynn, 2002). Ovo naglašava važnost uključivanja sustava za podršku odlučivanju (DSS) koji pomažu stručnjacima u procesu donošenja odluka.

U ovoj aplikaciji odabrali smo nekoliko ML algoritama za pomoć u optimizaciji operacija utovarnog skladišta. ML algoritmi sastavljeni su u jedinstveni okvir za donošenje odluka koji služi kao DSS za menadžere i stručnjake u određenoj tvrtki. Štoviše, cijeli DSS za donošenje odluka može se promatrati kao AI platforma, budući da stalno preračunava prijedloge iz nekoliko ML modela (koliko viličara koristiti i koje) i automatski odabire one najbolje, s obzirom na dostavljene unose operatera.

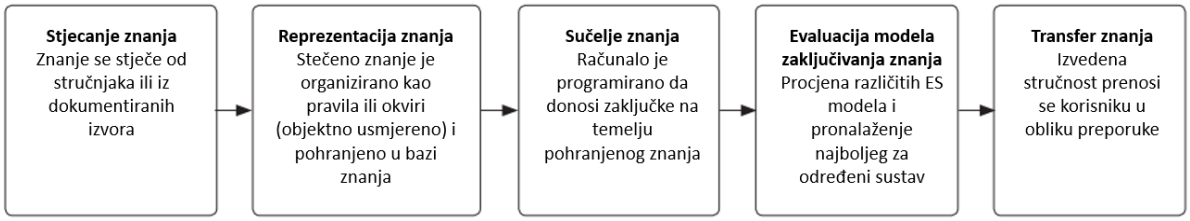
Opis problema

Proces utovara ključan je za skladišnu logistiku i utječe na razinu tržišne usluge. Tijekom otpreme, skladišni stručnjak određuje broj i izbor viličara za utovar, vođen trima čimbenicima: (1) dovršetak utovara unutar navedenog vremenskog okvira, (2) minimiziranje ometanja drugih zadataka viličara i (3) usklađivanje upotrebe viličara s održavanjem mogućnosti, koje mogu podnijeti dva remonta istovremeno. Svaki viličar prolazi četiri do pet remonta godišnje.

Viličari su vitalni za operacije utovara, koje moraju podržati marketinšku strategiju tvrtke, a istodobno osigurati nesmetan rad ostalih aktivnosti. Pogrešna raspodjela viličara može dovesti do neiskorištenosti resursa ili naštetiti ugledu tvrtke i razini usluge. Kašnjenja u utovaru povlače kazne. Upravitelj mora koordinirati korištenje viličara u svim aktivnostima kako bi izbjegao istovremene remonte i upravljao različitim potrebama održavanja. Iako menadžeri obično donose točne odluke, okruženja visokog stresa mogu dovesti do pogrešaka. Stoga je DSS potreban za povećanje povjerenja i pouzdanosti donošenja odluka.

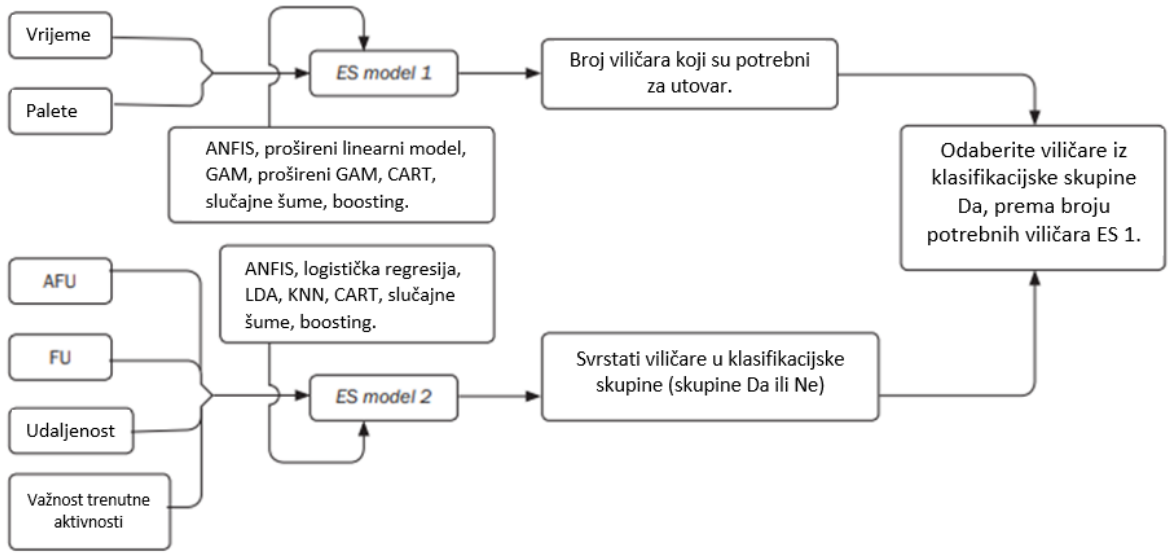
AI i ML kao DSS za centralno skladište

Prvi korak za generiranje AI i ML sustava je stjecanje stabilnog i ispravnog izvora znanja (baza podataka) i rasvjetljavanje poslovnih uloga koje treba podržavati. Stoga slika 10.3 predstavlja metodologiju za izgradnju AI i ML DSS sustava.



Slika 4.60Metodološki koraci za izgradnju AI i ML DSS sustava (Rainer i Turban, 2008; Turban i dr., 2005).

Stjecanje znanja ostvareno je razgovorima s menadžerima, promatranjem njihovih procesa donošenja odluka i pregledom skladišne dokumentacije. Za razvoj DSS-a za navedeni problem uspostavljene su dvije baze znanja. Prva baza znanja uključuje odluke o broju viličara raspoređenih u zoni utovara (434 stručne odluke), dok druga pokriva koji su viličari korišteni (368 stručnih odluka) u različitim operativnim scenarijima. Tijekom faze zaključivanja znanja primijenjeno je nekoliko ML algoritama pomoću Matlab softvera: Adaptivni neuro-fuzzy sustav zaključivanja (ANFIS), generalizirani aditivni modeli (GAM), slučajne šume, boosting, klasifikacijska i regresijska stabla (CART), proširena linearna regresija, logistička regresija, k-najbliži susjedi (KNN) i linearna diskriminantna analiza (LDA). Procijenjeni su različiti ML modeli i identificirani su oni s najboljim performansama. ANFIS i CART pokazali su vrhunske rezultate i odabrani su kao konačni DSS-ovi za praktičnu primjenu u tvrtki. Prijenos znanja je omogućen kroz korisničko sučelje finalnih DSS modela. Struktura i logika DSS-a ilustrirana je na slici 10.4.



Slika 4.61Izgradnja strukture skladišnog DSS-a na temelju AI i ML algoritama.

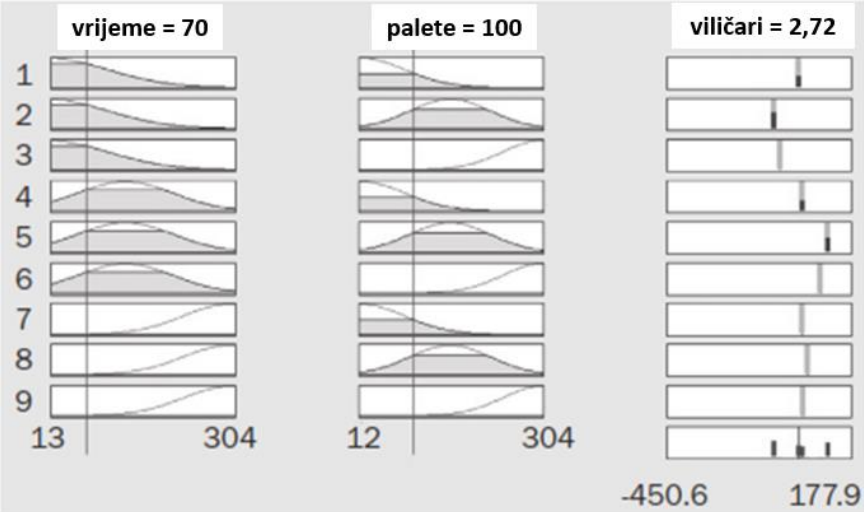
DSS okvir sastoji se od ulaznog sloja koji se sastoji od nekoliko ključnih čimbenika koji utječu na uključivanje viličara. ML sloj ima ML modele koji ponovno izračunavaju prijedlog o tome koliko i koje viličare koristiti u danom scenariju unosa. Modeli s najboljim učinkom biraju se kao modeli ekspertnih sustava (ES modeli) budući da se baza znanja na kojoj se stvaraju ML modeli izdvaja od stručnjaka. Prvi model fokusiran je na određivanje broja viličara potrebnih u zoni utovara (ES model 1). Drugi model bavi se problemom odabira koji će se specifični viličari angažirati (ES model 2). Oba modela razvijena su korištenjem nadziranih tehnika strojnog učenja. Prema Turbanu i dr. (2005.), strojno učenje je pokazalo izvrsne rezultate u dizajniranju inteligentnih sustava za podršku odlučivanju (DSS). ES modeli šalju signale (ML sugestije i prijedloge) dalje u operaciju sortiranja, gdje svaki viličar koji je klasificiran u grupi sortiranja „Da", može biti angažiran u određenoj operaciji utovara.

Konzultacijama su identificirani čimbenici koji utječu na odluke menadžera. Za određivanje broja viličara za raspoređivanje u zoni utovara, ključni čimbenici su navedeno vrijeme utovara (15 do 135 minuta) i količina tereta (15 do 225 paleta). Prilikom odabira viličara koje će angažirati, upravitelj uzima u obzir važnost trenutne aktivnosti (ocjenjuje se od 1 do 9 prema politici tvrtke), stopu iskorištenja viličara, njegovu udaljenost od pristaništa za utovar i prosječnu stopu iskorištenja svih viličara. Svaki viličar ima određeni broj radnih sati prije nego što je potreban remont, a njegova je uporaba ograničena izvan tog ograničenja. Iskorištenje

viličara (engl. *Forklift Utilization* - FU) je postotak radnih sati koje koristi pojedinačni viličar, dok je prosječna iskorištenost viličara (engl. *Average Forklift Utilization* - AFU) prosječno radno vrijeme svih viličara. Veći AFU sugerira da će većina viličara uskoro trebati remont.

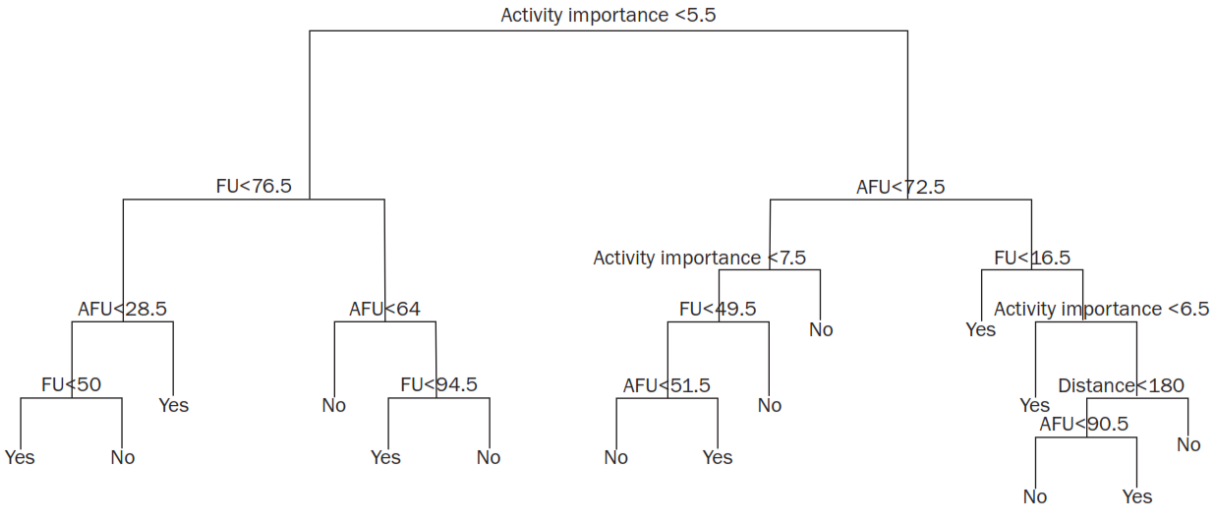
Korisničko sučelje DSS-a

U većini situacija unosa najbolja izvedba je pokazana putem ANFIS-a i CART-a. Sukladno tome, oni su odabrani kao pokretači danog DSS-a i njegovih ES-ova. Korisničko sučelje ES modela 1 prikazano je na slici 10.5, a operaterima omogućuje brzo i jednostavno donošenje odluka o broju viličara za postavljanje jednostavnim pomicanjem okomite linije kroz domenu ulaznih varijabli, na temelju navedenog vremena utovara i količine tereta.



Slika 4.62Sustav neizrazitog zaključivanja ES modela 1.

ES Model 2 služi kao dopunski alat ES Modelu 1, poboljšavajući donošenje odluka dajući informacije o tome treba li određeni viličar biti postavljen u zoni utovara (slika 10.6). Uzimajući u obzir položaj viličara (udaljenost od zone utovara), njegovu trenutnu aktivnost (važnost aktivnosti), njegovu iskorištenost radnih sati (FU) i prosječnu iskorištenost svih viličara (AFU), korisnici mogu lako utvrditi je li određeni viličar prikladan za utovar ili treba odabrati neki drugi. CART stablo odlučivanja je jednostavno za tumačenje, eliminirajući potrebu za unosom vrijednosti u softver. Umjesto toga, stablo sa slike 10.6 može se ispisati i istaknuti u skladištu za brzu referencu.



Slika 4.63Stablo odlučivanja ES modela 2 u vezi s uključivanjem viličara.

Menadžeri mogu svakodnevno koristiti predstavljeni DSS, pomažući u postizanju boljeg odgovora opskrbnog lanca na zahtjeve kupaca i osiguravajući visoku vjerojatnost isporuke na vrijeme. Predloženi AI & ML DSS pokazao je uspješne rezultate u stjecanju stručnjakovog "know-how" znanja i hvatanju njihove "logike zaključivanja". Korištenjem ove metode, stručnost menadžera može se izdvojiti i primijeniti na druge skladišne operacije. Ovo je osobito vrijedno za praktičare jer je angažiranje stručnjaka za skladištenje često skupo. Osim toga, DSS također može poslužiti kao alat za obuku menadžera početnika, pomažući im da steknu iskustvo i poboljšaju svoje vještine donošenja odluka tijekom vremena. Stoga su AI i ML sustavi koji mogu simulirati odluke menadžera ključni alati koji nude značajne uštede troškova i povećanu učinkovitost u skladišnim operacijama.

Literatura 10. poglavlja

- Athanasopoulou, K., Daneva, G. N., Adamopoulos, P. G., & Scorilas, A. (2022). Artificial intelligence: The milestone in modern biomedical research. BioMedInformatics, 2(4), 727-744. https://doi.org/10.3390/biomedinformatics2040049
- Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. Communications of the ACM, 64(7), 58-65.
- Chollet, F. (2021). Deep learning with Python. Simon and Schuster.
- Druzdzel, M. J., & Flynn, R. R. (2002). Decision support systems. In A. Kent (Ed.), Encyclopedia of library and information science (2nd ed.). Marcel Dekker, Inc.
- Gibbs, M. (2019, December 17). Table-driven programming and the weather forecasting stone. Global Nerdy. https://www.globalnerdy.com/2019/12/17/table-driven-programming-and-the-weather-forecasting-stone/

- Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural computation, 18(7), 1527-1554.
- Jackson, P. (1999). Introduction to expert systems. Addison-Wesley.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
- Mirčetić, D., Ralević, N., Nikoličić, S., Maslarić, M., & Stojanović, Đ. (2016). Expert system models for forecasting forklifts engagement in a warehouse loading operation: A case study. Promet-Traffic & Transportation, 28(4), 393-401.
- Mircetic, D., Lalwani, C., Lirn, T., Maslaric, M., & Nikolicic, S. (2014, July). ANFIS expert system for cargo loading as part of decision support system in warehouse. In 19th International Symposium on Logistics (ISL 2014).
- Rostami-Tabar, B., & Mircetic, D. (2023). Exploring the association between time series features and forecasting by temporal aggregation using machine learning. Neurocomputing, 548, 126376.
- Olson, D. L., & Courtney, J. F. (1992). Decision support models and expert systems. Macmillan.
- Rainer, R. K., & Turban, E. (2008). Introduction to information systems: Supporting and transforming business. John Wiley & Sons.
- Turban, E. (1998). Decision support and expert systems (2nd ed.). Macmillan.
- Turban, E., Aronson, J., & Liang, T.-P. (2005). Decision support systems and intelligent systems (7th ed.). Pearson Prentice Hall.
- Wenzel, H., Smit, D., & Sardesai, S. (2019). A literature review on machine learning in supply chain management. In W. Kersten, T. Blecker, & C. M. Ringle (Eds.), Artificial Intelligence and Digital Transformation in Supply Chain Management: Innovative Approaches for Supply Chains (Vol. 27, pp. 413-441). https://doi.org/10.15480/882.2478

POPIS SLIKA

POPIS TABLICA

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -