# STATISTIČKE METODE ZA ANALIZU LOGISTIČKIH PODATAKA

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

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

*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*
*Prof. dr.sc. Svetlana Nikoličić*
*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*
*Doc. dr. Dejan Mirčetić*
*Doc. dr. sc. Jelena Franjković*



*Tiskano izdanje*

*Online izdanje*

*ISBN: ……*

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:

*Predgovor*

Područje logistike sve se više oslanja na uvide temeljene na podacima kako bi se optimiziralo poslovanje, smanjili troškovi i osigurala učinkovitost u lancima snabdevanja. 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 lancem snabdevanja. 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 primenom. Deset poglavlja uključivalo je detaljno istraživanje ključnih tema kao što su predviđanje potražnje, simulacijsko modeliranje, regresiona analiza i integracija veštačke inteligencije i strojnog učenja u logističke operacije. Korišćenjem 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). Sprovedeno je nekoliko preliminarnih istraživanja kako bi se odredio sadržaj ovog udžbenika. Prvo je sprovedeno opsežno istraživanje kako bi se ispitali postojeći predmeti poslovne analitike, njihov sadržaj i veštine koje prenose studentima logistike širom Evropske unije, Sjedinjenih Američkih Država i Ujedinjenog Kraljevstva. Ova je analiza otkrila jaz između logističkog znanja i statističkih veština potrebnih na terenu te onih koji se trenutno nude studentima. Na temelju dubinskih intervjua sa univerzitetskim nastavnim osobljem, studentima i stručnjacima iz industrije, više od 100 veština poslovne analitike identifikovano je kao neophodno. Koristeći ABC metodu klasifikacije rangiranja, 33 veštine odabrane su za uključivanje u ovu knjigu, prvenstveno usmerene na matematiku, informatiku, menadžment, primenjenu matematiku i statistiku. Kombinacija ovih utvrđenih potreba i veština dovela je do razvoja deset poglavlja sadržaja koja se bave najkritičnijim veštinama potrebnih u tom području.

Prvo poglavlje pokriva *Uvodnu statistiku* i pruža sveobuhvatan pregled statističkih koncepata i njihovih primena, posebno unutar analize lanca snabdevanja. Započinje naglašavanjem ključne uloge statistike u optimizaciji lanaca snabdevanja. Koristi deskriptivnu statistiku kao što je srednja vrednost, medijana i standardna devijacija za analizu vremena isporuke, nivoa zaliha i troškova. Poglavlje predstavlja prediktivne tehnike poput regresije i analize vremenskih serija za prognoziranje potražnje i zaliha. Dalje istražuje važnost varijabli, razlikujući kvalitativne i kvantitativne tipove, te zadire u temeljne statističke mere kao što su srednja vrednost, medijana, mod, varijansa 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 verovatnoća, nudeći alate za razumijevanje odnosa između varijabli i modeliranja slučajnih pojava u podacima. Ove statističke tehnike pomažu da se poboljša donošenje odluka u lancu snabdevanja, učinkovitost i upravljanje rizikom.

Drugo poglavlje, *Statistika za poslovnu analitiku*, istražuje bitne statističke koncepte i tehnike za dobijanje 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 merenju varijabilnosti podataka. Takođe pokriva sampling-distribuciju i centralnu graničnu teoremu, objašnjavajući kako one 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 proračunima verovatnoće. Poglavlje završava raspravom o standardnoj grešci i intervalima pouzdanosti, koji pomažu kvantifikaciju nesigurnosti koja okružuje procene. U konačnici, ovo poglavlje čitataocima daje statističke alate potrebne za poslovnu analitiku, omogućujući im donošenje informisanih odluka vođenih podacima.

Poglavlje o *Upravljanju podacima* istražuje različite aspekte upravljanja podacima u logistici, fokusirajući se na formate podataka, organizaciju i tehnologije. Počinje s ulogom elektronske razmene podataka (engl. *Electronic Data Interchange* - EDI) u razmeni informacija unutar lanca snabdevanja korišćenjem standardizovanih alfanumeričkih formata. Objašnjava koncept informacija, podataka i znanja, raspravljajući o tome kako se podaci digitalizuju i organizuju u baze podataka, skladišta i baze znanja. Poglavlje se bavi logističkim podacima, posebno upotrebom linijskih kodova (bar kodova) i RFID oznaka za identifikaciju i praćenje u logistici. Takođe predstavlja tehnike organizacije podataka, u rasponu od proračunskih tabela do relacionih baza podataka, objašnjava ključne koncepte kao što su primarni i strani ključevi, normalizacija i programski jezici za upravljanje podacima kao što je SQL. Osim toga, u poglavlju se raspravlja o najboljim praksama za filtriranje podataka i sprečavanje grešaka tokom 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 *Simulacionom modeliranju i analizi* (engl. *Simulation Modelling and Analysis* - SMA) fokusirano je na stvaranje digitalnih modela za simulaciju sistema u stvarnom svetu za optimizaciju i donošenje odluka. Započinje objašnjenjem Conant-Ashby teoreme, koja sugeriše da simulacioni model mora odgovarati složenosti svog pandana iz stvarnog sveta kako bi se učinkovito regulisao. SMA optimizira lance snabdevanja i prometne mreže u logistici, omogućujući menadžerima da simuliraju i procene različite scenarije. Poglavlje opisuje ključne metodologije simulacije, kao što je simulacija diskretnih događaja (engl.

*Discrete Event Simulation* - DES) za analizu usmerenu na proces, sistemska dinamika (engl. *System Dynamics* - SD) za performanse sistema visokog nivoa, simulacija bazirana 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 sistema.

Poglavlje *Linearna regresija s jednim i više regresora* uvodi regresijsku analizu za razumevanje odnosa između zavisnih i nezavisnih varijabli. Započinje jednostavnom linearnom regresijom, gde jedna nezavisna varijabla, poput izdataka za oglašavanje, predviđa ishod kao što je prodaja. U poglavlju se objašnjava konstrukcija regresionog modela i regresione jednačine koja se koristi za predviđanje zavisne varijable na osnovu podataka uzorka. Takođe pokriva metodu najmanjih kvadrata, ključnu tehniku za procenu regresione linije minimiziranjem grešaka predviđanja. Zatim uvodi koeficijent determinacije ($R^2$) za merenje koliko dobro regresioni model odgovara podacima. Zatim se poglavlje bavi višestrukom regresijom, gde dve ili više nezavisnih varijabli predviđaju zavisnu varijablu, nudeći sveobuhvatniju analizu. Primeri uključuju predviđanje vremena putovanja na osnovu udaljenosti i broja dostava.

Poglavlje *Uvod u operaciona istraživanja* fokusirano je na korišćenje analitičkih metoda za poboljšanje donošenja odluka, posebno u logistici i upravljanju lancem snabdevanja. Operaciona istraživanja koriste tehnike modeliranja, statistike i optimizacije za pronalaženje optimalnih rešenja 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 kompanijama donošenje informisanih odluka korišćenjem prognoziranja, prediktivne analitike i vizualizacije podataka. Poglavlje takođe uvodi višekriterijumsko odlučivanje (engl. *multi-criteria decision-making* - MCDM), koje pomaže u ocenjivanju i izboru optimalnih rešenja na osnovu različitih kriterijuma. Konačno, sistemi za podršku odlučivanju (engl. *decision support systems* - DSS) i inženjerstvo bazirano 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. Objašnjeno je kako SPSS omogućava uvoz, manipulaciju i pripremu podataka putem korisničkog interfejsa. Poglavlje pokriva ključne funkcije poput deskriptivne statistike, izrade grafikona i vizualizacije podataka. Takođe predstavlja temeljne statističke testove - T-testove, korelaciju, hi-kvadrat i ANOVA - vodeći čitataoce 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 njenu primenu putem alata kao što su R i SQL za rešavanje poslovnih problema. BA ima za cilj poboljšati donošenje odluka i uspešnost kompanije korišćenjem metoda vođenih podacima. Uključuje deskriptivne, prediktivne i preskriptivne platforme za analizu podataka i donošenje informisanih 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 čuvani u SQL-u mogu analizirati pomoću R skripti za automatizaciju zadataka. Praktični primeri, poput postavljanja upita bazi podataka Chinook, ilustruju kako R i SQL rade zajedno za generisanje uvida, kao što je identifikacija 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 lancu snabdevanja, vizualizaciju i inženjering karakteristika. Prognoziranje potražnje predviđa potrebe kupaca, menja celi lanac snabdevanja i smanjuje logističke troškove. U poglavlju se obrađuju ključni koraci za prognoziranje potražnje, uključujući definisanje problema, prikupljanje podataka, analizu trendova, izbor modela i njihovu procenu. Vizualizacija pomaže u prepoznavanju obrazaca kao što su sezonska kretanja i trendovi, koji mogu uticati na izbor 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. Primer u prehrambenoj industriji pokazuje sposobnost modela S-ARIMA da predvidi potražnju i usmeri donošenje odluka. Konačno, poglavlje pokriva kako testirati i potvrditi modele predviđanja kako bi se osigurala njihova učinkovitost u aplikacijama u stvarnom svetu, koristeći metrike kao što su RMSE i MAPE za procenu učinka.

Posljednje poglavlje istražuje ulogu veštačke inteligencije (engl. *artificial intelligence* - AI) i mašinskog učenja (engl. *machine learning* - ML) u lancima snabdevanja, počevši s pregledom razvoja AI-a od simboličkih sistema do modernih pristupa ML-u. AI se odnosi na automatizaciju zadataka koji obično zahtevaju 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, primenjuju za rešavanje poslovnih problema, uključujući prognoziranje potražnje, upravljanje zalihama i optimizaciju. Bitna studija slučaja uključuje primenu AI i ML algoritama u centralnom skladištu fabrike hrane, optimizirajući korišćenje viljuškara. Uključivanjem sistema za podršku odlučivanju (DSS), AI/ML modeli pomažu menadžerima u izboru optimalnog broja viljuškara, poboljšavajući operativnu učinkovitost. Poglavlje naglašava kako AI/ML može obuhvatiti stručno znanje, smanjiti troškove i poboljšati procese donošenja odluka u upravljanju lancem snabdevanja.

# 1. Uvodna statistika

## 1.1 Uloga i važnost statistike u analizi podataka u lancima snabdevanja

Statistika igra ključnu ulogu u modernim lancima snabdevanja, gde su učinkovito upravljanje, planiranje i kontrola ključni. Statističke metode koriste se za prikupljanje, analizu i tumačenje podataka, omogućujući kompanijama da bolje razumeju i optimiziraju svoje lance snabdevanja.

Istaknimo neke od važnih uloga statistike u analizi lanca snabdevanja.

Deskriptivna statistika ključna je za opisivanje osnovnih svojstava podataka o lancu snabdevanja, kao što su srednja vrednost, standardna devijacija, medijana, kvartili i druge mere. Ovi alati nam pomažu razumeti distribuciju i karakteristike podataka kao što su prosečna vremena isporuke, količine na zalihama i prosečni troškovi, što doprinosi boljem razumevanju i upravljanju lancem snabdevanja.

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

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

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

Osim toga, statistika se takođe koristi za procenu rizika lanca snabdevanja, kao što je verovatnost kašnjenja u isporukama, oštećenja tokom transporta i drugih potencijalnih problema.

Statističkim praćenjem i kontrolom procesa identifikujemo odstupanja od standarda, što nam omogućuje poboljšanje kvaliteta i učinkovitosti procesa lanca snabdevanja.

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

Konačno, statistika je ključni alat za donošenje utemeljenih odluka o nabavci, zalihama, izboru dobavljača i drugim aspektima upravljanja snabdevanjem, doprinoseći učinkovitom i delotvornom radu celog lanca snabdevanja.

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

## 1.2 Osnovni pojmovi statistike

### Varijable

Varijable su osnovni gradivni blokovi u statistici jer predstavljaju svojstva ili karakteristike koje se mere ili posmatraju u anketi, eksperimentu ili uzorku podataka. Varijable su ključne za razumevanje i analizu podataka jer omogućuju istraživačima, analitičarima i statističarima da opisuju, analiziraju i razumeju fenomene.

Važno je razumeti 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 klasifikovati prema matematičkom redu. Primeri uključuju pol (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 vrednosti koje se mogu prebrojati ili izmeriti i mogu se sortirati po nekom matematičkom redu. Primeri uključuju starost, 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, meriti ili predvideti, dok je nezavisna varijabla ona koja treba uticati na zavisnu varijablu. Na primer, ako želimo istražiti da li nivo obrazovanja utiče na dohodak, dohodak je zavisna varijabla, a nivo obrazovanja nezavisna varijabla.

**Diskretne i kontinuirane varijable.** Varijable se takođe mogu podeliti na diskretne i kontinuirane. Diskretne varijable imaju ograničen skup mogućih vrednosti i obično su predstavljene celim brojevima. Primer je broj dece u porodici, gde su moguće vrednosti 0, 1, 2 itd. Kontinuirane varijable, s druge strane, imaju beskonačan broj mogućih vrednosti i obično se mere pomoću decimalnih brojeva. Primer je visina osoba, gde je moguć beskonačan broj vrednosti unutar zadatog raspona.

Varijable su osnovni alati za istraživanje i analizu podataka. Razumevanje i pravilno definisanje varijabli ključno je za sprovođenje statističkih analiza i proučavanje fenomena u istraživanju. Varijable omogućuju istraživačima izražavanje i kvantifikaciju različitih aspekata stvarnosti, omogućujući bolje razumevanje fenomena, donošenje odluka i predviđanje budućih događaja. Takođe omogućuju korišćenje različitih statističkih tehnika za testiranje hipoteza, predviđanja i bolje razumevanja uzročno-posledičnih veza između varijabli.

## 1.3 Osnovni statistički koncepti s primerima

### Prosek (srednja vrednost)

Srednja **vrednost** (engl. *mean*), takođe poznata kao **prosek**, jedna je od osnovnih statističkih mera. Srednja vrednost je aritmetički prosek svih vrednosti u skupu podataka. Izračunava se sabiranjem svih podataka, a zatim deljenjem s brojem podataka.

Izračunavanje proseka:

- Sabrati sve vrednosti u skupu podataka.
- Podelite zbir s brojem vrednosti u skupu.
- Jednačinaa za izračunavanje proseka $(\bar{x})$ je: $\bar{x} = (x_1 + x_2 + x_3 + \ldots + x_n) / n$

Gde je $\bar{x}$ prosek. $x_1, x_2, x_3, \ldots, x_n$ su vrednosti u skupu podataka. $n$ je broj vrednosti u skupu podataka.

Primer:

Zamislite skup podataka koji predstavlja ocene učenika na ispitu iz matematike: 80, 85, 90, 75, 95. Da biste izračunali prosek, saberite sve ove vrednosti i podelite s brojem ocena, koji je u ovom slučaju 5:

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

Dakle, prosečni studentski rezultat je 85. Prosek je koristan za merenje centralne tendencije podataka i daje nam grubu ideju o tome šta možemo očekivati kao "tipičnu" vrednost u skupu podataka. Međutim, srednja vrednost može se značajno promeniti ako su u podacima prisutni ekstremi. Stoga je važno poznavati druge statističke mere kao što su medijana i mod kako bi se bolje razumela distribucija podataka.

## Medijana

Medijana je statistički koncept koji se koristi za merenje položaja srednje vrednosti skupa podataka. To je vrednost koja deli uređene podatke na dve jednake polovine. To znači da polovina podataka ima vrednosti manje ili jednake medijani, a druga polovina ima vrednosti veće ili jednake medijani. Medijana je jedna od osnovnih mera centralne tendencije u statistici i koristi se za opisivanje distribucije podataka, posebno kada su podaci iskrivljeni ili sadrže ekstremne vrednosti.

Kako izračunati medijanu?

- Prvo morate sortirati skup podataka od najmanje do najveće vrednosti.
- Ako je broj podataka paran $(n)$, tada je medijana prosek dve srednje vrednosti. To znači da je medijana jednaka proseku vrednosti na poziciji $n/2$ i $(n/2 + 1)$ kada su podaci poređani uzlaznim redosledom.
- Ako je broj podataka neparan, tada je vrednost medijane na sredini.

Primer:

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

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

Budući da je broj podataka neparan (7), medijana će biti vrednost na srednjoj poziciji, što je četvrta vrednost u poređanom skupu podataka. Dakle, medijana je u ovom slučaju jednaka 7 sati. To znači da polovina ljudi u ovom skupu podataka spava 7 ili manje sati, dok druga polovina spava 7 ili više sati.

### Mod

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

Modus je koristan za identifikovanje najčešće vrednosti u skupu podataka i posebno je koristan pri analizi kvalitativnih (kategorijskih) varijabli gde su vrednosti nenumeričke.

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

Primer: zamislite skup podataka koji predstavlja boje automobila na parkingu:

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

U ovom slučaju mod je "Plava", jer se ova vrednost najčešće pojavljuje (četiri puta), dok se "Plava" i "Zelena" pojavljuju ređe.

Mod je jednostavan za izračunavanje jer jednostavno identifikuje vrednost s najvećom učestalošću pojavljivanja u skupu podataka. Mod se koristi za opisivanje karakterističnih vrednosti u podacima i može biti koristan u razumevanju koja je vrednost najkarakterističnija za određenu situaciju ili grupu.

### Raspon varijacije

Razlika između maksimalne i minimalne vrednosti u skupu podataka je statistički koncept koji se naziva raspon. Time se meri kolika je razlika između maksimalne i minimalne vrednosti u skupu podataka. Raspon je jednostavan način za procenu raspona vrednosti u skupu podataka i merenje varijabilnosti između minimalnih i maksimalnih vrednosti.

Izračunavanje raspona varijacije je jednostavno:

- Najpre pronađite minimalnu vrednost (min) i maksimalnu vrednost (max) u skupu podataka.
- Zatim izračunajte razliku između maksimalne i minimalne vrednosti (max - min).

Primer: zamislite skup podataka koji predstavlja starost učesnika događaja: 20, 25, 30, 35, 40. Da biste izračunali raspon varijacije, prvo pronađite minimalnu vrednost (20) i maksimalnu vrednost (40) u skupu podataka. Zatim izračunajte razliku između maksimalne i minimalne vrednosti: VR = 40 - 20 = 20

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

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

### Varijansa i standardna devijacija

**Varijansa** je prosečni zbir kvadrata odstupanja od srednje vrednosti. To je kvadrat standardne devijacije. **Standardna devijacija** je statistička mera koja se koristi za merenje disperzije ili varijabilnosti u skupu podataka. Govori koliko su vrednosti udaljene od srednje vrednosti (proseka) u skupu. Standardna devijacija jedna je od najčešće korišćenih mera disperzije u statistici i utvrđuje se izračunavanjem kvadratnog korena varijance.

Izračunavanje standardne devijacije:

- Prvo izračunajte varijansu. Varijansa se izračunava uzimanjem proseka svih vrednosti u skupu za svaku vrednost u skupu, zatim kvadriranjem i sabiranjem tih razlika.
- Kada dobijete vrednost varijanse ($\sigma^2$), izračunajte standardnu devijaciju izračunavanjem kvadratnog korena varijanse. To se radi vađenjem kvadratnog korena iz $\sigma^2$:

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

Standardna devijacija meri koliko su vrednosti disperzovane oko srednje vrednosti u skupu podataka. Veća vrednost standardne devijacije znači da su vrednosti raširenije i više se razlikuju od srednje vrednosti, dok niža vrednost standardne devijacije označava manju raspršenost.

Primer: zamislite skup podataka koji predstavlja ocene učenika na ispitu iz matematike: 80, 85, 90, 75, 95. Formula koja će biti prikazana u nastavku važi samo ako pet vrednosti s kojima smo započeli čine celokupnu populaciju. Prvo izračunavate prosek (srednju vrednost), koji iznosi 85. Zatim izračunavate varijansu, koja iznosi 50.

Prvo izračunajte odstupanja svakog podatka od srednje vrednosti i kvadrirajte rezultat svake:

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

Varijansa je srednja vrednost ovih vrednosti:

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

Na kraju, izračunavate standardnu devijaciju uzimajući kvadratni korijen varijanse:

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

Dakle, standardna devijacija je u ovom slučaju oko 7,07. To znači da su u proseku rezultati učenika udaljeni oko 7,07 jedinica od proseka. Standardna devijacija se često koristi u analizi distribucije podataka i u proceni varijabilnosti vrednosti u skupu.

### Kvartili

Kvartili su vrednosti koje dele uređene podatke u određene delove. Na primer, kvartili dele podatke na četiri jednaka dela. Prvi kvartil (Q1) deli donjih 25% podataka, drugi kvartil (Q2) jednak je medijani, a treći kvartil (Q3) deli gornjih 25% podataka.

Primer: 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 korišćenje različitih metoda i alata, s ciljem da se podaci prezentuju na jasan, transparentan i informativan način.

Evo nekoliko uobičajenih načina za prikaz statistike:

### Tabele

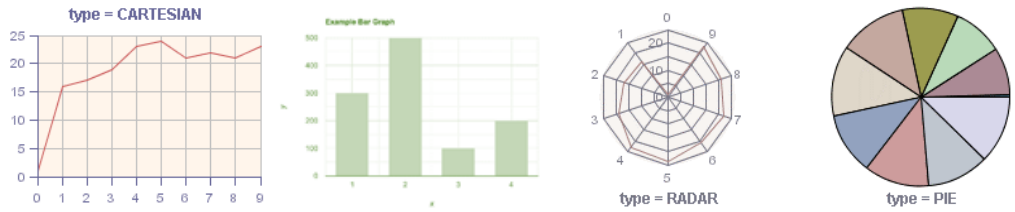
Tabele su osnovna metoda za prikaz podataka. Primeri uključuju tabele frekvencija, koje pokazuju broj pojavljivanja za različite vrednosti, i tabele podataka koje prikazuju više informacija o podacima.

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

**Slika 1.1 Primer tabele.**

### Grafički prikazi

Grafički prikazi su koristan alat za vizualizaciju podataka. Uključuju različite vrste grafikona kao što su stubičasti grafikoni, linijski grafikoni, pita grafikoni, histogrami, kutijasti dijagram itd.




**Slika 1.2 Primeri grafičkih prikaza podataka.**

**Linijski grafikoni** koriste se za vizualizaciju trendova i promena tokom 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 finansije, nauka i poslovanje za analizu vremenskih serija podataka, poređenje trendova u kategorijama ili predviđanje budućeg razvoja na osnovu istorijskih podataka.

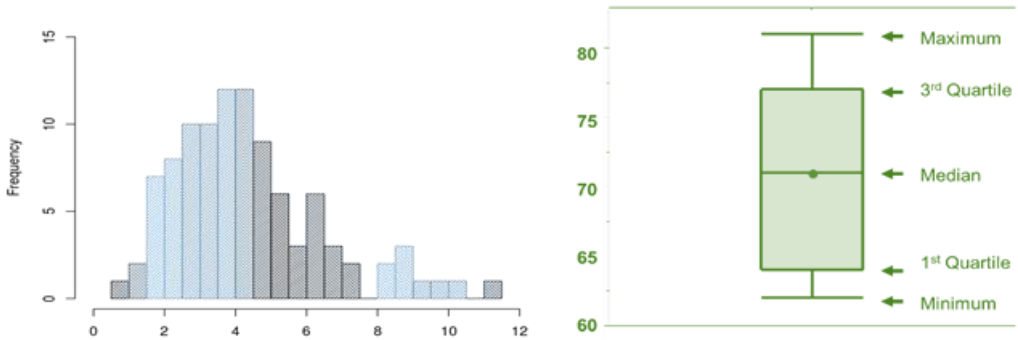
**Stubičasti grafikoni** koriste se za poređenje količina u različitim kategorijama, što ih čini idealnim za predstavljanje diskretnih podataka. Posebno su delotvorni za isticanje razlika, sličnosti i trendova među grupama. Stubičasti grafikoni obično se koriste kada treba prikazati frekvencije, procente ili druge numeričke mere na jasan i vizualno jednostavan način. Široko se primenjuju u poslovanju, obrazovanju i istraživanju za analizu i iznošenje kategoričkih podataka.

**Polarni grafikoni**, takođe poznati kao radarski ili paukovi grafikoni, koriste se za prikaz višestrukih podataka u više dimenzija u kružnom formatu. Idealni su za poređenje nekoliko varijabli ili entiteta prema istim kriterijima, ističući prednosti i slabosti na jasan, vizualan način. Često se koriste u analizi učinka, donošenju odluka i konkurentskim poređenjima, kao što je procena karakteristika proizvoda, timskih veština ili rezultata ankete u različitim kategorijama.

**Pita grafikoni** koriste se za predstavljanje proporcija ili procenata celine, što ih čini idealnim za vizualizaciju relativnih veličina različitih kategorija. Posebno su učinkoviti kada treba pokazati kako delovi doprinose ukupnom iznosu ili na prvi pogled uporediti proporcije. Tortni grafikoni obično se koriste u izveštajima, prezentacijama i istraživanjima za prikaz podataka poput tržišnog učešća, raspodele proračuna ili demografske distribucije.

**Histogrami**

Histogrami su grafički prikazi distribucije podataka. Koriste se za prikaz frekvencija vrednosti varijable u različitim intervalima.




**Slika 1.3 Histogram i kutijasti dijagram (Box-Plot).**

**Kutijasti dijagram (Box plot)**

Kutijasti dijagram, 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, medijana, treći kvartil i maksimum.

Izbor metode za prikaz statistike zavisi od prirode podataka, ciljevima analize i ciljanoj publici. Važno je odabrati metodu koja najbolje odgovara vašoj poruci i čini podatke razumijivijim.

## 1.5 Distribucija frekvencija

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

Proces stvaranja distribucije frekvencija uključuje sledeće korake:

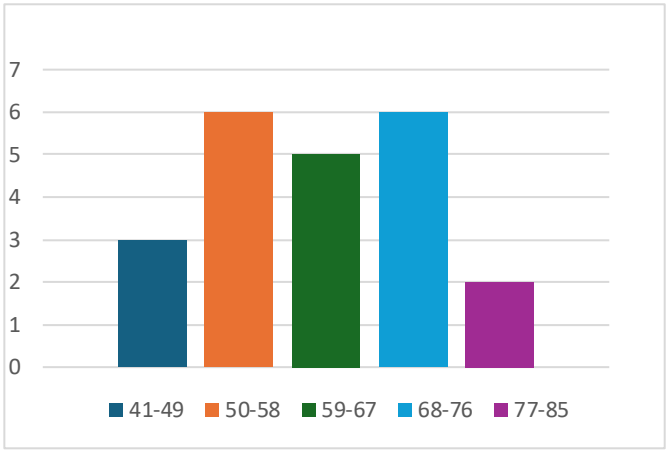
- Prikupljanje podataka: prvo prikupite podatke za koje želite uraditi distribuciju frekvencija.
- Identifikujte različite vrednosti: identifikujte različite vrednosti koje se pojavljuju u vašim podacima. Ovo su kategorije ili diskretne vrednosti koje želite analizirati.
- Brojanje pojavljivanja: izbrojite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

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

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

Slika 1.22 Tabela distribucije frekvencija.

Grafikon distribucije frekvencija (histogram) bi prikazao stupce za svaki raspon ocena s visinom koja predstavlja broj učenika (frekvenciju) u svakoj klasi. Na taj način možemo jasno videti 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 razumevanje podataka, ali ne i za donošenje zaključaka o populaciji kao celini. Glavni cilj deskriptivne statistike je opisati karakteristike podataka, na primer izračunati srednju vrednost, medijanu, raspon, standardnu devijaciju i dati 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 celini iz analize uzorka. Koristi različite statističke metode kao što su testiranje hipoteza, intervali pouzdanosti i regresiona analiza kako bi se razumelo da li se promatrani rezultati uzorka mogu generalizivati na populaciju. Na primer, ako želimo otkriti je li srednja starost u uzorku reprezentativna za populaciju u celini, koristićemo 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 opše zaključke o populaciji ili uzorku iz analize uzorka podataka.

Glavni ciljevi inferencijalne statistike su:

**Procena parametara populacije:** inferencijalna statistika omogućuje nam procenu parametara populacije kao što su srednja vrednost, varijansa, proporcije i druge karakteristike iz uzorka.

**Testiranje hipoteza:** inferencijalna statistika može se koristiti za testiranje hipoteza o populaciji na osnovu uzorkovanih podataka. To uključuje statističko testiranje, gde upoređujemo uzorak s pretpostavkama o populaciji.

**Stvaranje intervala pouzdanosti:** inferencijalna statistika omogućuje izračunavanje intervala koji sadrže procenjene vrednosti parametara populacije s određenom nivom pouzdanosti.

Primer inferencijalne statistike: pretpostavimo da želimo proceniti prosečnu visinu svih studenata na univerzitetu. Budući da je nemoguće proveriti sve učenike, uzimamo uzorak od 100 učenika i merimo njihovu visinu.

Zatim koristimo inferencijalnu statistiku za izračunavanje intervala pouzdanosti za prosečnu 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 grešku srednje vrednosti za izračunavanje intervala

pouzdanosti. Na primer, ako želimo interval pouzdanosti od 95%, koristimo standardnu grešku i kvantile normalne distribucije.

Približan interval pouzdanosti od 95% za prosečnu visinu svih studenata na univerzitetu 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 prosečna 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 univerzitetu iz ukupnog uzorka.

Zajedno ove statističke metode omogućuju logističkim kompanijama bolje razumevanje njihovih procesa, predviđanje budućih događaja i donošenje informisanih 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 vrednosti. Obe metode pomažu da se razume kako jedna varijabla utiče na drugu i koliko se dobro jedna varijabla može koristiti za predviđanje druge. Evo objašnjenja svake od ove dve metode:

### Korelacija

Korelacija se koristi za merenje stepena povezanosti između dve kvantitativne (numeričke) varijable. Ona ukazuje da li postoji linearna veza između dve varijable i koliko je jaka ta veza. Korelacija se meri koeficijentom korelacije, koji uzima **vrednosti između -1** i **1.**

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

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

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

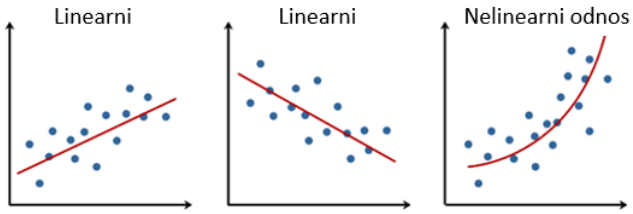
Primer: korelacija između broja sati učenja i ocena koje studenti postižu biće pozitivna ako povećanje broja sati učenja obično odgovara višim ocenama.

### Regresija

Regresija se koristi za modeliranje i predviđanje vrednosti jedne kvantitativne varijable (zavisne varijable) iz vrednosti 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 jednačinom prave linije ($y = a + bx$), gde je a sečište s $y-$osom, a $b$ nagib krive.

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




**Slika 1.24 Grafikon jednostavne linearne regresije.**

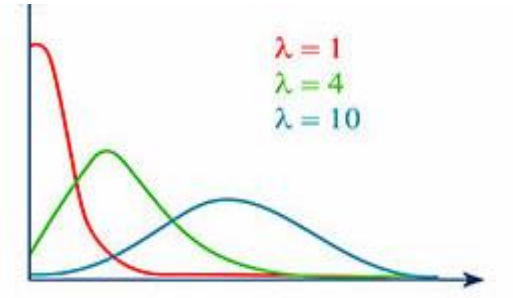
Primer: jednostavna linearna regresija može se koristiti za modeliranje odnosa između broja obavljenih zadataka učenja (nezavisna varijabla) i završne ocene ispita (zavisna varijabla).

## 1.8 Distribucija verovatnoća

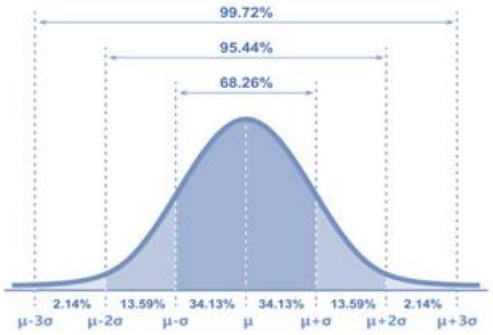
U statistici, distribucija verovatnoća opisuje verovatnost različitih vrednosti koje varijabla može poprimiti. To je matematički model koji pomaže da se

razumeju i analiziraju slučajne pojave i da se predvidi kako će se vrednosti raspodeliti pod određenim okolnostima. Postoji više različitih distribucija verovatnoća, svaka sa svojim karakteristikama i primenama u različitim situacijama. Evo nekih od najpoznatijih distribucija verovatnoća u statistici:

**Normalna (Gaussova) distribucija:** normalna distribucija jedna je od najvažnijih i najčešće korišćenih distribucija. Opisuje simetričnu i zvonastu distribuciju s poznatim parametrima: srednjom vrednošću (μ) i standardnom devijacijom ($\sigma$). 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 uspeha (npr. broja "glava") u određenom broju neovisnih Bernoullijevih eksperimenata. Ima dva parametra: broj pokušaja ($n$) i verovatnost 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 retkih događaja kao što su nesreće, pozivi hitnim službama itd. Parametar distribucije je prosečna stopa događaja ($\lambda$).

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

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

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

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

Ove distribucije verovatnoća osnovni su gradivni blokovi u statistici i koriste se za modeliranje i analizu različitih vrsta podataka u različitim kontekstima. Izbor ispravne distribucije verovatnoće ključan je pri sprovođenju 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 svet poslovne statistike, gde se podaci pretvaraju u značajne uvide, usmeravajući donošenje odluka i otkrivajući skrivene istine. U ovom sveobuhvatnom istraživanju krećemo na putovanje kako bismo demistifikovali bitne statističke koncepte i tehnike koji podržavaju rigoroznu analizu poslovnih podataka. Od razumevanja zamršenosti distribucija do primene testiranja hipoteza i kreiranja intervala pouzdanosti, svako poglavlje otkriva novi aspekt statističke pismenosti.

U srcu statističke analize leži normalna distribucija, kriva u obliku zvona koja prožima bezbrojne pojave u prirodi i ljudskom ponašanju. U ovom delu ulazimo u srž normalne distribucije, razotkrivajući njena svojstva i značaj u statističkom zaključivanju. Kroz vizualizaciju primera iz stvarnog sveta, rasvetljavamo sveprisutnost ove temeljne distribucije i njenu ulogu kao kamena temeljca statističke teorije.

Standardna devijacija služi kao kompas u statističkom ambijentu, vodeći nas kroz varijabilnost svojstvenu skupovima podataka. U ovom poglavlju razlažemo koncept standardne devijacije, otkrivajući njenu važnost u kvantifikovanju disperzije i proceni raspršenosti podataka. Opremljeni dubljim razumijevanjem standardnih odstupanja, kretaćete se podacima s poverenjem, precizno uočavajući uzorke i netipične vrednosti.

Varijable čine gradivne blokove statističke analize, a svaka poseduje različite karakteristike i implikacije. Ovo poglavlje pojašnjava dihotomiju između kontinuiranih i diskretnih varijabli, prikazujući njihovu ulogu u modeliranju i interpretaciji podataka. Shvatanjem nijansi tipova varijabli, iskoristićete pun potencijal statističkih tehnika prilagođenih različitim strukturama podataka.

Sampling-distribucija služi kao osnov statističkog zaključivanja, premošćavajući jaz između promatranja uzorka i parametara populacije. U ovom poglavlju razotkrivamo koncept sampling-distribucije, razjašnjavajući njegovu relevantnost u izradi verovatnosti o karakteristikama populacije. Kroz konkretne primere razvićete intuitivno razumevanje uloge sampling-distribucije uzorkovanja u robusnoj statističkoj analizi.

Centralna granična teorema je ključni koncept u statistici koji nam pomaže da shvatimo nesigurnost. Ovo poglavlje objašnjava centralnu graničnu teoremu na jednostavan način, pokazujući kako proseke uzorka čini predvidljivijima i pomaže u testiranju hipoteza. Razumevanjem ovog koncepta moći ćete izvući smislene zaključke iz podataka.

Razumevanje testiranja hipoteza bitno je za donošenje odluka na osnovu podataka. Omogućuje nam da utvrdimo jesu li uočeni obrasci u podacima smisleni ili su jednostavno slučajni. Primenom testiranja hipoteza možemo proceniti pretpostavke, uporediti grupe i proceniti statistički značaj rezultata, što ga čini vitalnim alatom u naučnom istraživanju, poslovnoj analizi i mnogim drugim područjima.

Z-standardizovana vrednost i z-tabele služe kao navigaciona pomoć u moru standardne normalne distribucije, olakšavajući standardizovana poređenja i izračunavanja verovatnoće. Ovo poglavlje pojašnjava zamršenost z-standardizovanih vrednosti, jačajući vas da tumačite standardizovane rezultate i koristite Z-tabele za statističku analizu. Uz veštinu o z-standardizovanim vrednostima, kretaćete se ogromnim prostranstvom normalne distribucije s poverenjem i preciznošću.

U situacijama kada su veličine uzorka male ili su standardne devijacije populacije nepoznate, t-rezultati i t-tabele pojavljuju se kao nezamenjivi alati za statističku analizu. Ovo poglavlje razotkriva misterije t-rezultata, vodeći vas kroz njihov izračunavanje i tumačenje pomoću t-tabela. Opremljeni ovim znanjem, lako ćete se snalaziti u nijansama t-distribucija, osiguravajući zaključivanje u različitim statističkim scenarijima.

Normalna i t-distribucija predstavljaju stubove teorije verovatnoćei, a svaka poseduje jedinstvene karakteristike i primene. U ovom poglavlju razjašnjavamo razlike između ovih distribucija, omogućavajući vam da shvatite kada svaku od njih da upotrebite u statističkoj analizi. Kroz praktične primere i komparativne analize, razvićete razumevanje normalne i t-distribucije, obogaćujući svoj skup statističkih alata.

Intervali pouzdanosti pružaju uvid u neizvesnost oko parametara populacije, omogućavajući nam da kvantifikujemo preciznost naših procena. U ovom poglavlju istražujemo konstrukciju intervala pouzdanosti za srednje vrednosti i proporcije, razotkrivajući metodologiju i tumačenje ovih bitnih statističkih alata. Savlađivanjem intervala pouzdanosti, transparentno i kritički ćete preneti neizvesnost koja je svojstvena vašim nalazima.

Dok p-vrednosti nude pristup statističkim zaključivanjima, njihovo pogrešno tumačenje može dovesti do pogrešnih zaključaka i pogrešno informisanih odluka. Ovo poglavlje ispituje potencijalne zamke preteranog oslanjanja na p-vrednosti, naglašavajući važnost konteksta i veličine učinka u statističkoj analizi. Kroz kritičko ispitivanje i praktične uvide, pažljivo ćete se kretati kroz složenost p-vrednosti, osiguravajući integritet svojih statističkih zaključaka.

Unutar ovih stranica leže ključevi za otključavanje misterija statističke analize, što vam omogućava da pouzdano i precizno upravljate složenošću podataka. Dok zajedno krećemo na ovo putovanje, neka nam znatiželja bude kompas, a istraživanje naše svetlo vodilja, osvetljavajući put prema dubljem razumevanju i delotvornim zaključcima.

## 2.1 Normalna distribucija

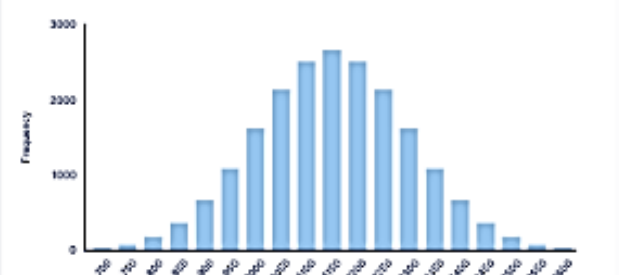
U središtu statističke analize nalazi se normalna distribucija, sveprisutna distribucija verovatnoća koja služi kao merilo za mnoge statističke tehnike. Udubićemo se u njene karakteristike, njenu simetričnu krivu liniju u obliku zvona i značaj u razumevanju distribucije podataka.

Normalna distribucija nalazi primenu u raznim područjima, uključujući finansije, psihologiju, inženjerstvo i biologiju. Od modeliranja cena deonica do razumevanja distribucije visine ljudi, normalna distribucija služi kao svestran alat za analizu i tumačenje podataka.

Kroz ovo poglavlje proučiće se matematička svojstva normalne distribucije, istražujući kako izračunati verovatnoće, percentile i z-središnje vrednosti. Raspravljaćemo o praktičnim tehnikama za vizualizaciju i interpretaciju normalnih distribucija pomoću histograma, dijagrama gustine i funkcija kumulativne distribucije.

Do kraja ovog poglavlja duboko ćete vrednovati normalnu distribuciju i njen značaj u statističkoj analizi. Bićete spremni za rešavanje naprednijih statističkih koncepata i njihovu primenu na skupove podataka u stvarnom svetu. Krenimo na ovo putovanje kako bismo zajedno razotkrili misterije normalne distribucije.

Normalna distribucija, takođe poznata kao Gaussova distribucija ili zvonasta kriva, pokazuje simetričnu distribuciju podataka bez asimetrije. Kada su grafički prikazani, podaci grade krivu liniju u obliku zvona, s većinom vrednosti koje se skupljaju oko središta i smanjuju kako se udaljavaju od njega.

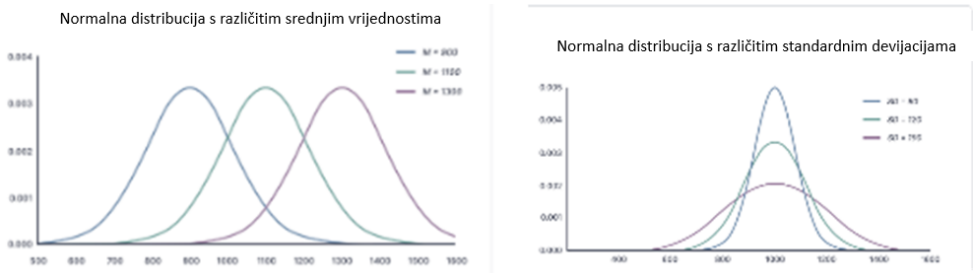


Slika 2.1 Primer Gaussove distribucije ili zvonaste krive

Različite varijable u prirodnim i društvenim naukama obično pokazuju normalnu distribuciju ili su joj blizu. Primeri uključuju visinu, porođajnu težinu, sposobnost čitanja, zadovoljstvo poslom i SAT rezultate. Zbog učestalosti normalno raspodeljenih varijabli, brojni statistički testovi prilagođeni su takvim populacijama. Veština u razumevanju karakteristika normalne distribucije osnažuje pojedince da koriste inferencijalnu statistiku za poređenje grupa i generisanje procena populacije iz uzoraka.

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

- Srednja vrednost, medijana i mod su potpuno isti.
- Distribucija je simetrična u odnosu na srednju vrednost - polovina vrednosti nalazi se ispod, a polovina iznad srednje vrednosti.
- Distribucija se može opisati s dve vrednosti: srednjom vrednošću i standardnom devijacijom.




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

Srednja vrednost služi kao lokacijski parametar koji diktira središte vrha krive. Podešavanje srednje vrednosti pomera krivu u skladu s tim: povećanje pomera krivu udesno, dok smanjenje pomera krivu ulevo. U međuvremenu, standardna devijacija funkcioniše kao parametar razmera, utičući na širenje ili širinu krive.

Standardna devijacija širi ili sužava krivu. Mala standardna devijacija rezultira uskom krivom linijom, dok velika standardna devijacija dovodi do široke krive.

## 2.2 Empirijsko pravilo

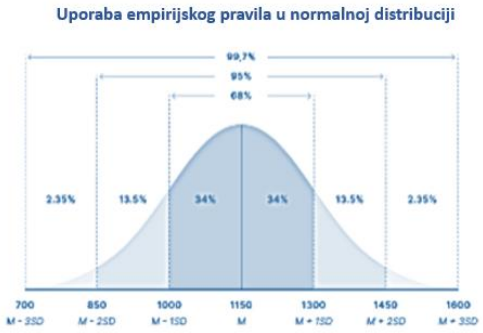
Empirijsko pravilo, takođe poznato kao pravilo 68-95-99.7, daje uvid u raspodelu vrednosti unutar normalne distribucije:

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

Na primer, razmotrite scenario u kojem se prikupljaju rezultati SAT-a od učenika na novom kursu pripreme za ispit, a podaci su u skladu s normalnom distribucijom sa srednjom ocenom (M) od 1150 i standardnom devijacijom (SD) od 150.

Primenom empirijskog pravila može se zaključiti:

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




**Slika 2.54 Empirijsko pravilo u normalnoj distribuciji.**

Empirijsko pravilo

nudi brzu metodu za procenu podataka, omogućavajući otkrivanje outliera ili netipičnih vrednosti 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. Identifikovanje distribucije varijable omogućava primenu relevantnih statističkih testova.

## 2.3 Formula krive linije normalne distribucije

Za konstruiranje krive linije normalne distribucije na osnovu poznate srednje vrednosti i standardne devijacije, može se upotrebiti funkcija gustine verovatnoća, čime se tačno predstavlja distribucija podataka.


**Slika 2.74 Kriva normalne distribucije prilagođena podacima SAT rezultata.**

Unutar funkcije gustine verovatnoća, područje ispod krive linije predstavlja verovatnoću. S obzirom da normalna distribucija služi kao distribucija verovatnoće, kumulativna površina ispod krive linije uvek iznosi 1 ili 100%. Iako se formula za normalnu funkciju gustine verovatnoća može činiti zamršenom, njeno korišćenje zahteva samo poznavanje srednje vrednosti populacije i standardne devijacije. Zamenom ovih parametara u formuli, može se odrediti gustina verovatnoće povezana s bilo kojom datom vrednošću x.

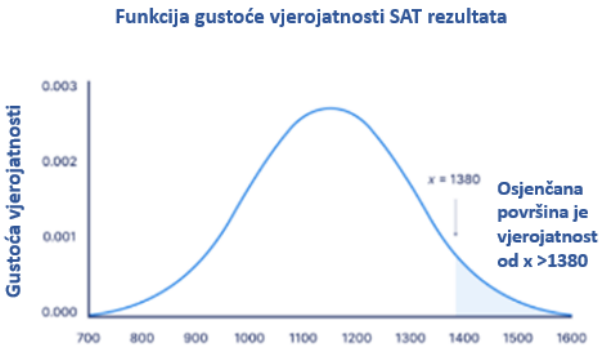
- $f(x)$ = verovatnoća
- $x$ = vrednost varijable
- $\mu$ = srednja vrednost
- $\sigma$ = standardna devijacija
- $\sigma^2$ = varijansa

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

Primer:

Koristeći funkciju gustine verovatnoća, želite znati verovatnoću da SAT rezultati u vašem uzorku premašuju 1380.

Na vašem grafikonu funkcije gustine verovatnoće, verovatnoća je osenčeno područje ispod krive linije koje se nalazi desno od mesta gde je vaš SAT rezultat jednak 1380.



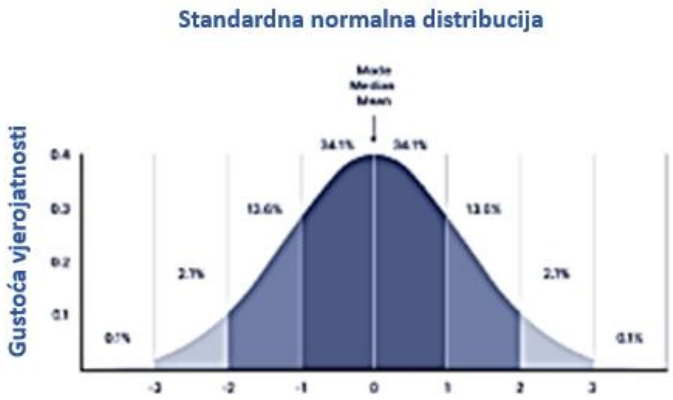

**Slika 2.100 Grafikon funkcije gustine verovatnoće SAT rezultata.**

Vrednost verovatnoće 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 vrednost od 0 i standardnu devijaciju od 1. Svaka normalna distribucija može se promatrati kao transformacija standardne normalne distribucije, koja prolazi kroz prilagođavanja u merama, položaju ili oba.

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




**Slika 2.120 Grafikon standardne normalne distribucije.**

Trebate znati samo srednju vrednost i standardnu devijaciju vaše distribucije da biste pronašli $z$-skor vrednosti.

Objašnjenje formule z-skora

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

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

Normalne distribucije pretvaramo u standardnu normalnu distribuciju iz nekoliko razloga:

- kako bismo pronašli verovatnoću opažanja u distribuciji koja pada iznad ili ispod zadate vrednosti;
- kako bismo pronašli verovatnoću da se srednja vrednost uzorka značajno razlikuje od poznate srednje vrednosti populacije.
- za upoređivanje rezultata na različitim distribucijama s različitim srednjim vrednostima i standardnim odstupanjima.

## 2.5 Određivanje verovatnoće korišćenjem $z$-distribucije

Svaki z-rezultat odgovara verovatnoći, koja se često naziva p-vrednost, koja ukazuje na verovatnoću opažanja vrednosti ispod tog specifičnog z-skora. Transformacijom pojedinačne

vrednosti u z-skor, može se odrediti verovatnoća da se sve vrednosti do te tačke pojave unutar normalne distribucije.

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

**Izračun formule**

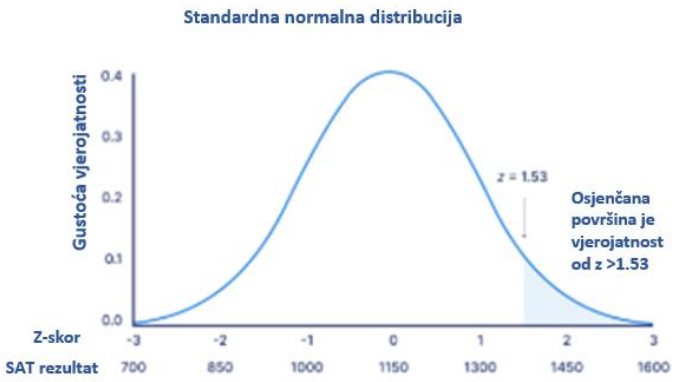
$$z = \frac{x - \mu}{\sigma} = \frac{1380 - 1150}{150} = 1.53$$

Za $z$ -skor od 1,53, $p$ -vrednost je 0,937. Ovo je verovatnoća da će SAT rezultati biti 1380 ili manje (93,7%), a to je područje ispod krive linije levo od osenčanog područja.

Da biste pronašli osenčano područje, oduzmite 0,937 od 1, što je ukupna površina ispod krive linije.

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

To znači da je verovatno 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, omogućavajući nam izvođenje zaključaka o populacijama na osnovu podataka iz uzorka. Udubićemo se u zamršenost

sampling-distribucija, da bi razumeli kako se odražava varijabilnost statistike uzorka i njihova ključna uloga u testiranju hipoteza.

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

## 2.7 Centralna granična teorema i sampling-distribucija

Centralna granična teorema (engl. Central Limit Theorem - CLT) je osnovni koncept u statistici koji podržava ponašanje sampling-distribucije. Navodi se da se sampling-distribucija srednje vrednosti uzorka približava normalnoj distribuciji kako se veličina uzorka povećava, bez obzira na oblik distribucije populacije. Ova teorema nam omogućava da izvedemo čvrste zaključke o parametrima populacije iz uzorka podataka.

Centralna granična teorema služi kao kamen temeljac razumevanja normalnih distribucija u statistici. U uslovima istraživanja, dobijanje tačne procene srednje vrednosti populacije često uključuje prikupljanje podataka iz brojnih slučajnih uzoraka unutar populacije. Te pojedinačne srednje vrednosti uzoraka zajedno čine ono što je poznato kao sampling-distribucija srednje vrednosti.

Centralna granična teorema ističe dva ključna principa:

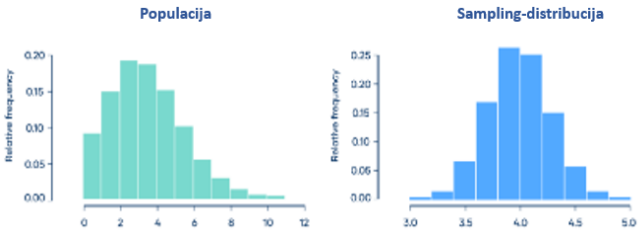
1. **Zakon velikih brojeva:** kako se veličina uzorka ili broj uzoraka povećava, srednja vrednost uzorka nastoji se približiti srednjoj vrednosti populacije.
2. **Normalnost sampling-distribucije:** uprkos izvornoj distribuciji varijable, kada se radi s višestrukim velikim uzorcima, sampling-distribucija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

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

Nasuprot tome, za male uzorke, osiguravanje pretpostavke normalnosti je ključno zbog nesigurnosti koja okružuje sampling-distribuciju srednje vrednosti. Tačni rezultati zahtevaju potvrdu da se populacija pridržava normalne distribucije pre korišćenja parametarskih testova s malim uzorcima.

Ilustrativno, centralna granična teorema tvrdi da će dobijanjem dovoljno velikih uzoraka iz populacije, srednje vrednosti tih uzoraka pokazati normalnu distribuciju, čak i ako osnovna distribucija populacije odstupa od normalnosti.

Primer: Razmotrite populaciju prema Poissonovoj distribuciji (prikazano na levoj slici). Nakon izvlačenja 10 000 uzoraka iz ove populacije, od kojih se svaki sastoji od 50 opažanja, distribucija srednjih vrednosti uzorka blisko je usklađena s normalnom distribucijom, u skladu s centralnom graničnom teoremom (kao što je ilustrovano na desnoj slici).




**Slika 2.155 Primer populacije u Poissonovoj distribuciji i normalnoj distribuciji.**

Centralna granična teorema zavisi od pojma sampling-distribucije, koja predstavlja distribuciju verovatnoće statistike izračunate iz brojnih uzoraka izvučenih iz populacije.

Konceptualizacija eksperimenta može pomoći u shvatanju sampling-distribucije:

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

Združivanje ovih srednjih vrednosti uzoraka predstavlja primer sampling-distribucije. Prema centralnoj graničnoj teoremi, sampling-distribucija srednje vrednosti teži 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 vrednosti pokazuje normalnost.

Srećom, ne treba više puta uzorkovati populaciju da bi se utvrdio oblik sampling-distribucije. Umesto toga, parametri sampling-distribucije srednje vrednosti zavise od parametara same populacije.

- Srednja vrednost sampling-distribucije je srednja vrednost populacije.

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

- Standardna devijacija sampling-distribucije je standardna devijacija populacije podeljena s kvadratnim korenom veličine uzorka.

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

Sampling-distribuciju srednje vrednosti možemo opisati pomoću ove oznake:

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

gde:

- $\bar{X}$ je sampling-distribucija srednjih vrednosti uzorka
- ~ znači "sledi distribuciju"
- $N$ je normalna distribucija
- $\mu$ je srednja vrednost populacije
- $\sigma$ 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 utiče na sampling-distribuciju srednje vrednosti 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 vrednosti može odstupati od normalnosti. Ovo odstupanje nastaje jer valjanost centralne granične teoreme zavisi od "dovoljno velike" veličine uzorka.
- Uobičajeno, uzorak od 30 ili više smatra se "dovoljno velikim".
- Kada je n<30, centralna granična teorema se ne primenjuje, a sampling-distribucija odražava distribuciju populacije. Stoga je sampling-distribucija normalna samo ako je distribucija populacije normalna.
- Nasuprot tome, kada je n≥30, centralna granična teorema važi, a sampling-distribucija približava se normalnoj distribuciji.

2. Veličina uzorka i standardna devijacija:

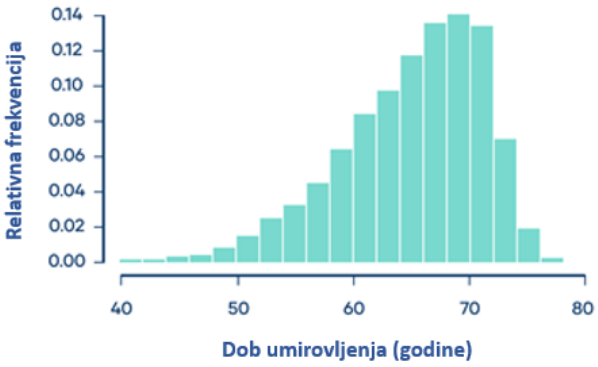
- Veličina uzorka direktno utiče na standardnu devijaciju sampling-distribucije, odražavajući varijabilnost ili raspršenost distribucije.
- S manjim uzorcima, standardna devijacija obično je viša, što ukazuje na veću varijabilnost među srednjim vrednostima uzorka zbog njihove neprecizne procene srednje vrednosti populacije.
- Suprotno tome, veći uzorci odgovaraju nižim standardnim devijacijama, što ukazuje na manju varijabilnost među srednjim vrednostima uzorka zahvaljujući njihovoj tačnijoj proceni srednje vrednosti populacije.

### Važnost centralne granične teoreme:

Parametarski testovi kao što su t-testovi, ANOVA i linearna regresija imaju veću statističku snagu u poređenju s većinom neparametarskih testova. Ova povećana statistička snaga proizilazi iz pretpostavki o distribuciji populacija, koje su zasnovane na centralnoj graničnoj teoremi.

### Kontinuirana distribucija:

Razmotrimo starost za odlazak u penziju pojedinaca u Sjedinjenim Američkim Državama. Stanovništvo se sastoji od svih penzionisanih Amerikanaca, a distribucija ovog stanovništva može se predstaviti na sledeći način:




**Slika 2.182 Grafikon kontinuirane distribucije**

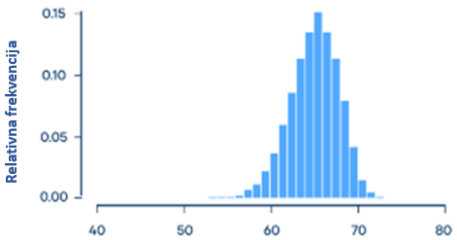
Distribucija starosti za penzionisane iskrivljena je ulevo, pri čemu većina odlazi u penziju unutar približno pet godina od prosečne starosti za penzionisanje od 65 godina. Međutim, postoji prošireni rep pojedinaca koji odlaze u penziju puno ranije, npr. s 50 ili čak 40 godina. Populacija pokazuje standardnu devijaciju od 6 godina.

Zamislite sprovođenje malog uzorkovanja ove populacije. Nasumično se izabere pet penzionera i beleži se njihova starost za odlazak u penziju. Na primer: 68, 73, 70, 62, 63.

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

Sada pretpostavimo da se ovaj proces uzorkovanja ponovi 10 puta, a svaki uzorak uključuje pet penzionera. Izračunava se srednja vrednost svakog uzorka, što rezultuje distribucijom poznatom kao sampling-distribucija srednje vrednosti. Na primer: 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 vrednosti ovih uzoraka približno će odgovarati normalnoj distribuciji.



**Slika 2.208 Normalna distribucija srednjih vrijednosti**

Uprkos tome što sampling-distribucija pokazuje nešto normalniji oblik u poređenju s populacijom, još uvek zadržava blagi zaokret ulevo. Osim toga, evidentno je da je varijabilnost u sampling-distribuciji uža od one populacije.

Prema centralnoj graničnoj teoremi, sampling-distribucija srednje vrednosti nastoji se približiti normalnoj distribuciji kako se veličina uzorka povećava. Međutim, trenutna sampling-distribucija srednje vrednosti odstupa od normalne zbog relativno male veličine uzorka.

## 2.8 Testna statistika

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

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

Ali šta tačno čini testnu statistiku?

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

Testna statistika sažima vaše opažene podatke u jedinstvenu vrednost, koristeći mere 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), podeljenih s varijansom podataka (tj. standardnom devijacijom).

Razmotrite ovaj primer:

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

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

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

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

## 2.9 Vrste testne statistike

U nastavku je prikazan sinopsis preovlađujućih 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, osnovne hipoteze i tumačenja testne statistike ostaju dosledni.

| **Testna statistika** | **Nulta i alternativna hipoteza** | **Statistički testovi** |
|---|---|---|
| ***t* vrednost** | **Nulta:** Srednje vrednosti dve grupe su jednake.<br>**Alternativna:** Srednje vrednosti dve grupe nisu jednake. | • *T* test<br>• Regresijski testovi |
| ***z* vrednost** | **Nulta:** Srednje vrednosti dve grupe su jednake.<br>**Alternativna:** Srednje vrednosti dve grupe nisu jednake. | • *Z* test |
| ***F* vrednost** | **Nulta:** Varijacija između dve ili više grupa veća je ili jednaka varijaciji između grupa.<br>**Alternativna:** Varijacije između dve ili više grupa su manje od varijacija između grupa. | • ANOVA<br>• ANCOVA<br>• MANOVA |
| ***X*$^2$-vrednost** | **Nulta:** Dva su uzorka nezavisna. | • Hi-kvadrat test |

| Testna statistika | Nulta i alternativna hipoteza | Statistički testovi |
|---|---|---|
| | **Alternativna:** Dva uzorka nisu nezavisna (tj. koreliraju). | • Neparametarski korelacioni testovi |

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

Na primer, u testiranju vaše hipoteze o temperaturi i datumima cvetanja, provodite regresionu analizu. Regresioni test daje: • regresioni koeficijent od 0,36 • t-vrednost koja upoređuje ovaj koeficijent s očekivanim rasponom regresionih koeficijenata pod nultom hipotezom nepostojanja veze.

Rezultujuća t-vrednost iz regresionog testa od 2,36 predstavlja vašu testnu statistiku.

## 2.10 Standardna greška

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

Dok je standardna greška srednje vrednosti najčešće citirani oblik standardne greške, slične mere postoje za druge statističke parametre kao što su medijana ili proporcije. Standardna greška funkcionše kao prevlađujuća mera greške uzorkovanja, prikazujući nejednakost između parametra populacije i statistike uzorka.

Kako bi se ublažila standardna greška, preporučuje se povećanje veličine uzorka. Korišćenje velikog, slučajnog uzorka služi kao najučinkovitija strategija za smanjenje pristranosti uzorkovanja i povećanje pouzdanosti nalaza.

**Standardna greška i standardna devijacija** su mere varijabilnosti:

- **Standardna devijacija** opisuje varijabilnost **unutar jednog uzorka**.
- **Standardna greška** procenjuje varijabilnost **u višestrukim uzorcima** populacije.

Standardna devijacija služi kao deskriptivna statistika izvedena direktno iz podataka uzorka, dok standardna greška predstavlja inferencijalnu statistiku, obično procenjenu, osim ako nije poznat tačan parametar populacije.

## 2.11 Formula standardne greške

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

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

**Ako su poznati parametri populacije**

Kada je poznata standardna devijacija populacije, možete je koristiti u donjoj formuli za tačno izračunavanje standardne greške.

**Formula** **Obrazloženje**

$$SE = \frac{\sigma}{\sqrt{n}}$$

- $SE$ je standardna greška
- $\sigma$ je standardna devijacija populacije
- $n$ je broj elemenata u uzorku

**Ako su parametri populacije nepoznati**

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

Primer: Korišćenje formule standardne greške za procenu standardne greške za rezultate SAT-a iz matematike. Sledite sledeća dva koraka.

Prvo pronađite kvadratni koren veličine uzorka (*n*).

| **Formula** | **Izračun** |
|---|---|
| $n = 200$ | $\sqrt{n} = \sqrt{200} = 14.1$ |

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

| **Formula** | **Proračun** | | |
|---|---|---|---|
| $SE = \frac{s}{\sqrt{n}}$ | $s = 180$ | $\sqrt{n} = 14.1$ | $\frac{s}{\sqrt{n}} = \frac{180}{14.1} = 12.8$ |

Standardna greška rezultata SAT iz matematike je 12,8.

Možete predstaviti standardnu grešku uz srednju vrednost ili je uključiti u interval pouzdanosti kako biste preneli nesigurnost koja okružuje srednju vrednost.

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

Izveštavanje o standardnoj grešci unutar intervala pouzdanosti je poželjno jer eliminiše potrebu čitaoca za izvođenje dodatnih izračunavanja kako bi dobili smisleni raspon.

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

Na nivou pouzdanosti od 95%, očekuje se da će 95% svih srednjih vrednosti uzorka pasti unutar intervala pouzdanosti koji obuhvata ± 1,96 standardnih grešaka srednje vrednosti uzorka. Ovaj interval služi kao procena unutar koje se veruje da se stvarni parametar populacije nalazi unutar 95% pouzdanosti.

Na primer: Konstruisanje intervala pouzdanosti od 95%. Konstruišete interval pouzdanosti od 95% (CI) da biste procenili srednju vrednost matematičke SAT ocene populacije. S obzirom na normalno raspodijeljenu karakteristiku kao što su SAT rezultati, otprilike 95% svih srednjih vrednosti uzorka pada unutar približno 4 standardne greške srednje vrednosti uzorka.

**Formula intervala pouzdanosti**

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

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

$SE$ = standardna greška = 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 slučajnim uzorkovanjem, 95% CI [525 575] ukazuje da postoji verovatnoća od 0,95 da je srednja vrednost 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 elektronskoj razmeni podataka (engl. *Electronic Data Interchange* - EDI), poruke koje sadrže kodove proizvoda ili usluga, identifikaciju transportnih jedinica, kao i pravne dokumente, razmenjuju se među partnerima u lancu snabdevanja. Obično imaju oblik alfanumeričkih nizova.

## 3.1 Informacije-Podaci-Znanje

U računarskim informacionim sistemima prikaz informacija varira zavisno od njihove namene i upotrebe. Prikaz može biti alfanumerički ili binarni s obzirom na to da li predstavlja tekst, sliku, zvuk ili izvršni program. Da bi se podaci različitih vrsta (npr. brojeva, znakova, datuma, valuta itd.). mogli sačuvati, preuzeti, obraditi i preneti) moraju biti ispravno kodirani. Alfanumerički formati predstavljanja podataka potiču 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 formate. Stoga omogućavaju zajedničko tumačenje podataka od strane poslovnih partnera koji pripadaju različitim etničkim gripama i geografskim sredinama. Dok alfanumerički nizovi zavise od njihovog kodiranja, numerički podaci se uglavnom razlikuju po veličini i/ili preciznosti.

Proces transformisanja podataka iz izvornog analognog u digitalni oblik popularno se naziva digitalizacija. Odgovarajuće dizajnirani pomoćni i aplikacioni programi koji prihvataju podatke iz različitih izvora (npr. optički skeneri, elektronski senzori, EDI, itd.) omogućavaju organizacijama da automatizuju svoje prikupljanje podataka, čuvanje, obradu i prenos unutar i između svojih računarskih informacionih sistema.

Kada se prikupe, podaci različitih vrsta mogu se spojiti i organizovati u tabele podataka, baze podataka, skladišta podataka (hronološki redosled) ili baze znanja (konceptualni redosled). Viši nivoi organizacije podataka omogućavaju automatizovano razvrstavanje, zaključivanje i predstavljanje tako akumuliranog znanja u analitičke svrhe.

## 3.2 Logistički podaci

U logistici se EDI koristi za prenos transakcionih 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 informacioni sistemi partnera (W3schools, 2023). Za brzu identifikaciju i manipulaciju osmišljeni su bar kodovi i RFID oznake.

Kako bi se omogućila međunarodna saradnja, trebalo je definisati 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 standardizovani i kodirani kao optički čitljivi bar-kodovi ili elektromagnetski čitljivi radiofrekventni identifikacioni (RFID) kodovi.

Bar-kod (EAN/UCC) je višesektorski i međunarodni oblik numerisanja stavki (POS EAN-8 i EAN-13, promenjivi EAN-128, traka podataka, ITF-14, QR, matrica podataka 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 primene.

RFID kodovi se prvenstveno koriste na isti način kao bar kodovi. Oni jedinstveno identifikuju artikle ili usluge. Obično, RFID nalepnice nose još više informacija na čipu veličine glave čiode. Osim identifikacije, RFID oznake omogućavaju beleženje podataka o praćenju, što je često potrebno u logističkim aplikacijama. Za razliku od linijskih kodova, RFID omogućava njihovo skeniranje bez direktnog vidnog polja, kao i skeniranje više nalepnica 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 direktnu, automatiziranu identifikaciju i praćenje logističkih jedinica unutar lanca snabdevanja.

EPCglobal standardi takođe predstavljaju osnovu GDSN-a (engl. *Global Data Synchronization Network*). Omogućavaju automatizivano prikupljanje i razmenu specifikacijskih podataka o proizvodima i njihovoj ambalaži, čime se preduzećima omogućava centralizovano upravljanje tim podacima kako bi ih oni i njihovi partneri naizmenično koristili.

Tabela 3.1 sažima različite identifikacione tehnologije s njihovim primenama. Otkriva niz jednodimenzionalnih i dvodimenzionalnih bar kodova, kao i različite klase RFID kodova s njihovim mogućnostima.

**Tabela 3.1Tehnologije označavanja.**

| **Tehnologija** | **Primena** |
|---|---|
| Bar kod 1D | Maloprodajni artikli i komponente proizvoda |
| Bar kod 2D | Usluge (npr. UPS, avionske karte), veleprodajni artikli koji zahtevaju praćenje |
| RFID klasa 1 (pasivno, R-oznake) | Predmeti koji zahtevaju masovnu identifikaciju, kontrolu pristupa |
| RFID klasa 2 (pasivno, RW-tagovi) | Stavke koje zahtevaju praćenje |
| RFID klasa 3 (poluaktivan, RW-tagovi) | Kontrola pristupa s dodatnim 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 zasnovane na lokaciji |

Budući trendovi u označavanju, praćenju i sledivosti slede dva glavna smera: minijaturizacija i raznolikost. Bar kodovi (1D) takođe moraju omogućiti označavanje minijaturnih predmeta (npr. medicinskih kapsula). Novi kodovi matrice podataka (2D) ne samo da će omogućiti ispravljanje grešaka tokom skeniranja, već i šifrovanje podataka.

RFID se nastavlja širiti na druga područja upotrebe kao što je identifikacija od strane davaoca usluga (npr. železnička kartica, prijava na posao, itd.), beskontaktna plaćanja (npr. bežični prenos novca, plaćanja na automatima) i pametna rešenja (npr. upravljanje pametnom kućom, daljinski upravljani pametni uređaji itd.) kao i e-valute.

## 3.3 Organizacija podataka

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

### Proračunske tabele

Prvi oblik organizacije podataka su dvodimenzionalni nizovi polja, koji se takođe nazivaju tabele ili proračunske tabele. Obično prvi red proračunske tabele označava značenja vrednosti unetih u osnovne kolone, nakon čega slede redovi podataka.

Polje ili ćelija tabele je najmanja jedinica podataka. Ima određeni tip (broj, datum, valuta itd.). Njegov sadržaj se može adresirati oznakama reda i kolone (npr. A1, koji predstavlja prvi red kolone A).

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

### Baze podataka

Datoteka ili tabela baze podataka je zbirka zapisa iste vrste. Baza podataka (engl. *Data Base* - DB) sastoji se od više međusobno povezanih tabela. 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 sačuvana u računaru.

Iz gornje definicije mogu se izvući neki zaključci o klijent-server arhitekturi gde server drži DB, kojoj pristupaju njegovi klijenti. Naravno, za pristup DB-u mora biti uspostavljena komunikaciona mreža između servera i njegovih klijenata. DB server se obično naziva njen "*back-end*", dok klijenti predstavljaju njen "*front-end*". Sistem za upravljanje bazom podataka (engl. *Data Base Management System* - DBMS) na serveru omogućava svojim klijentima pristup podacima skladištenim u DB-u putem svojih aplikacija programskog interfejsa (API) i DBM funkcija. DBM funkcije su mehanizmi koji omogućavaju unos, preuzimanje, obradu i prezentaciju podataka u DB-u. Za pozivanje ovih funkcija definisani su standardni jezici upita (engl. *Standard Query Languages* - SQL).

### Model relacione baze podataka

Postoje različiti oblici organizacije DB-a, a najčešći je relacioni model (RDB). Osnovna ideja ovog modela je činjenica da korisnik ne može unapred znati sve moguće upotrebe podataka koji se skladište u DB-u. Budući da obično ne postoje fiksni putevi pretraživanja kroz datoteke baze podataka, osmišljeni su različiti jezici upit za preuzimanje i manipulaciju podacima. RDB model se bazira na konceptu entiteta i odnosa:

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

RDB tabele, koje predstavljaju entitete ili relacije, međusobno su povezane pomoću ključeva. Skup atributa koji jedinstveno identifikuje entitet naziva se njegov primarni ključ. Kada se primarni ključ pojavljuje kao polje u drugoj tabeli s ciljem ispunjavanja relacije s izvornom tabelom, naziva se sekundarnim ili stranim ključem. Tabela može sadržati samo jedan primarni ključ za jedinstvenu identifikaciju svojih zapisa, dok može sadržati više sekundarnih ključeva.

Generalno, postoje dva pristupa konstruisanju RDB-a: analitički i sintetički.

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

1. Analiza stvarnog sveta - globalni model;
2. Određivanje entiteta i odnosa – konceptualni model (npr. E-R dijagram);

3. Određivanje logičkog modela – relaciona shema;

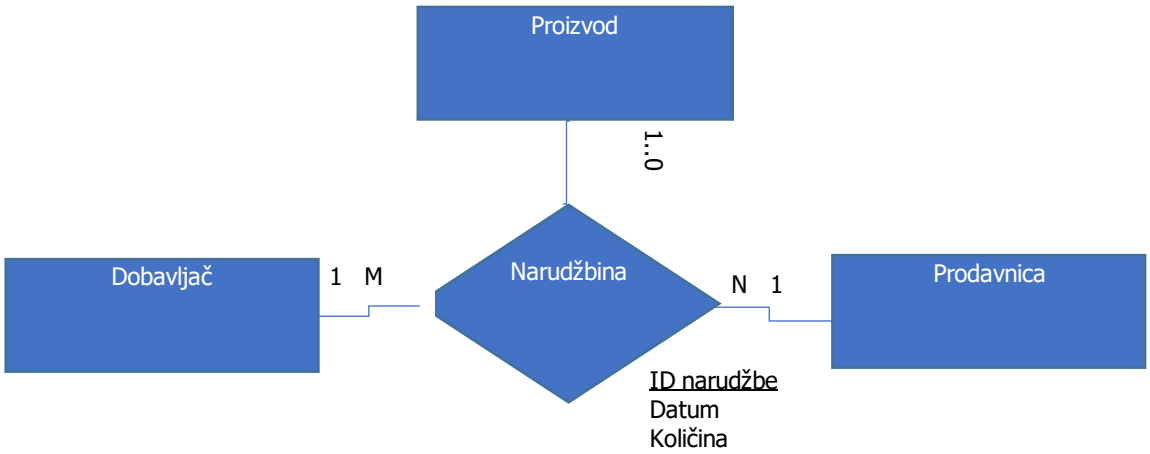
4. Izgradnja baze podataka (DBMS) – fizički model.

Kako bismo ilustrovali pristup, razmotrimo primer trgovinskog lanca i njegovih dobavljača (slika 3.1). Svaka prodavnica ima više dobavljača. Svaki dobavljač može snabdevati različite prodavnice. Ciklus popunjavanja započinje narudžbinom iz prodavnice. Zauzvrat dobavljač isporučuje robu u prodavnicu. Narudžbine su transakcije u kojima se spajaju podaci o prodavnici, dobavljaču i isporučenoj robi (slika 3.2).


**Slika 3.1 Globalni 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.**




**Slika 3.13 Konceptualni model.**

Logički model predstavlja RDB tabele prema vrstama njihovih zapisa. U logičkom modelu (slika 3.3) određeni atributi sadrže simbol tarabe (engl. *hashtag*) (#). To znači da polje koje oni predstavljaju jeste ili pripada (kompozitnom) primarnom ključu. Neka polja su podvučena. Nazivaju se sekundarnim ili stranim ključevima, budući da referenciraju primarne ključeve povezanih tabela.

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

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

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

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

Tabela u 2NF je u 1NF. Dodatno, nijedan atribut ključa ne sme delomično funkcionalno zavisiti od primarnog ključa. Tako se svi ključevi koji jedinstveno identifikuju određene atribute čuvaju odvojeno. Ovo je uglavnom da bi se osiguralo da se samo atributi koji zavise od svih (delova) primarnog ključa čuvaju u jednoj tabeli.

Tabela u 3NF je u 2NF. Osim toga, nijedan atribut koji nije ključ nije tranzitivno zavisan od primarnog ključa. To znači da se svi ne-ključni atributi koji mogu predstavljati ključ za određene druge ne-ključne atribute čuvaju u zasebnoj tabeli, pri čemu se samo njihov ključ održava u izvornoj tabeli kao strani ključ.

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

Tabela u Boyce-Codd NF (BCNF) je u 3NF; dodatno, svaka odrednica je ključ. Ovo uklanja sve odnose koji nisu obuhvaćeni postojećim redosledom ključeva iz izvorne tabele, čime se formiraju dodatne tabele za svaki ključ kandidata. Tabela u 4NF je u 3NF i BCNF. Osim toga, svaki atribut s više vrednosti koji delimično zavisi od ključa nalazi se u sopstvenoj tabeli. 4NF je namenjen uklanjanju svih mogućih preostalih atributa s više vrednosti iz originalne tabele. Tabela u 5NF je u 4NF; dodatno, svaka JOIN operacija je predviđena ključevima. Tabela u 6NF je u 5NF; dodatno, uzimaju se u obzir sve netrivijalne JOIN-zavisnosti.

Može se uočiti da kod viših oblika organizacije broj tabela raste sa svakim korakom. Stoga je razumno posmatrati fragmentaciju podataka kako bi se sprečilo stvaranje nepotrebnih tabela kojima se retko pristupa.

**Tabela 3.2Primer normalizacije RBD-a.**

<table>
<tr><td>ONF<br><br>NARUDŽBINA<br><br>• Dobavljač_ID *<br>• Ime_dobavljača<br>• Dobavljač_kontakt<br>• Prodavnica_ID *<br>• Naziv_prodavnice<br>• Adresa_prodavnice<br>• ID_narudžbine*<br>• Datum<br>• EPC<br>• Količina<br>• Naziv_proizvoda<br>• Cena_proizvoda</td><td>1NF<br><br>NARUDŽBINA<br><br>• Dobavljač_ID *<br>• Ime_dobavljača<br>• Dobavljač_kontakt<br>• Prodavnica_ID *<br>• Naziv_prodavnice<br>• Adresa_prodavnice<br>• ID_narudžbine*<br>• Datum<br>• EPC<br>• Količina<br>• Naziv_proizvoda<br>• Cena_proizvoda</td></tr>
<tr><td colspan="2">* Ključevi kandidata za ponavljajuću grupu atributa, koji jedinstveno identifikuju narudžbinu</td></tr>
<tr><td>2NF<br><br>DOBAVLJAČ<br><br>• Dobavljač _ID #<br>• Ime_dobavljača<br>• Dobavljač_kontakt<br><br>PRODAVNICA<br><br>• Prodavnica _ID #<br>• Naziv_prodavnice<br>• Adresa_prodavnice</td><td><br><br>NARUDŽBINA<br><br>• Narudžbina_ID#<br>• Dobavljač_ID#<br>• Prodavnica _ID #<br>• EPC<br>• Naziv_proizvoda<br>• Cena_proizvoda<br>• Datum<br>• Količina</td></tr>
</table>

| | |
|---|---|
| 3NF<br><br>NARUDŽBINA<br><br>• Dobavljač _ID #<br>• Prodavnica _ID #<br>• Narudžbina_ID#<br>• Datum<br>• Količina<br>• ID_proizvoda | <br><br>PROIZVOD<br><br>• EPC #<br>• Naziv_proizvoda<br>• Cena_proizvoda |

## Programski jezici za rad sa DB

Za upravljanje podacima uglavnom se koriste dva programska jezika: 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 direktno za upravljanje bazom podataka i skladištenje podataka.

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

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

Svaka manipulacija bazom podataka počinje njenim stvaranjem. Rečenica

CREATE DATABASE *baza podataka_naziv*;

stvara novu praznu bazu podataka s navedenim imenom.

Kao što je gore objašnjeno, podaci unutar baza podataka organizovani su u tabele zapisa podataka određenog tipa gde svi redovi dele zajedničku strukturu. Za izradu tabele koristi se sledeća rečenica:

CREATE TABLE *tabela_naziv* (
*kolona1 tip1,*
*kolona2 tip2, kolona3 tip3,*
…. );

Svaka imenovana kolona predstavlja atribut s određenim tipom podataka. Na primer, u:

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

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

kreiraju se dve tabele. Prva sadrži podatke o kupcima, dok druga sadrži podatke o njihovim narudžbinma, pozivajući se na prvu tabelu preko broja kupca kao stranog ključa.

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

CREATE INDEX *indeksa_naziv* ON *tabela_naziv* (*naziv_kolone*);

Svaka manipulacija podacima na indeksiranoj tabeli traje malo duže, budući da za njenu konzistentnost treba ne samo proveriti podatke koji daju ključeve i pravilno poređati podatke, već i druge atribute iz navedenog indeksa.

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

SELECT *kolona1, kolona2*, …
FROM *tabela_ naziv*;

Ovaj upit vraća podatke u koloni1, koloni2 itd. iz tabele. Rečenice upita obično se formiraju pružanjem dodatnih opcija, filtriranjem podataka, ispunjavanjem navedenih uslova:

WHERE navodi uslov koji određuje kriterijume izbora zapisa.

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

HAVING specificira agregatne funkcije na grupama definiranim naredbama GROUP BY.

ORDER BY specificira atribute prema kojima su povratni zapisi poređani.

Na primer:

```
SELECT "Prodavnica" . " Prodavnica_naziv " , "Proizvod" . " Proizvod_naziv " , " Narudžbina" .
"Količina " FROM "Narudžbina" , "Proizvod" , "Dobavljač" , " Prodavnica " WHERE
"Narudžbina" . " ID_proizvoda " = " Proizvod" . "EPC " AND "Narudžbina" . " Dobavljač_ID "
= "Dobavljač" . "Dobavljač_ID" AND "Narudžbina" . " Prodavnica _ID" =
"Prodavnica"."Store_ID" ORDER BY "Prodavnica"."Prodavnica_naziv" ASC
```

vraća popis prodavnica s njihovim naručenim proizvodima i količinama, poređanih prema nazivu prodavnice.

Najvažnija operacija u procesu selekcije je operacija JOIN. Često zamenjuje uslov WHERE kao JOIN ON, nakon čega sledi uslov. Upoređuje vrednosti kolona i na osnovu poređenja određuje da li ih treba uključiti u rezultat ili ne. U LEFT JOIN zapis se vraća ako su kriterijumi ispunjeni u levoj tabeli i obrnuto u operaciji RIGHT JOIN. Kao što je gore navedeno, uslov mora biti ispunjen u obe tabele kako bi bio u skladu s operacijom INNER JOIN ili FULL JOIN. Budući da se ovo drugo najčešće koristi, kao sinonim može se koristiti JOIN. Pozivajući se na uslove 5 i 6 normalne forme, ovo je ista JOIN operacija, koju je potrebno ispuniti da bi se ispunili uslovi odgovarajućeg NF-a.

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

```
INSERT INTO tabela_naziv (kolona1 , [kolona2 , ... ] )
VALUES ( vrednost1 , [vrednost2 , ...]);
```

Da bi bile uspešne, vrednosti u operaciji trebaju ispuniti sve uslove atributa označenih nazivima kolona. Nazive kolona ne treba navesti u slučaju da su sve vrednosti navedene. U slučaju da su u tabeli predviđene neke DEFAULT vrednosti, ne treba ih navoditi, osim ako se razlikuju.

Kada se podaci unesu, mogu se modifikovati naredbom UPDATE :

```
UPDATE tabela_naziv
SET kolona=vrednost1, kolona2=vrednost 2 ,...
WHERE neka_kolona = neka_vrednost;
```

U izjavi su date nove vrednosti za polja u navedenim kolonama. Kriterijumi izbora reda označeni su specifikatorom WHERE, koji određuje sve vrednosti kolona na koje se odnosi naredba UPDATE. Kako bi se sprečile neželjene promene, potreban je dodatni oprez pri formulisanju kriterijuma izbora.

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

DELETE FROM *tabela_naziv*
WHERE *neka_kolona = neka_vrednost*;

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

Naravno, upravljanje bazom podataka se ne završava ovde. Svaki element baze podataka takođe se može ukloniti, izmeniti i/ili zameniti novim. U slučaju da se indeks, tabela ili baza podataka trebaju ukloniti, mogu se primeniti sledeće izjave:

DROP INDEX *indeks_ime* ON *tablic_ime*;

DROP TABLE *tabela_naziv*;

DROP DATABASE *baza podataka_naziv*;

Ako neko samo želi ukloniti podatke iz tabele, može se koristiti naredba TRUNCATE:

TRUNCATE TABLE *tabela_naziv* ;

U slučaju da želite dodati ili ukloniti atribut (kolonu u/iz tabele, to možete učiniti naredbom ALTER:

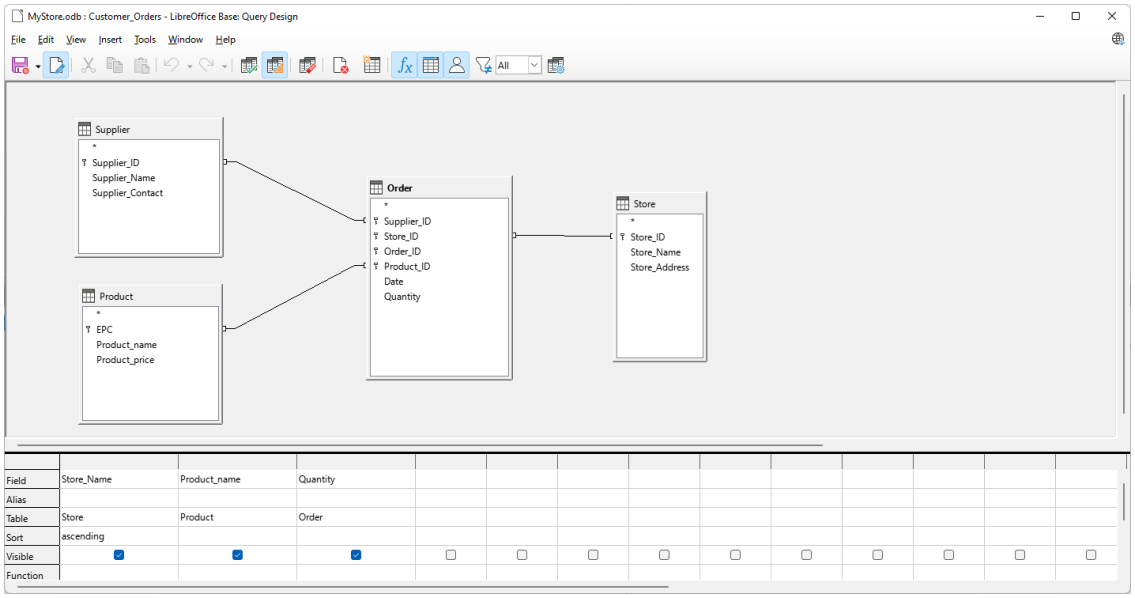
ALTER TABLE *tabela_naziv* ADD kolona_*naziv vrsta podataka*;

ALTER TABLE *tabela_naziv* DROP COLUMN kolona_*naziv;*

Ovim objašnjenjima se zaključuje kratki pregled SQL jezika i njegovih najčešćih scenarija upotrebe. SQL se obično koristi u klijent-server arhitekturama s DBMS-om koji je na glavnom računaru servera. Da bi mu se pristupilo, izdaju se SQL naredbe, bilo od strane aplikativnog programa klijenta ili web-interfejsa DBMS servera.

S druge strane, QBE se takođe često koristi s relacionim DBMS-om s grafičkim korisničkim interfejsom (engl. *Graphical User Interface* - GUI), kao što su MS Access ili LibreOffice Base. S QBE-om se baza podataka i njene tabele stvaraju mnogo interaktivnije, a njihovu strukturu je lakše održavati. Kao što mu ime govori, nudi i jednostavniji oblik unosa i otkrivanja podataka. Za izvođenje pretraživanja potrebno je sastaviti sve tabele koje se koriste u pretraživanju i zatim uspostaviti uslove kao uzorke u poljima kolona kako bi se filtrirali relevantni podaci (Slika 3.4). Formulacija upita dopunjena je postojećim relacijama između

tabela. Kao i obično, rezultat takvog upita je još jedna tabela s rezultirajućim podacima, koji se kasnije mogu dalje obraditi. Na ovaj način takođe se mogu formirati kaskadni ili višefazni upiti.




**Slika 3.15QBE upit ekvivalentan gornjem SQL upitu.**

## Filteri i maske podataka

Kako bi se sprečili pogrešni ili nepotpuni podaci, koji bi mogli ometati njihovu obradu i tumačenje, potrebno je primeniti dodatne mere opreza:

1. Filtriranje praznih redova i kolona;
2. Primena stroge tipizacije podataka za sprečavanje računarskih grešaka;
3. Definisanje maski za unos kako bi se sprečio unos pogrešnih podataka;
4. Pristranost podataka kako bi se sprečili pogrešni rezultati.

Prazni redovi i kolone čest su izvor grešaka koje uglavnom potiču od lošeg interfejsa u aplikacijama za prikupljanje podataka. Otkazane ili nedovršene transakcije obično rezultiraju praznim redovima ili nedostajućim podacima u zapisima transakcija. Oni se samo delimično mogu rešiti aplikacijama proračunskih tablica gde se prazni redovi i podaci koji nedostaju mogu otkriti proverom ukupnog broja u odnosu na broj podataka koji nisu nula u redovima/kolonama. Mogu se filtrirati uklanjanjem praznih redova/kolona, ali možda to nije uvek željena aktivnost jer bismo mogli izgubiti i neke vredne podatke. Najbolji način da se to

spreči je korišćenjem DBMS-a koji bi s jedne strane omogućio unos samo kompletnih transakcionih podataka, dok bi s druge strane takođe sprečio prazne redove/kolone, budući da ih u bazama podataka nema.

Slabo upisivanje podataka je još jedan uobičajeni izvor grešaka. Ako se umesto numeričkih podataka unesu alfanumerički podaci, poput datuma ili iznosa valute, to bi rezultiralo greškama prilikom obrade tih podataka. U proračunskim tabelama, kao i u bazama podataka, pojedinačnim ćelijama podataka, koje predstavljaju vrednosti atributa u slogovima podataka, mogu se dodeliti tipovi podataka, što bi nas alarmiralo pri unosu podataka u pogrešnom formatu. Stoga se ovakvom merom može sprečiti da pogrešni podaci ometaju njihovu obradu.

Prilikom konstruisanja baza podataka, ograničenja se mogu primeniti na polja podataka koja predstavljaju atribute entiteta ili relacije. Osim što im se može dodeliti odgovarajuća vrsta, maske za unos, mogu se definisati čime se omogućava unos podataka, poput datuma, valuta, EAN kodova itd., samo u određenom formatu. To obično rešava mnoge pogrešne percepcije koje bi inače mogle nastati tokom obrade podataka.

Još jedan čest izvor grešaka su netačni podaci, koji predstavljaju podatke koji su s aspekta veličine veći ili manji od očekivanog. Ovakvi podaci bi mogli ometati našu obradu, dajući pogrešne rezultate. Teže ih je otkriti i mogu se filtrirati samo gledanjem podataka. U aplikacijama za proračunske tabele dobra uobičajena praksa bila bi određivanje minimalnih, maksimalnih i srednjih vrednosti podataka u odgovarajućim kolonama kako bi se otkrila moguća odstupanja. Ako se otkriju, mogu se naglasiti i ručno obraditi, ako ih je malo, ili filtrirati i modifikovati upitom u tabeli baze podataka ako ih je puno. U svakom slučaju treba ih pažljivo proceniti, 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 katalogizuju hronološkim redom. Obično se na podacima sprovode neke poslovne analize, a odeljenje analitike takođe čuva rezultate za kasnije potrebe. Nakon što se memorišu, ti se podaci obično ne menjaju kako bi se očuvala njihova doslednost.

Osim hronološkog reda, prilikom izgradnje baza znanja uzimaju se u obzir i kontekstualni redovi. Ovde su subjekti predstavljeni kao derivati entiteta najvišeg nivoa ili njegovih jedinica. Onosi između njih se uspostavljaju slobodnije jer su namenjeni ažuriranju i nadogradnji kako se koriste. Odnosi se uspostavljaju u obliku pravila i zasnovani su na svojstvima entiteta. Stoga je oblik u kojem se skladište nešto drugačiji. Često se skladište u obliku ontologija koje sadrže dublje znanje o prikupljenim podacima. Slično čuvanju rezultata upita u skladištima podataka, upiti u bazama znanja takođe se čuva za kasniju upotrebu za prikaz trenutnih rezultata, kako se menjaju entiteti, relacije i instance podataka.

Za razliku od baza podataka i skladišta podataka, koji su specifični za aplikaciju, baze znanja mogu biti nezavisne od aplikacije i često ih različite aplikacije koriste među domenama. Primer 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 skladištenje podataka, prikazana je organizacija podataka i mehanizmi pronalaženja. Konačno, neke uobičajene greške u automatizovanoj obradi podataka su rešene kako bi čitalas ostao na oprezu. Osim navedenih primera, 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. Simulaciono modeliranje i analiza

Simulacionim modeliranjem i analizom (engl. *simulation modelling and analysis* - SMA) nastoje se ispuniti zahtevi Conant–Ashby teoreme (Conant i Ashby, 1970), definišući model sistema kao dobrog regulatora, koji ima onoliko ručki, delova i stanja koliko i njegov izvorni fizički pandan. Time pruža mogućnost izgradnje digitalnog modela i uspostavljanja digitalne laboratorije koja će omogućiti njegovo istraživanje, prilagođavanje i optimizaciju. Rezultujući simulacioni modeli su apstraktni, dinamički i u većini slučajeva stohastički, budući da su njihove varijable sistema modelirane distribucijama verovatnoće.

## 4.1 Simulacija u logistici

U logistici SMA može pružiti vredne inpute za optimizaciju lanaca snabdevanja (engl. *supply chain* - SC) i transportnih mreža (engl. *traffic network* - TN). Simulaciono modeliranje može se koristiti za grafičku vizualizaciju vremenskih tokova kroz složene SC i TN procese i resurse, omogućavajući predviđanje i kvantifikaciju mogućih ishoda iz različitih scenarija. Ovo pomaže SC i TN entitetima da izvuku vredne zaključke i razumeju učinke svojih potencijalnih odluka na SC i TN verzije uključujući SC vreme isporuke (engl. *lead-times*), TN vreme putovanja i troškove. Stoga, SMA u SC i TN modeliranju može doprineti SC i TN analizi i poboljšanju njihovog dizajna usmerenog na postizanje veće učinkovitosti i održivosti.

Postoje mnogi aspekti SC-a, koji predstavljaju različite perspektive upravljanja lancem snabdevanja. Aspekt posmatranja menadžera proizvodnje na lanac snabdevanja razlikuje se od menadžera marketinga, koji se opet razlikuje od menadžera nabavke, itd. Dakle, korišćeni modeli su različiti, čak i za istu kompaniju, a posebno za ceo SC.

Prilikom rešavanja problema SMA u logistici, menadžeri treba da donose odluke na strateškom, taktičkom i operativnom nivou, zavisno od njihovog učinku na SC ili TN u celini. Zbog svoje međuovisnosti, menadžeri često nisu u mogućnosti rešiti probleme ni na jednom nivou. Istovremeno, takođe je teško posmatrati sva tri nivoa iz perspektive bilo koje pojedinačne celine. Iz perspektive SMA, SC ili TN se mogu posmatrati na dva nivoa:

1. Makro nivo

- samoorganizacija,
- koevolucija entiteta,
- zavisnost od veza/transportnih ruta.

2. Mikro nivo

- 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. Zavisno od nivoa i perspektive, trajanje operacija može se meriti u danima, nedeljama ili čak mesecima kada su u pitanju među-organizacione aktivnosti, dok se s druge strane, unutar-organizacione operacije mere u satima ili čak sekundama. Zavisno od prirode modeliranog problema, trajanje najkraće operacije ili maksimalna učestalost dolaznih/odlaznih zahteva određuje ne samo prikaz vremena u SMA modelu, već i njegovu granularnost. Što je kraće minimalno trajanje najkraće operacije ili što je veća učestalost zahteva, to je finija granularnost vremena, odnosno preciznost vođenja vremena u modelu. Ovo je važno za modelara, budući da vreme reakcije modela ne može biti kraće od unapred definisane vremenske granularnosti. Stoga je potrebno unapred proceniti trajanje svih operacija i vremena između dolazaka dolaznih/odlaznih signala kako bi se mogle ispravno odrediti vremenske jedinice modela sistema.

U simulacionom modelu vreme može napredovati kritičnim događajima od transakcije do transakcije ili kontinualno. U poslednjem slučaju, tok vremena u modelu ne zavisi od 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 tokom simulacije može ubrzati tok vremena u modelu, tako da se procesi izvode brže nego u stvarnom vremenu. Stoga se mogu napraviti rana predviđanja sledećih događaja.

Vremena između dolaznih simulaconih jedinica i vremena njihove obrade/tranzita mogu proizaći iz posmatranja i merenja. Ako ne variraju, onda su deterministička. Međutim, obično su po svojoj prirodi stohastička. Stoga je potrebno uvođenje konstrukata koji modeliraju njihove funkcije distribucije verovatnoća (npr. trougaone, uniformne, eksponencijalne, itd.).

## 4.2 Simulacija diskretnih događaja

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

Konstrukti:

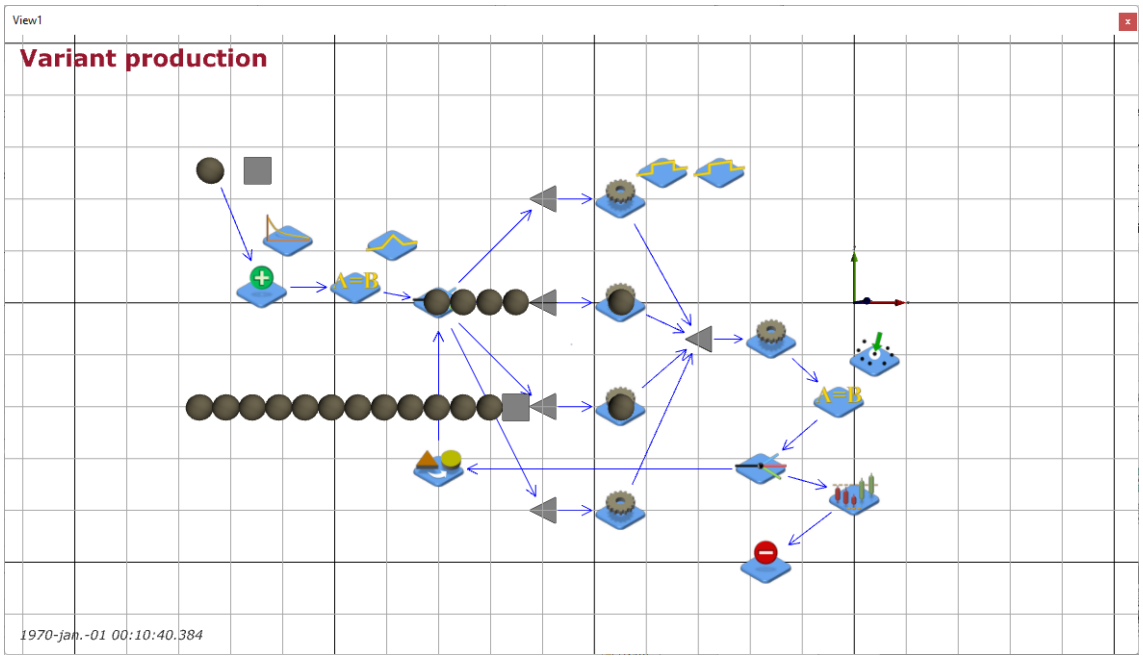
- Jedinice toka predstavljaju jedinice simulacije (npr. narudžbine, materijali itd.) koje ulaze u sistem na ulazu(ima) i napreduju kroz model sistema;
- Procesori predstavljaju mobilne (npr. ljudi, viljuškari itd.) i fiksne (npr. mašine, proizvodne linije itd.) resurse koji obrađuju simulacione jedinice;
- Redovi čekanja čuvaju jedinice toka do njihovog prelaza na sledeći dostupni procesor;
- Konektori definišu promicanje jedinica kroz model sistema.

Svojstva:

- Orijentisan na proces;
- Fokusira se na detaljno modeliranje procesa;
- Heterogeni entiteti;
- Mikro-entiteti su pasivni objekti;
- Događaji unose dinamiku u sistem;
- Diskretna vremenska progresija; od jednog (vremenskog) događaja do sledećeg;
- Fleksibilnost se postiže promenom strukture modela; struktura sistema tokom simulacije je fiksna.

## Primer

Primer DES-a (Slika 4.1, iz simulacionog okruženja JaamSim (JaamSim Development Team, 2023)) uključuje model varijante proizvodnje, gde 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. Izbor vrste proizvoda određen je trougaonom raspodelom između 1 i 4 s modulom na 3. Svaka vrsta proizvoda ima namensku proizvodnu liniju. Proizvodni nalozi se ispunjavaju po eksponencijalnoj distribuciji sa srednjom vrednošću od 30 s. Proizvodnja svakog pojedinačnog proizvoda traje 100-120s prema ravnomernoj distribuciji. Nakon što su finalizirani, kvalitet proizvoda se proverava na namenskom kontrolonom mestu. Provera kvaliteta traje 10 s. Iz iskustva kompanije, u proseku jedan od 10 proizvoda ne prođe kontrolu. Proizvodi nezadovoljavajućeg kvaliteta transportuju se nazad na izvornu proizvodnu liniju. Njihova ponovna obrada traje 120-130s prema ravnomernoj raspodeli. Trajanje proizvodnje i kontrole kvaliteta te ponovne obrade ne zavise od vrste proizvoda. Nakon što su uspešno prošli kontrolu kvaliteta, gotovi proizvodi se transportuju s mesta proizvodnje u skladište gotovih proizvoda. Ponovna obrada neispravnih proizvoda dok su još u proizvodnji je način da se smanji uticaj na životnu sredinu i troškove proizvodnje.




**Slika 4.1Varijanta proizvodnje s kontrolom kvaliteta.**

Sinopsis

DES može da analizira i optimizira sledeće procesne parametre:

- Vreme proizvodnog ciklusa i učinak.
- Iskorišćenost proizvodnih ćelija i prostora.
- Kapacitet skladišnih prostora kao i vreme zadržavanja skladišnih jedinica.
- Korišćenje mobilnih resursa (npr. operateri, pokretne trake, viljuškari).

## 4.3 Sistemska dinamika

Analiza sistemske dinamika (engl. *system dynamics* – SD) predstavlja pogled menadžera SC-a na proizvodni proces pomoću doslednog i koherentnog modela. SD se smatra najprikladnijim alatom za određivanje strukture, kao i optimalnih količina za pojedinačne lokacije (kada i koliko inputa, zaliha i outputa). Stoga, omogućava učinkovito korišćenje proizvodnih i skladišnih objekata.

Konstrukti:

- Zalihe predstavljaju tampone koji mogu čuvati stavke isporuke u lancu snabdevanja.
- Tokovi predstavljaju kanale snabdevanja.
- Petlje povratne sprege predstavljaju parametre finog podešavanja za popunjavanje zaliha.

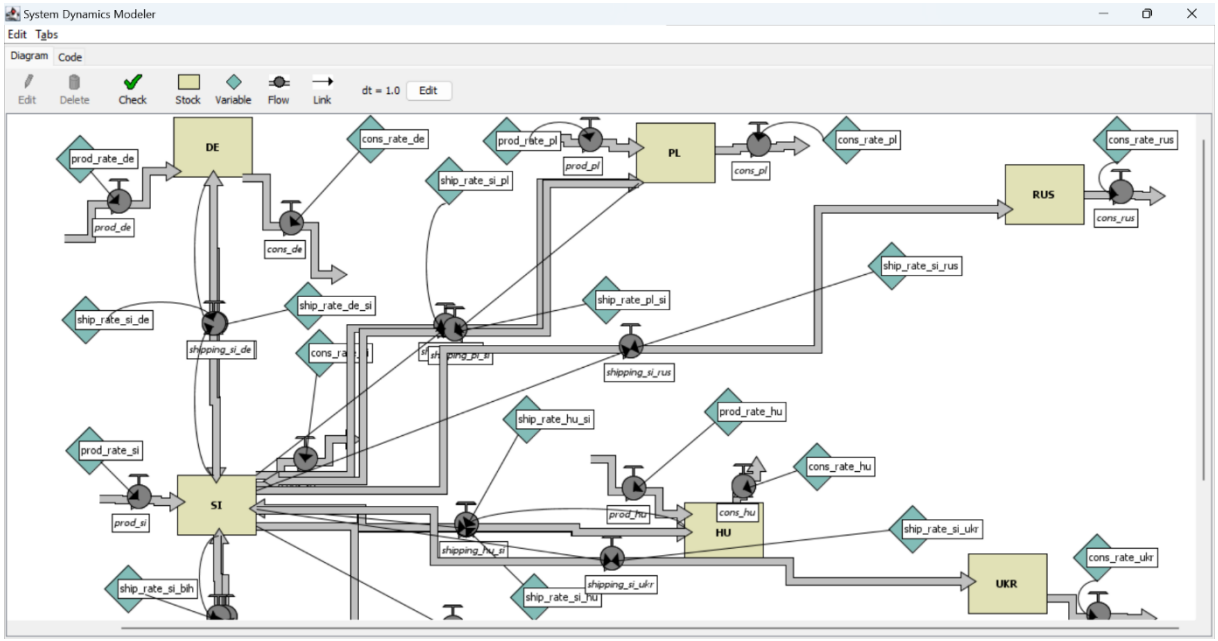
Svojstva:

- Usmeren na sistem.
- Modeliranje varijabli sistema je usmereno na ključne pokazatelje učinka.
- Homogeni entiteti.
- Entiteti na mikronivou se zanemaruju.
- Dinamika se uvodi petljom povratne sprege.

- Kontinuirana vremenska progresija; vreme napreduje sinhronizovano za sve komponente modela sistema.
- Fleksibilnost se postiže promenom strukture modela.
- Struktura sistema tokom simulacije je fiksna.

### Primer

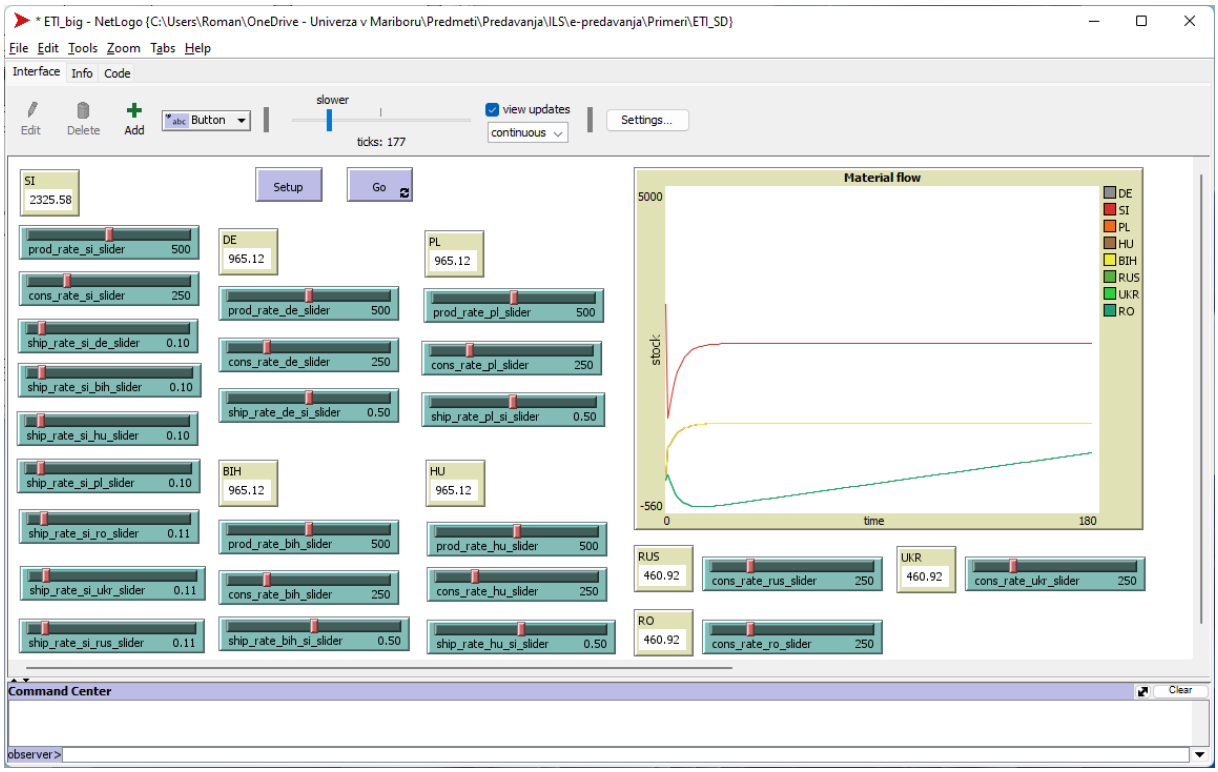
Primer SD-a (Slika 4.2, iz simulacionog okruženja NetLogo (Wilensky, 1999)) obuhvata lanac snabdevanja kompanije kućnih aparata i opisuje tokove materijala između njenih podružnica (Gumzej i Rakovska, 2020). Kompanija ima više proizvodnih lokacija: glavnu lokaciju u Sloveniji (SI) kao i podružnice u Nemačkoj (DE), Poljskoj (PL), Mađarskoj (H) i Bosni i Hercegovini (BIH). Uz proizvodna mesta, veleprodajna mesta kompanije nalaze se u Rusiji (RUS), Ukrajini (UKR) i Rumuniji (RU). Proizvodne lokacije snabdevaju sopstvena tržišta gotovim proizvodima i jedna drugu komponentama proizvoda.




**Slika 4.2Layout lanca snabdevanja.**

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 tok je neprekidan kroz svakodnevne transakcije, tj. svaki dan se određeni broj komponenti isporučuje između proizvodnih mesta i određeni broj gotovih proizvoda se troši na licu mesta ili se otprema na mesta distribucije. Na osnovu

početnih zaliha od 300 jedinica na SI lokaciji, 0 zaliha na drugim lokacijama i modela distribucije, količine zaliha na pojedinačnim lokacijama predstavljaju prosečne zalihe prema definisanoj proizvodnji (kom), potrošnji (%) i otpremi (%).




**Slika 4.3Nadzorna ploča lanca snabdevanja.**

### Sinopsis

Simulacija sistemske dinamike omogućava:

- Planiranje layouta SC-a.
- Optimizaciju proizvodnih i distributivnih kapaciteta.
- Procenu opterećenja kanala distribucije i povezanih troškova.

## 4.4 Simulacija zasnovana na agentima

Analiza simulacijama zasnovanim na agentima (engl. *agent-based simulation* – ABS) 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, kvalitet, svest, itd.).

Konstrukti:

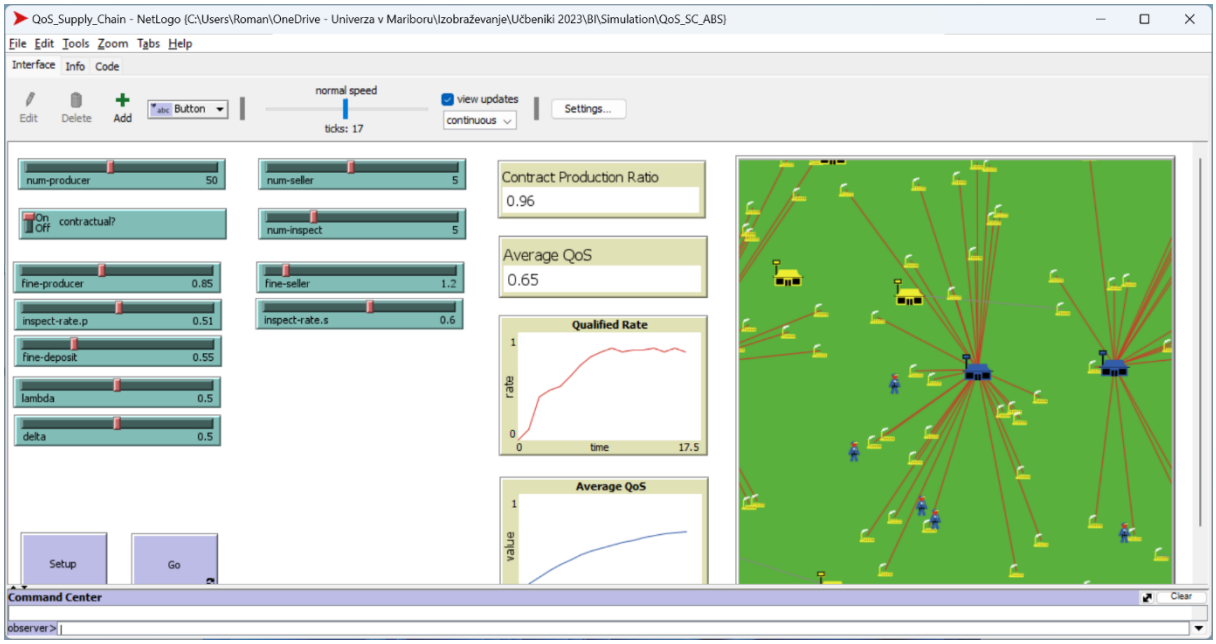
Agenti koji predstavljaju čvorove lanca snabdevanja (npr. dobavljači, trgovci na malo i inspektori) sa svojim svojstvima, odnosima i ponašanjem.

Svojstva:

- Fokusiran na entitet.
- Problemski orijentirano modeliranje entiteta i njihovih interakcija.
- Heterogenost entiteta.
- Mikro-entiteti su aktivni objekti koji deluju u svom okruženju, međusobno komuniciraju i samostalno donose odluke.
- Odluke i interakcije između agenata unose dinamiku u sisteme.
- Agenti i njihovo okruženje čine formalne modele.
- Protok vremena je diskretan i univerzalan na nivou modela; vreme modela je u skladu s učestalošću SC transakcija i životnim ciklusima SC čvorova.
- Fleksibilnost modela postiže se promenom strukture sistema i ponašanja agenata.
- Struktura sistema tokom simulacije je promenjiva.

### Primer

Primer ABS-a (slika 4.4, iz simulacionog okruženja NetLogo (Wilensky, 1999)) korišćen je za analizu ponašanje ešalona lanca snabdevanja na otvorenom tržištu (Gumzej i Rakovska, 2020), s obzirom na njihov kvalitet usluge (engl. *Quality of Service* - QoS). U primeru, različite politike koje su vezane za ukupno upravljanje kvalitetom kompanije istraživane su modelom koji se sastoji od dobavljača, kupaca i regulatora tržišta.




**Slika 4.4Regulacija tržišta.**

### Sinopsis

Simulacija zasnovana na agentima omogućava:

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

## 4.5 Simulacija mreže

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

Konstrukti:

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

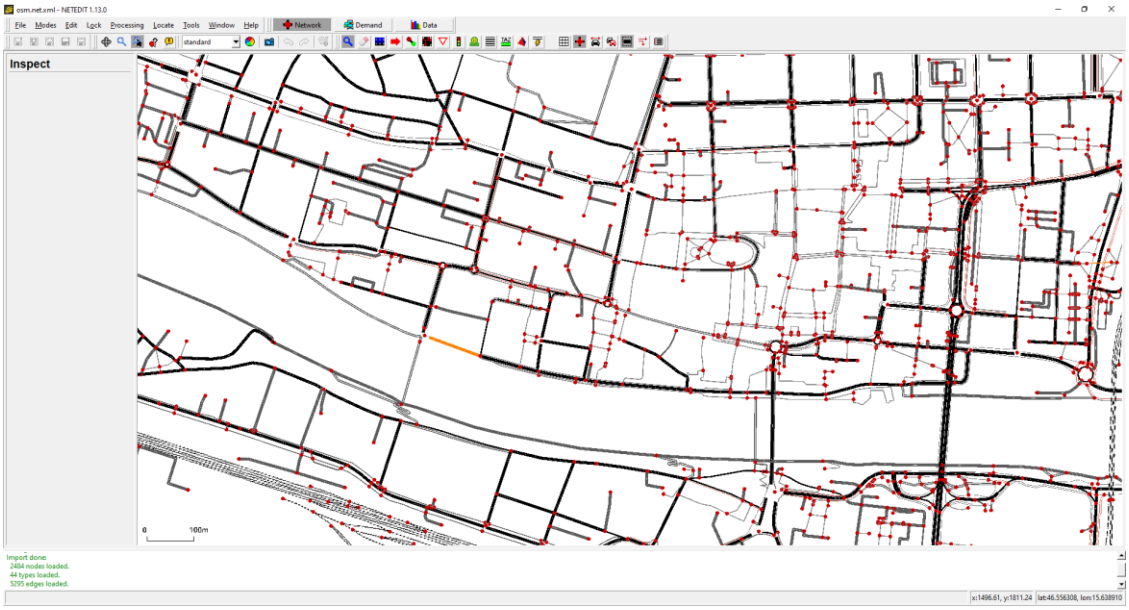
Mreža koja predstavlja mrežu preklapanja (npr. saobraćajnu mrežu) na kojoj se objekti toka kreću.

Svojstva:

- Usmeren na sistem.
- Problemski orijentisano modeliranje entiteta i njihovih interakcija.
- Heterogenost entiteta.
- Mikro-entiteti su aktivni objekti koji deluju u svom okruženju, međusobno komuniciraju i samostalno donose odluke.
- Odluke i interakcije između agenata unose dinamiku u sisteme.
- Agenti i njihovo okruženje čine formalne modele.
- Protok vremena je diskretan i univerzalan na nivou modela; vreme je u skladu s relativnim brzinama objekata toka.
- Fleksibilnost modela postiže se promenom mrežne strukture koja se fiksira tokom simulacije i ponašanja agenata koje varira zavisno od stanja (prometa) mreže i njihovih ciljeva.

### Primer

Predstavljeni primer (slika 4.5, iz SUMO (Pablo et.al., 2018.) simulacionog okruženja) korišćen je za određivanje saobraćajnih tokova i propusnosti ulica u središtu grada pogođenih planiranom blokadom saobraćajnica (Šinko i Gumzej, 2021). Osim toga, mereni su pokazatelji povezani sa saobraćajem, poput vremena putovanja, potrošnje goriva i emisija.




**Slika 4.5Saobraćajna situacija i mreža.**

Sinopsis

Mrežna simulacija omogućava:

- 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

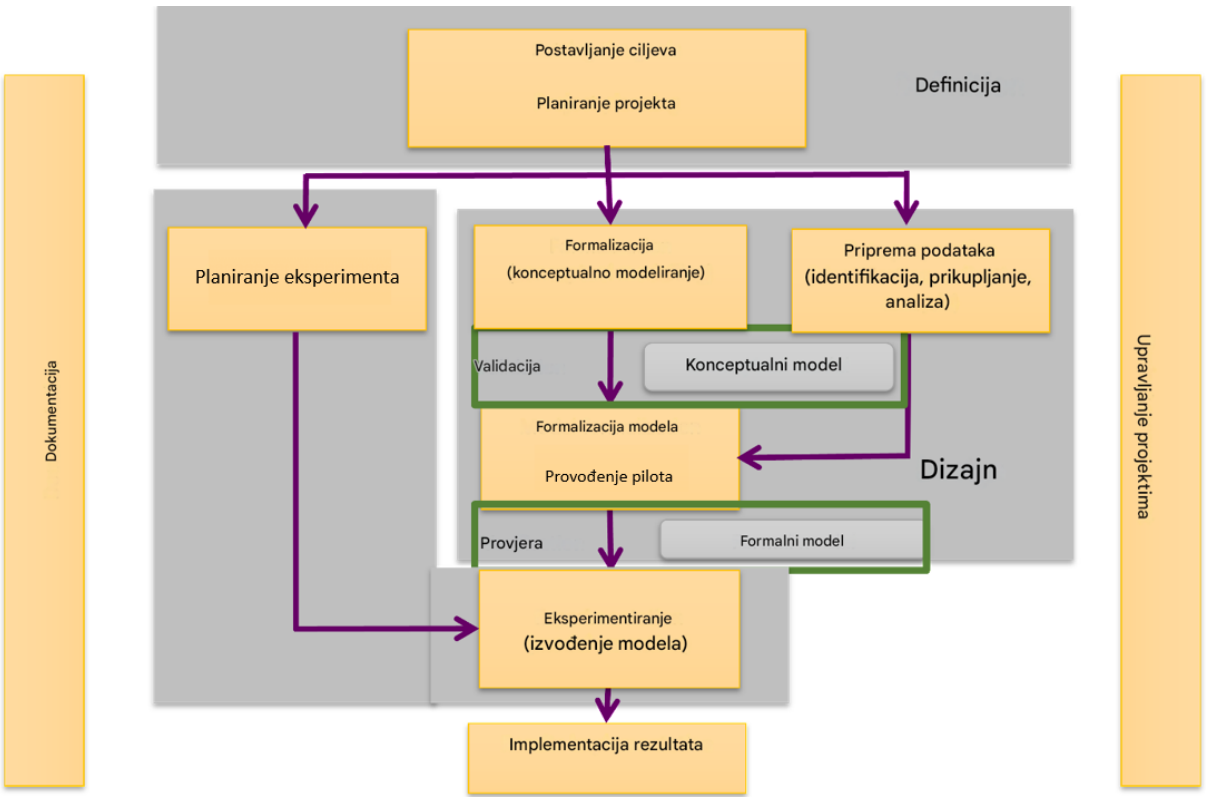
Projekti logističke simulacije dizajnirani su u skladu s paradigmom Šest sigma (engl. *Design for Six Sigma* - DFSS) i zasnivaju se na Demingovom ciklusu poboljšanja:

- Planiranje: definiranje sistema i ciljeva.
- Postavka: dizajn simulacionog modela.
- Analiza: eksperimentisanje sa simulacionim modelom i procena alternativa.
- Akcija: korišćenje rezultata simulacije za

implementaciju poboljšanja.

Svaki projekat 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 sistema i definicija predispozicija, prikupljanje podataka.
3. Logički model: tok objekta, zalihe i tok ili mrežni dijagram modela sistema.
4. Simulacioni model: izrada odgovarajućeg simulacionog modela.
5. Verifikacija i validacija simulacionog modela: provera konzistentnosti i koherentnosti modela.
6. Analiza na osnovu simulacionog modela: dizajn i izvođenje eksperimenta.
7. Korišćenje rezultata simulacije za izradu akcionog plana: projekcija poboljšanja sistema.




**Slika 4.6Proces simulacionog modeliranja i analize (SMA).**

## 4.7 Zaključak

U logistici je SMA važna komponenta operacionih istraživanja (OR) koja omogućava optimizaciju procesa.

Sistemska dinamika (SD) je metodologija za analizu složenih, dinamičkih i nelinearnih interakcija u sistemima, što rezultuje novim strukturama i politikama za poboljšanje ponašanja sistema. Ovde se rešavaju fizički i informacioni tokovi s ciljem smanjenja njihovog kašnjenja i konačno zaliha u lancu snabdevanja.

Druga popularna procesno orijentisana metodologija je simulacija diskretnih događaja (DES). To je jedan od najčešće korišćenih i najfleksibilnijih analitičkih alata u SMA proizvodnih sistema. Uspešno se nosi s neizvesnošću i pruža mogućnosti upoređivanja alternativnih načina za smanjenje vremena isporuke kao i optimizaciju korišćenja mašina i resursa.

Korisna metodologija za razumevanje ponašanja organizacija i njihovih interakcija (npr. lanaca snabdevanja i njihovih entiteta) je simulacija zasnovana na agentima (ABS). Simulacija mreže (NS), kao posebna vrsta ABS-a, omogućava modeliranje i optimizaciju (saobraćajnih) mreža.

Zaključak je da holistički pristup primene SMA u logistici uveliko doprinosi složenim odlukama o dizajnu sistema, gde postoji mnogo varijabli koje međusobno deluju. Koristan integrisani pristup, uključujući metodologije SD, DES i ABS, koji može kvantifikovati protoke na različitim nivoima lanca snabdevanja, predstavljen je u (Gumzej & Rakovska, 2020). U analizi i optimizaciji saobraćajnog toka, NS metodologija pruža potreban okvir u vezi s praćenjem i finim podešavanjem ključnih pokazatelja uspešnosti (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 regresionih varijabli

Odluke menadžmenta često se zasnivaju na odnosima između dve ili više varijabli. Na primer, menadžer marketinga može pokušati da predvidi prodaju za određeni nivo troškova oglašavanja nakon ispitivanja odnosa između tih izdataka i prodaje.

U drugom slučaju, javno poduzeće može koristiti odnos između maksimalne dnevne temperature i potrebe za električnom energijom, za predviđanje potrošnje električne energije. Menadžer se ponekad oslanja na intuiciju i intuitivno procenjuje kako su dve varijable povezane. Međutim, ako je moguće dobiti podatke, ima smisla koristiti statistički postupak koji se zove regresiona analiza kako bi se pokazalo kako su te dve varijable međusobno povezane.

U terminologiji regresije, varijabla čija se vrednost predviđa naziva se zavisna varijabla.

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

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

U ovom odjeljku će se objasniti najjednostavnija vrsta regresione analize koja uključuje jednu nezavisnu varijablu i jednu zavisnu varijablu. Odnos između dve varijable se aproksimira pravom linijom. Naziva se jednostavnom linearnom regresijom. Regresiona analiza koja uključuje dve ili više nezavisnih varijabli naziva se višestruka regresiona analiza.

## 5.1 Jednostavni linearni regresioni model

Best Burger je lanac restorana brze hrane koji se nalazi u području s više država. Najbolje lokacije Burgera nalaze se u blizini univerzitetskih kampusa. Menadžeri veruju da je tromesečna prodaja u ovm restoranima (označeno s $y$) u pozitivnoj

korelaciji s veličinom studentske populacije (označeno s $x$). Restorani u blizini kampusa s velikim brojem studenata obično generišu veću prodaju od onih u blizini kampusa s malim brojem studenata. Pomoću regresione analize možemo razviti jednačinu koja pokazuje kako je $y$ zavisna varijabla povezana s nezavisnom varijablom $x$.

## 5.2 Regresioni model i regresiona jednačina

U slučaju Best Burgera, populaciju čine svi restorani Best Burger. Za svaki restoran u populaciji postoji vrednost $x$ (studentska populacija) i odgovarajuća vrednost $y$ (tromesečna prodaja). Jednačina koja opisuje kako je $y$ povezana s $x$ zove se regresioni model.

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

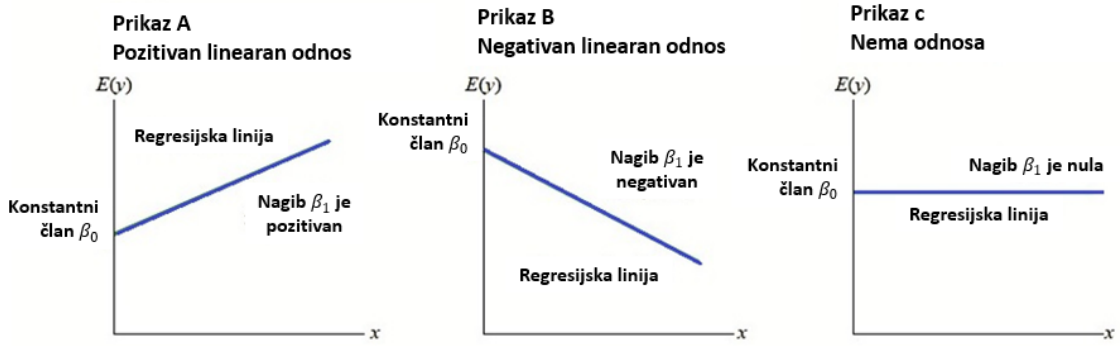
$\beta_0$ i $\beta_1$ nazivaju se parametri modela, $\epsilon$ (grčko slovo epsilon) je slučajna varijabla koja se naziva greška modela. Greška predstavlja varijabilnost $y$ što se ne može objasniti linearnim odnosom između $x$ i $y$.

Populacija svih restorana Best Burger takođe se može posmatrati kao zbirka podpopulacija, jedna za svaku zasebnu vrednost $x$. Na primer, jednu podpopulaciju čine svi restorani Best Burger u blizini univerzitetskih kampusa s 8000 studenata. Drugu podpopulaciju čine svi restorani Best Burger koji se nalaze u blizini univerzitetskih kampusa s 9000 studenata i tako dalje. Svaka subpopulacija ima odgovarajuću raspodelu vrednosti $y$. Svaka raspodela vrednosti $y$ ima svoju srednju ili očekivanu vrednost. Jednačina koja opisuje šta je očekivana vrednost $y$, označena s $E(y)$, čime je povezana s $x$ naziva se regresiona jednačina. Regresiona jednačina za jednostavnu linearnu regresiju je sledeća:

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

Grafikon jednostavne jednačine linearne regresije je prava linija. $\beta_0$ predstavlja početnu vrednost regresione linije, $\beta_1$ je koeficijent smera linije i $E(y)$ srednju vrednost ili očekivanu vrednost $y$ za datu vrednost x.

Primeri mogućih regresionih linija prikazani su na slici 5.1 u nastavku. Regresiona linija $y$ u slučaju A pokazuje da je vrednost $y$ u pozitivnoj korelaciji s $x$. Kako se vrednosti $x$ povećavaju, vrednosti $E(y)$ se takođe povećavaju. Tamo gde su manje vrednosti $E(y)$ povezane su s višim vrednostima $x$. Regresiona linija na prikazu C prikazuje slučaj kada vrednost $y$ nije povezana s $x$. To znači da je vrednost $y$ ista za svaku vrednost $x$.




**Slika 4.7Primeri grafikona linearnog odnosa.**

## 5.3 Procenjena regresiona jednačina

Kad bi bile poznate vrednosti parametara populacije $\beta_0$ i $\beta_1$, mogli bismo koristiti gornju jednačinu za izračunavanje vrednosti $y$ za zadatu vrednost $x$. U praksi je tim parametrima teško pristupiti, pa se jednostavno procenjuju korišćenjem podataka uzorka. Statistika uzorka (označena s $b_0$ i $b_1$) utvrđena je kao procena parametara populacije $\beta_0$ i $\beta_1$. Zamenom vrednosti statistike uzorka $b_0$ i $b_1$ umesto $\beta_0$ i $\beta_1$ u regresionoj jednačini dobija se nova, procenjena regresiona jednačina. Procenjena regresiona jednačina za jednostavnu linearnu regresiju je sledeća:

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

Grafikon procenjene jednostavne linearne regresije naziva se procenjena regresiona linija. $b_0$ predstavlja početnu vrednost regresione linije, $b_1$ je koeficijent smera linije.

U nastavku ćemo pokazati kako koristiti metodu najmanjih kvadrata za izračunavanje vrednosti $b_0$ i $b_1$ u procenjenoj regresionoj jednačini.

Generalno je $\hat{y}$ (rezultat za $E(y)$) prosečna vrednost $y$ za datu vrednost $x$. Ako sada želimo proceniti očekivanu vrednost tromesečne prodaje za sve restorane Best Burger koji se nalaze u blizini kampusa s 10 000 studenata, vrednost $x$ bi bila zamijenjena vednošću 10 000 u posljednjoj jednačini. U nekim slučajevima, međutim, možda ćemo biti više zainteresovani za predviđanje prodaje samo za jedan određeni restoran, na primer, pretpostavimo da želite predvideti 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 vrednosti $y$ za datu $x$ vrednost $\hat{y}$.

## 5.4 Metoda najmanjih kvadrata

Metoda najmanjih kvadrata je postupak u kojem se pomoću uzoraka podataka nalazi jednačina procenjene regresione linije. Kako bismo ilustrovali metodu najmanjih kvadrata, pretpostavimo da su podaci prikupljeni iz uzorka od 10 restorana s najboljim hamburgerima u blizini univerzitetskih kampusa. Sa $x_i$ će se označiti veličina studentske populacije (u hiljadama) i veličina $y_i$ tromesečna prodaja (u hiljadama EUR). Vrednosti za $x_i$ *i* $y_i$ za 10 uzoraka restorana sažeti su u tabeli u nastavku. Vidimo da je restoran 1, za $x_1$= 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 vrednošću je restoran 10, koji se nalazi u blizini kampusa s 26.000 studenata i ima kvartalnu prodaju od 202.000 €.

Sledi dijagram rasipanja podataka na slici 5.2 u nastavku. Studentska populacija prikazana je na horizontalnoj osi, a kvartalna prodaja na vertikalnoj osi. Dijagrami rasipanja za regresionu analizu konstruisani su s nezavisnom varijablom $x$ na horizontalnoj osi i zavisnom varijablom $y$ na vertikalnoj osi. Stoga, dijagram rasipanja omogućava 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 rasipanja podataka.**

Koji se preliminarni zaključci mogu izvući sa slike 5.3? Veća tromesečna prodaja događa se u kampusima s većom populacijom studenata. Osim toga, postoji konstantan odnos između veličine studentske populacije i tromesečne prodaje, koji se može opisati pravom linijom. Između $x$ *i* $y$ zaista postoji pozitivan linearni

odnos. Stoga smo odabrali jednostavan linearni regresioni model za prikaz odnosa između tromesečne prodaje i studentske populacije. S obzirom na ovaj izbor, naš sledeći zadatak je koristiti tabelu podataka uzorka za određivanje vrednosti $b_0$ i $b_1$, koji su važni parametri u proceni jednostavne jednačine linearne regresije. Za i-ti restoran procenjena regresiona jednačina je:

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

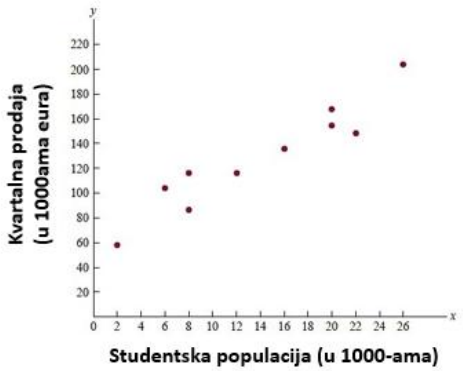
gde je

$\hat{y}_i$_ procenjena vrednost tromesečne prodaje (1000 €) za i-ti restoran

$b_0$_ početna vrednost procenjene regresione linije

$b_1$_ koeficijent smera procenjene regresione linije

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




**Slika 4.9 Grafikon rasipanja.**

$y_i$ označava opaženu (stvarnu) prodaju za restoran $i$ i $\hat{y}_i$, predstavljajući procenjenu vrednost prodaje za restoran $i$, svaki restoran u uzorku će imati opaženu prodajnu vrednost od $y_i$ i predviđenu prodajnu vrednost $\hat{y}_i$. Kako bi procenjena regresiona linija osigurala dobro uklapanje u podatke, želimo da razlike između opaženih prodajnih vrednosti i predviđenih prodajnih vrednosti budu što manje.

Metoda najmanjih kvadrata koristi uzorke podataka za dobijanje vrednosti $b_0$ i $b_1$.

Minimizirajte zbir kvadrata odstupanja između posmatranih vrednosti zavisne varijable $y_i$ i predviđene vrednosti zavisne varijable $\hat{y}_i$. Polazna tačka za izračunavanje minimalnog zbira metodom najmanjih kvadrata data je izrazom

Kriterijum minimalnog iznosa: $\min \sum(y_i - \hat{y}_i)^2$

gde je

$y_i$=posmatrana vrednost zavisne varijable za i-to opažanje

$\hat{y}_i$ =predviđena vrednost zavisne varijable za i-to opažanje

Koeficijent smera regresione linije i početna vrednost:

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

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

$x_i$_ vrednost nezavisne varijable za i-to opažanje

$y_i$_ vrednost zavisne varijable za i-to opažanje

$\bar{x}$ _ prosečna vrednost za nezavisnu varijablu

$\bar{y}$ _ prosečna vrednost za zavisnu varijablu

$n$_ukupan broj opažanja

Neki od proračuna potrebnih za izradu procenjene linije regresije najmanjih kvadrata prikazani su u nastavku. Na uzorku od 10 restorana imamo n=10 opažanja. Gornje jednačine prvo zahtevaju utvrđivanje srednje vrednosti $x$ i prosečne vrednosti $y$.

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

Alternativna jednačina izračunava $b_1$:

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

Koristeći poslednje jednačine i informacije na slici 5.4, možemo izračunati usmereni koeficijent regresione linije za primer restorana Best Burger. Izračunavanje nagiba ($b_1$) je kako slijedi.

Slika 5.5 prikazuje dijagram ove jednačine na dijagramu rasipanja.

Nagib procenjene regresione jednačine ili koeficijent smera jednačine ($b_1 = 5$) 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 |
| | $\Sigma x_i$ | $\Sigma y_i$ | | | $\Sigma(x_i - \bar{x})(y_i - \bar{y})$ | $\Sigma(x_i - \bar{x})^2$ |

**Slika 4.10 Prikaz jednačine na dijagramu rasipanja.**

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

Nakon toga sledi utvrđivanje početne vrednosti ($b_0$).

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

Ovako se procenjuje regresiona jednačina:

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

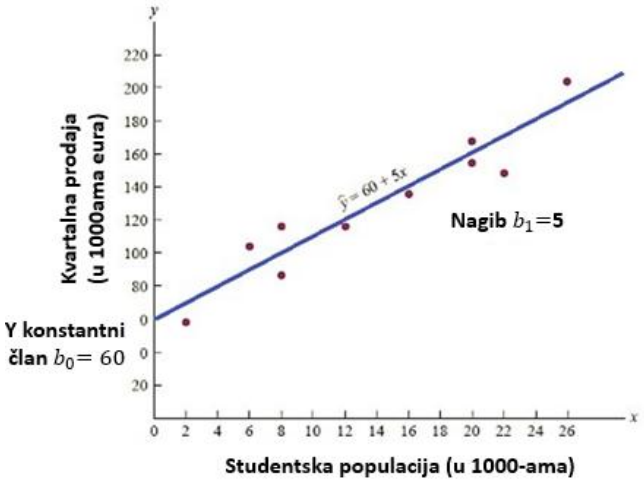
Slika prikazuje dijagram ove jednačine na dijagramu rasturanja.

Nagib procenjene regresione jednačine ($b_1 = 5$) je pozitivan, što znači da kako se broj studenata povećava, prodaja se povećava. Zapravo, možemo zaključiti (na osnovu izmerene 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 tromesečne prodaje za 5 € po studentu.

Ako verujemo da regresiona jednačina, procenjena najmanjim kvadratima, adekvatno opisuje odnos između $x$ i $y$, čini se razumnim koristiti procenjenu regresionu jednačinu za predviđanje vrednosti $y$ za datu vrednost $x$. Na primer, ako želite predvideti tromesečnu 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 sledećim odjeljcima raspravljamo o metodama za procenu prikladnosti korišćenja procenjene regresione jednačine za procenu i predviđanje.




**Slika 4.11Dijagram rasipanja studentske populacije i tromesečne prodaje.**

## 5.5 Koeficijent determinacije

Za primer restorana Best Burger razvili smo procenjenu regresionu jednačinu: $y = 60 + 5x$ za približno linearni odnos između veličine studentske populacije $x$ i tromesečne prodaje $y$. Sada je pitanje: koliko dobro procenjena regresiona jednačina odgovara podacima? U ovom odjeljku pokazujemo da koeficijent determinacije daje meru dobrog uklapanja za procenjenu regresionu jednačinu. Za i-to opažanje, razlika između opažene vrednosti zavisne varijable $y_i$ i predviđene vrednosti zavisne varijable naziva se i-to rezidualno odstupanje.

Zbir kvadrata ovih rezidualnih odstupanja ili grešaka je vrednost koja je minimizirana metodom najmanjih kvadrata. Ova vrednost, takođe poznata kao zbir kvadrata reziduala, označava se sa SSE.

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

SSE vrednost je mera greške u korišćenju procenjene regresione jednačine za predviđanje vrednosti zavisne varijable u uzorku. Slika 5.6 prikazuje proračune potrebne za izračunavanje zbira kvadrata zbog greške 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 grešaka u slučaju Best Burger.**

Pretpostavimo da se od nas traži da uradimo procenu tromesečne prodaje pri čemu ne znamo veličinu studentske populacije. Bez poznavanja bilo koje povezane varijable, koristili bismo prosek uzorka kao procenu tromesečne prodaje u bilo kom restoranu. Tabela na slici 5.6 pokazala je da je podatke o prodaji $y_i$=1300. Stoga je prosečna tromesečna vrednost prodaje za uzorak od 10 najboljih restorana s hamburgerima $y_i$/n = 1300/10 = 130. Na slici 5.7 prikazujemo zbir kvadrata odstupanja dobijenih korišćenjem srednje vrednosti uzorka od 130 za predviđanje vrednosti kvartalne prodaje za svaki restoran u uzorku. Za i-ti restoran u uzorku razlika $y_i$ daje meru greške koja je uključena u aplikaciju za predviđanje prodaje. Odgovarajući zbir kvadrata, koji se naziva ukupan zbir 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 $y_i - y$ | 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 |

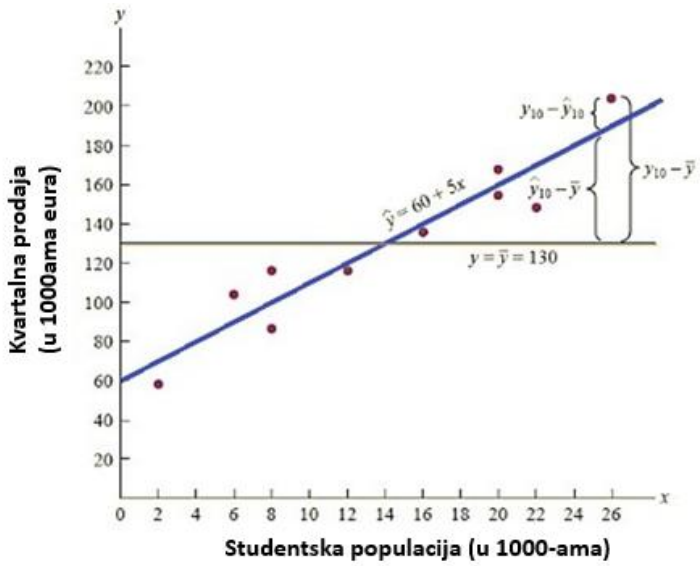
**Slika 4.13 Zbir kvadrata odstupanja.**

Zbir na dnu poslednje kolone na slici 5.7 je ukupni zbir kvadrata za BestBurgerove restorane SST = 15,730. Na slici 5.8 prikazujemo procenjenu regresionu liniju $y = 60 + 5x$ i liniju koja odgovara y = 130. Imajte na umu da se tačke grupišu bliže oko procenjene regresione linije

nego oko linije y = 130. Na primer, za 10. restoran u uzorku, vidi se da je greška puno veća kada se 130 koristi za predviđanje y = 10 nego kada se 130 koristi $y = 60 + 5x$ i iznosi 190. Možemo se setiti SST kao mere koliko dobro se opažanja grupišu oko linije i SSE kao mere koliko dobro se opažanja grupišu oko linije.

Da bi se izmerilo koliko vrednosti na procenjenoj regresionoj liniji odstupaju od sledećeg, izračunava se još jedan zbir kvadrata. Ovaj zbir kvadrata, koji se naziva regresionim zbirom kvadrata, označava se kao SSR.

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




**Slika 4.14 Regresiona linija u slučaju Best Burgera.**

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

## 5.6 Odnos između SST, SSR i SSE

$$SST = SSR + SSE$$

Gde je:

SST = ukupan zbir kvadrata

SSR = regresioni zbir kvadrata

SSE = rezidualni zbir kvadrata

Jednačina ( $SST = SSR + SSE$) pokazuje da se ukupni zbir kvadrata može podeliti na dve komponente, regresioni zbir kvadrata i rezidualni zbir kvadrata. Dakle, ako su poznate vrednosti bilo koja od ova dva zbira kvadrata, treći zbir kvadrata može se lako izračunati. Na primer, u slučaju Best Burger restorana, već znamo da je SSE = 1530 i SST = 15,730; stoga, rešavanjem za SSR u gornjoj jednačini, nalazimo da je regresioni zbir kvadrata

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

Sada ćemo videti kako možemo koristiti tri zbira kvadrata, SST, SSR i SSE, da bismo dobili kriterijum prilagođavanja za procenjenu regresionu jednačinu. Procenjena regresiona jednačina bi savršeno odgovarala ako bi svaka vrednost zavisne varijable $yi\_$ ležala slučajno na procenjenoj regresionoj 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 poklapanje SSR mora biti jednak SST i odnos (SSR/SST) mora biti jednak jedinici. Lošije prilagođavanje rezultiraće većim vrednostima za SSE. Rešavajući SSE u jednačini, vidimo da je SSE = SST - SSR. Stoga se najveća vrednost za SSE (a time i najlošije uklapanje) javlja kada je SSR = 0 i SSE = SST.

Za procenu se koristi odnos SSR/SST, koji ima vrednosti između nula i jedan koji odgovara procenjenoj regresionoj jednačini.

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

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

Za primer restorana Best Burger, vrednost koeficijenta determinacije je

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

Kada se koeficijent determinacije izrazi kao procenat, $r^2$ se može tumačiti kao procenat ukupnog zbira kvadrata koji se može objasniti pomoću procenjene regresione jednačine. Za najbolje restorane s hamburgerima možemo zaključiti da se 90,27% ukupnog zbira kvadrata može objasniti korišćenjem procenjene regresione jednačine $y = 60 + 5x$ 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 procenjenu regresionu jednačinu.

## 5.7 Koeficijent korelacije

Koeficijent korelacije može se smatrati opisnom merom snage linearnog odnosa između dve varijable, x i y. Vrednosti koeficijenta korelacije su uvek između -1 i +1. Vrednost +1 znači da su $x\ i\ y$ dve varijable u savršenoj korelaciji u pozitivnom linearnom smislu. To znači da su svi podaci na liniji a da je prava s pozitivnim nagibom. Vrednost -1 znači da su $x\ i\ y$ savršeno povezane u negativnom linearnom smislu, sa svim podacima na pravoj liniji s negativnim nagibom. Vrednosti koeficijenta korelacije blizu nule znače da x i y nisu linearno povezani.

Ako je regresiona analiza već sprovedena i koeficijent determinacije $r^2$ je izračunat, koeficijent korelacije uzorka može se izračunati na sledeći način.

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

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

PEARSONOV KOEFICIJENT KORELACIJE: UZORCI PODATAKA

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

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

gde 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 procenjena regresiona jednačina ima pozitivan nagib ($b_1 > 0$) i negativan ako procenjena regresiona jednačina ima negativan nagib ( $b_1 < 0$).

Za slučaj Best Burger, vrednost koeficijenta determinacije koji odgovara procenjenoj regresionoj jednačini $y = 60 + 5x$ je 0,9027. Budući da je nagib procenjene regresione jednačine pozitivan, jednačina 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 dve varijable, oba koeficijenta determinacije i koeficijent korelacije uzorka daju meru jačine odnosa.

Koeficijent determinacije daje meru između 0 i 1, dok koeficijent korelacije uzorka daje meru između -1 i +1. Iako je koeficijent korelacije uzorka ograničen na linearni odnos između dve varijable, koeficijent determinacije može se primeniti na nelinearne odnose i na odnose koji imaju dve ili više nezavisnih varijabli. Dakle, koeficijent determinacije pruža širi raspon primenjivosti.

## 5.8 Model višestruke regresije

U sledećim odjeljcima nastavljamo naše proučavanje regresione analize razmatrajući situacije koje uključuju dve ili više nezavisnih varijabli. Ovo predmetno područje, koje se naziva višestruka regresiona analiza, omogućava nam da uzmemo u obzir više faktora i tako dobijemo bolja predviđanja nego što je to moguće s jednostavnom linearnom regresijom.

Višestruka regresiona analiza je proučavanje o tome kako je zavisna varijabla y povezana s dve ili više nezavisnih varijabli. U opštem slučaju, s p ćemo označiti broj nezavisnih varijabli.

## 5.9 Regresioni model i regresiona jednačina

Koncepti regresionog modela i regresione jednačine uvedeni u prethodnom odjeljku primenjuju se u slučaju višestruke regresije. Jednačina koja opisuje kako je zavisna varijabla y povezana s nezavisnim varijablama $x_1, x_2, \ldots, x_p$ i greškom naziva se modelom višestruke regresije. Počinjemo s pretpostavkom da model višestruke regresije ima sledeći 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 greška ($\epsilon$) je slučajna varijabla. Pažljivo ispitivanje ovog modela otkriva da je y linearna funkcija varijabli $x_1, x_2, \ldots, x_p$ plus

greška $\epsilon$. Izraz greške 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 vrednost ($\epsilon$) nula. Implikacija ove pretpostavke je da je srednja ili očekivana vrednost $y$, označena s $E(y)$, jednaka $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$. Jednačina koja opisuje kako je srednja vrednost $y$ povezana s $x_1, x_2, \ldots, x_p$ naziva se jednačina višestruke regresije:

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

## 5.10 Procenjena jednačina višestruke regresije

Ako su vrednosti $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ poznate, jednačina iz 5.9 se može koristiti za izračunavanje prosečne vrednosti y pri datim vrednostima $x_1, x_2, \ldots, x_p$. Nažalost, ove vrednosti parametara generalno neće biti poznate i moraju se proceniti iz podataka uzorka. Jednostavan slučajni uzorak koristi se za izračunavanje statistike uzorka $b_0, b_1, b_2, \ldots, b_p$ koja se koristi kao tačkasti procenitelj parametara $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$. Ovi uzorci statistike daju sledeću procenu jednačine višestruke regresije:

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

gde su

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

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

### Primer: Frigo transportna kompanija

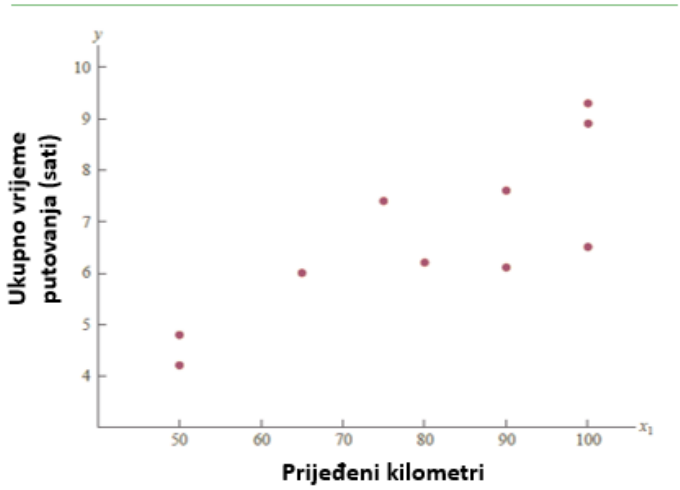
Kao ilustraciju višestruke regresione analize, razmotrićemo problem s kojim se susreće Frigo transportna kompanija, nezavisni autoprevoznik u južnoj Italiji. Najveći deo poslovanja kompanije Frigo odnosi se na dostavu na celom lokalnom području. Za bolji raspored rada, menadžeri žele planirati zajedničko dnevno vreme putovanja za svoje vozače.

U početku su menadžeri verovali da će ukupno dnevno vreme putovanja biti usko povezano s brojem pređenih kilometara u dnevnim isporukama. Jednostavan slučajni uzorak od 10 dodeljenih vozača dao je podatke prikazane na slici 5.9 i dijagramu rasipanja. Nakon

pregleda ovog dijagrama rasipanja, menadžeri su pretpostavili da se jednostavni linearni regresioni model može koristiti za opisivanje odnosa $y = \beta_0 + \beta_1 x_1 + \epsilon$ između ukupnog vremena putovanja $(y)$ i broja pređenih 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 kompanije.**




**Slika 4.16 Dijagram rasipanja za Frigo transportnu kompaniju.**

Za procenu parametara $\beta_0$ i $\beta_1$ ,jednačina najmanjih kvadrata korišćena je za izradu procenjene regresije.

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

Prethodna slika daje prikaz softverskog programa Minitab-a korišćenjem jednostavne linearne regresije na podatke u prethodnoj tabeli. Procenjena regresiona jednačina je:

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

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

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 identifikovati drugu nezavisnu varijablu, menadžeri su smatrali da broj dostava takođe može uticati na ukupno vreme putovanja. Podaci Frigo transportne kompanije, s dodanim brojem dostava, prikazani su na slici 5.12. - ($x_1$) i broj isporuka ($x_2$), kao nezavisne varijable. Procenjena regresiona jednačina 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 transportnekompanije i nezavisne varijable.**

Pogledajmo pažljivije vrednosti $b_1$= 0,0611 i $b_2$= 0,923 u poslednjoj jednačini.

## Napomena o tumačenju koeficijenata

U ovom trenutku možemo dati jedan komentar o odnosu između procenjene regresione jednačine sa samo pređenim kilometrima kao nezavisnom varijablom i jednačine koja uključuje broj isporuka kao drugu nezavisnu varijablu. Vrednost $b_1$ nije ista u oba slučaja. U jednostavnoj linearnoj regresiji tumačimo $b_1$ kao procenu promene $y$ za jednu jediničnu promenu nezavisne varijable. U višestrukoj regresionoj analizi ovo tumačenje treba malo modifikovati. To jest, u višestrukoj regresionoj analizi, svaki regresioni koeficijent se tumači na sledeći način: predstavljao bi procenu promene u $y$ koja odgovara promeni u $x_i$ za jednu jedinicu kada se sve ostale nezavisne varijable konstantne.

U slučaju Frigo transportne kompanije, to uključuje dve 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 |

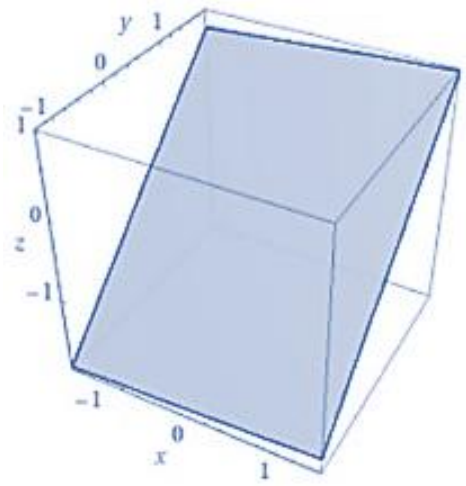
S = 0.573142 R-Sq = 90.4% R-Sq(adj) = 87.6%

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 kompaniju s dve nezavisne varijable.**

Stoga je 0,0611 sati procena očekivanog povećanja vremena putovanja koje odgovara povećanju od jednog kilometra pređene udaljenosti kada je broj dostava konstantan. Slično, budući da je $b_2$ = 0,923, procena očekivanog povećanja vremena putovanja koje odgovara povećanju jedne isporuke kada je broj pređenih kilometara konstantan iznosi 0,923 sata.




**Slika 4.20 Vizualni prikaz rezultata za Frigo transportnu kompaniju.**

# 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 operaciona istraživanja

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

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

Koristeći tehnike iz drugih matematičkih nauka, kao što su modeliranje, statistika i optimizacija, operaciona istraživanja dolaze do optimalnih ili gotovo optimalnih rešenja za probleme donošenja odluka. Zbog svog naglaska na praktičnim primenama, operaciona istraživanja preklapaju se s mnogim drugim disciplinama, posebno industrijskim inženjerstvom i logistikom, što ga čini sastavnim delom njihovih sistema 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 sistema. Karakteriše ih varijabilnost, međuzavisnost komponenti te strukturna i bihevioralna složenost. Da bi se upravljalo ovim karakteristikama, osmišljen je širok niz metoda za njihovo prikladno rešavanje kao celine. Koriste se u strateškom planiranju za:

- omogućavanje složene šta – ako analize;
- upravljanje složenošću: međuzavisnost + varijabilnost + dinamika;
- uključivanje manje troškova i smetnji u procesu nego eksperimentisanje sa stvarnim sistemom;
- fokus na detalje;

- poboljšanje razumevanja sistema;
- poboljšanje komunikacije između menadžmenta i stručnjaka.

Strateško logističko planiranje (slika 6.1) obuhvata sve aktivnosti koje je potrebno izvesti na strateškom, taktičkom i operativnom nivou kako bi se osiguralo upravljanje potpunim kvalitetom (engl. *Total Quality Management* - TQM) (Ciampa, 1992) i proizvodnja tačno na


**Slika 4.21 Strateško logističko planiranje.**

vreme ( engl. *Just in Time* - JIT) (Britannica, 2023).

## 6.2 Šest sigma

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

SCOR proces planiranja obuhvata sve aktivnosti povezane s razvojem planova za upravljanje i poboljšanje lanca snabdevanja. Neprekidni napori da se postignu stabilni i predvidljivi

rezultati procesa smanjenjem varijacija procesa (6-sigma) od vitalne su važnosti za poslovni uspeh.

### DMAIC i DMADV

Proizvodni procesi (*sourcing*, proizvodnja, isporuka, omogućavanje, kao i upravljanje povratima) imaju karakteristike koje se mogu definisati, meriti, analizirati, poboljšati i kontrolisati. Stoga ove faze čine metodologiju upravljanja proizvodnim procesom, skraćeno DMAIC.

Neki praktičari kombinovali 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 promovišu poslovnu i operativnu izvrsnost.

Metodologija DMADV (definiši, meri, analiziraj, dizajniraj i proveri), takođe poznata kao DFSS (engl. *Design for Six Sigma*, tj. Dizajn za šest sigma), u skladu je s KBE (engl. *Knowledge Based Engineering*, tj. Inženjering zasnovan na znanju). Faze DFSS metodologije (Chowdhury, 2002) su:

1. Definišite ciljeve dizajna koji su u skladu sa zahtevima kupaca i strategijom preduzeća.
2. Merite i identifikujte karakteristike koje su ključne za kvalitet (engl. *Critical to Quality* - CTQ), merite mogućnosti proizvoda, kapacitet proizvodnog procesa i merite rizike.
3. Analizirajte kako biste razvili i dizajnirali alternative.
4. Dizajnirajte poboljšanu alternativu, najprikladniju za analizu u prethodnom koraku.
5. Proverite dizajn, postavite probne radove, implementirajte proizvodni proces i predajte ga vlasniku (vlasnicima) procesa.

Projekti poboljšanja poslovanja *Six Sigma* (Tennant, 2001), inspirisani ciklusom „Planiraj–Radi–Proučavaj–Deluj" (engl. Plan–Do–Study–Act) W. Edwardsa Deminga (Tague, 2005), zavisno od svoje prirode, slede jednu od gore navedenih metodologija, a svaka ima pet faza:

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

## 6.3 Poslovna inteligencija

Poslovna inteligencija (engl. *Business Intelligence* - BI) obuhvata sve strategije i tehnologije koje preduzeća koriste za analizu podataka o prošlim i trenutnim poslovnim informacijama (Tableau, 2023.). Podržavaju je sistemi upravljanja znanjem (KMS) koji predstavljaju deo logističkih informacionih sistema (engl. *Logistics Information Systems* - LIS) a koji stručnjacima iz različitih područja i na različitim nivoima menadžmenta omogućavaju savetovanje i podršku.

### Poslovna analitika

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

BA proces sadrži sledeće:

1. Agregacija podataka: pre analize, podaci se prvo moraju prikupiti, organizovati i filtrirati, bilo pomoću dobrovoljnih podataka ili transakcijskih zapisa.
2. Rudarenje podataka: razvrstava velike skupove podataka koristeći baze podataka, statistiku i mašinsko učenje za prepoznavanje trendova i uspostavljanje odnosa.
3. Identifikacija pridruživanja i redosleda: identifikacija predvidljivih akcija koje se izvode zajedno s drugim akcijama ili slede jedna drugu.
4. Rudarenje tekstualnih podataka: istražuje i organizuje velike, nestrukturirane tekstualne skupove podataka u svrhu kvalitativne i kvantitativne analize.

5. Predviđanje: analizira istorijske podatke iz određenog perioda kako bi se napravile informisane procene koje su prediktivne u određivanju budućih događaja ili ponašanja.
6. Prediktivna analitika: prediktivna poslovna analitika koristi različite statističke tehnike za formiranje prediktivnih modela, koji izvlače informacije iz skupova podataka, identifikuju obrasce i daju prediktivnu ocenu za niz organizacionih ishoda.
7. Optimizacija: nakon što se identifikuju trendovi i naprave predviđanja, kompanije mogu koristiti simulacione tehnike 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 procesa (engl. *Sales and operations planning* - SOP) je fleksibilan alat za predviđanje i planiranje proizvodnih aktivnosti. SOP koraci:

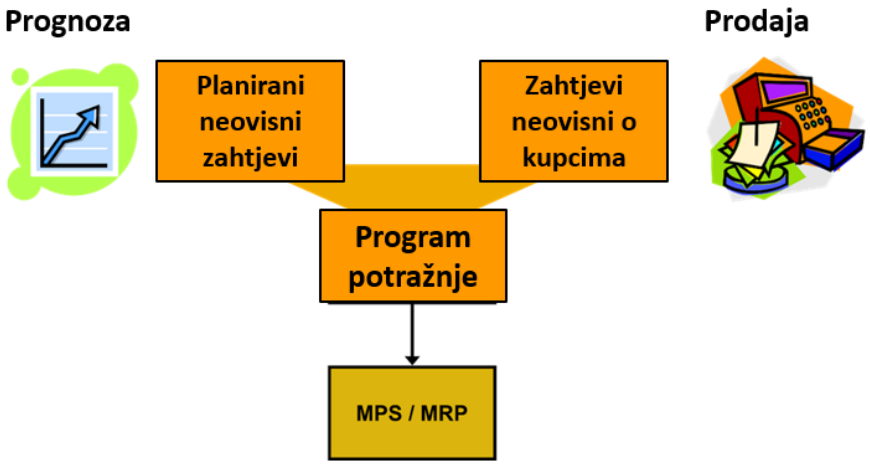
1. plan prodaje,
2. plan proizvodnje i
3. planiranje kapaciteta.

SOP radi na podacima iz različitih izvora informacija u celoj kompaniji: prodaja, marketing, proizvodnja, računovodstvo, ljudski resursi i nabavkaa. Obično ih osiguravaju odgovarajuće službe pomoću sistema za planiranje resursa poduzeća (engl. *enterprise resource planning* - ERP).

Dok SOP deluje na strateškom nivou, ERP deluje na taktičkom nivou logističkog informacionog sistema kompanije. Njima se pridružuje i *Demand Management* (upravljanje potražnjom) program koji povezuje strateško planiranje prodaje i procesa (SOP) i detaljno planiranje proizvodnje (*Master Production Scheduling* - glavno planiranje proizvodnje) i *Material Requirements Planning* - planiranje potreba za materijalima) na operativnom nivou. Ovde prethodno spomenuto planiranje i simulacija stupaju na scenu kako bi se napravio izvodljiv i optimalni plan proizvodnje.

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

1. planirani nezavisni zahtevi (engl. *planned independent requirements* - PIR) od projektovanih količina prodaje na osnovu marketinga i
2. zahtevi nezavisni od kupaca (engl. *customer independent requirements* - CIR) iz podataka na osnovu postojećih i planiranih prodajnih naloga.




**Slika 4.22Upravljanje potražnjom.**

## Primer SOP-a

Poduzeće u svojoj prodajnoj mreži trguje s više vrsta proizvoda. Prodajne transakcije unose se i čuvaju centralno, kako bi se moglo pratiti stanje zaliha, kao i izvršiti analiza prodaje za upravljanje potražnjom. Zapisuju se u CSV (engl. *Comma Separated Values*) formatu, koji je lako obraditi u ERP sistemu kompanije, kao i u službi za analitiku, koja koristi proračunske tabele.

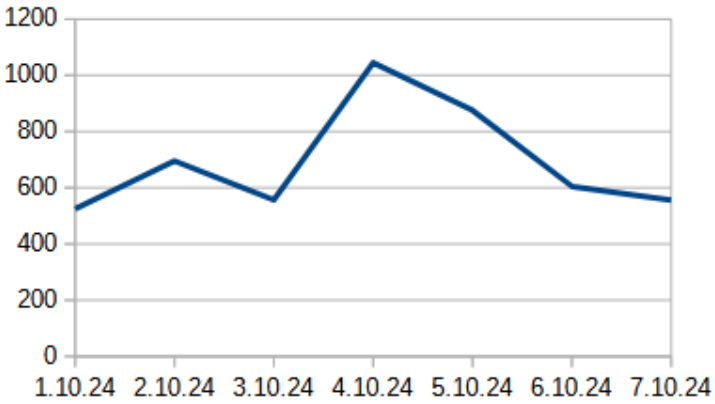
U analizi prodaje, prodajne transakcije se inicijalno filtriraju kako bi se utvrdilo jesu li potpune i ispravno formatirane. Tek tada su spremni za statističku procenu, budući da bi u protivnom nedostajući ili loše oblikovani podaci mogli rezultovati pogrešnim tumačenjem rezultata.

| 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 nedeljnoj prodaji.**

Uglavnom se analitikom omogućavaju različiti senzibilni uvidi u prikupljene "sirove podatke" (slika 6.3). To se može postići pivot tabelama koje omogućavaju grupisanje podataka prema odabranim atributima i sprovođenje statističke analize. U principu, bilo koji atribut (kolona) ulaznih podataka može se smatrati pivotom. Stoga često govorimo o više dimenzionoj "kocki podataka". Budući da ne možemo grafički prikazati više od dve ili tri dimenzije, najjednostavniji, ali obično najkorisniji, prikazi ulaznih podataka formiraju se od dva ili tri pivot atributa. Na osnovu zadatog uzorka podataka, u nastavku su navedeni neki primeri pivot tabela.

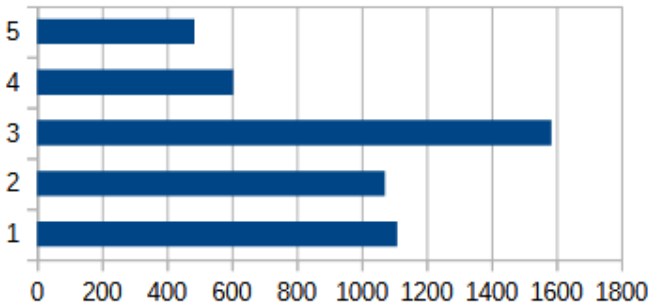
| Date | Vsota - Product Price |
|---|---|
| 1.10.24 | 525,00 € |
| 2.10.24 | 695,00 € |
| 3.10.24 | 556,00 € |
| 4.10.24 | 1.045,00 € |
| 5.10.24 | 875,00 € |
| 6.10.24 | 605,00 € |
| 7.10.24 | 556,00 € |
| **Skupaj Rezul** | **4.857,00 €** |




**Slika 4.24Statistika prodaje po radnim danima.**

Statistika prodaje po danima u nedelji ili mesecu (slika 6.4) omogućava uvid u sezonska kretanja.

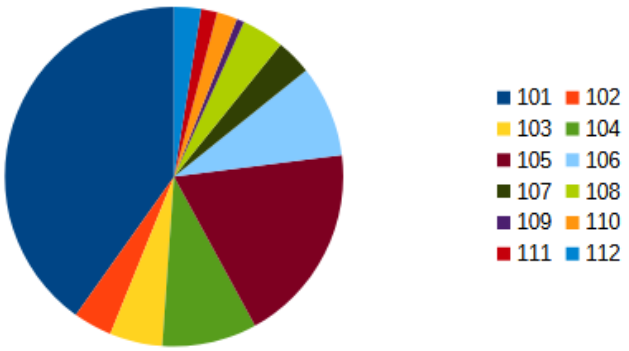
| Seller ID | Vsota - Product Price |
|---|---|
| 1 | 1.110,00 € |
| 2 | 1.072,00 € |
| 3 | 1.585,00 € |
| 4 | 605,00 € |
| 5 | 485,00 € |
| **Skupaj Rezul** | **4.857,00 €** |




**Slika 4.25Statistika prodaje po prodajnim mestima.**

Statistika prodaje po prodajnom mestu (slika 6.5) utvrđuje koji su prodajna mesta najzaposlenija 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 udeo u portfoliu.

| 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 mestu, transakciji i kupcu nudi strukturirani uvid u podatke koji je koristan prilikom rešavanja upita za određeni proizvod ili prodajnu transakciju.

## 6.4 Sistemi za podršku odlučivanju

Sistemi za podršku odlučivanju (engl. *Decision Support System* - DSS) su interaktivni sistemi koji pomažu donosiocima odluka da koriste podatke i modele za rešavanje nestrukturiranih ili delomično strukturiranih problema. Ekspertski sistem (ES) je aplikacioni program ili okruženje, koje uspešno podržava rešavanje problema u specijalizovanom problemskom području, zahtevajući stručno znanje i veštine.

Uz statističke i simulacione metode koje se koriste u BA, DSS često uključuju modeli koji omogućavaju donošenje odluka, na osnovu skupa razlikovnih kriterijuma. Ovi modeli mogu biti jednostavni (npr. tabele odlučivanja, stabla itd.) koji vode do jednog ispravnog rešenja. S

druge strane, kada se radi o višestrukim (moguće suprotstavljenim) kriterijumima i višestrukim rešenjima, nužan je razrađeniji model. Prikladan model koji se često koristi u profesionalnom i privatnom životu je višekriterijumski model odlučivanja.

## Višekriterijumsko odlučivanje

Višekriterijumsko odlučivanje (engl. *multi criteria decision making* - MCDM) ili analiza višekriterijumskog odlučivanja (engl. multiple-criteria decision analysis - MCDA) je poddisciplina OR-a koja eksplicitno procenjuje više (eventualno) suprotstavljenih kriterijuma u donošenju odluka nad skupom mogućih rešenja ili varijanti.

MCDM model odlučivanja sadrži sledeće:

- Kriterijum – parametri ulaznih varijanti, kritični za naš dizajn.
- Ponderi – relativna važnost odabranih kriterijuma.
- Funkcija korisnosti – funkcija koja kombinuje ponderisane parametre varijanti u upotrebljivu vrednost.
- Podaci – podaci koji predstavljaju varijante; unos podataka u naš MCDM model.

Vrste podataka u MCDM mogu biti:

- Kvantitativni – predstavljaju vrednosti koje se kao takve mogu uporediti.
- Kvalitativni – predstavljanje relativnih uporednih vrednosti (npr. visoka, prijatna, niska temperatura, itd.) koje je potrebno kvantifikovati kako bi se dobile jedinstvene vrednosti.
- Binarni – predstavljanje binarnih kriterijuma; svojstvo ispunjeno (1) ili ne (0).

Postupak višekriterijumskog 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 ocene $p_{i,j}$ za svaki $P_{i,j}$ $(j=1\dots n)$, u odnosu na maksimalnu vrednost $j^{-tog}$ parametra $P_{i,j}$ iz svih i uzoraka:
   a) $p_{i,j} = P_{i,j} / \max \{P_{i,j}\}$ ako je veća vrednost $P_{i,j}$ korisnija.
   b) $p_{i,j} = 1 - P_{i,j} / \max \{P_{i,j}\}$ ako je manja vrednost Pi,j korisnija.

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

4. Ponderisane ojene svih varijanti se sabiraju: $X_i = \Sigma x_{i,j}$ za svaki i=1...m kako bi se dobile kompozitne ocene prema našoj funkciji korisnosti.

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

## Primer MCDM-a

Prilikom izbora nove opreme u preduzeću često moramo sprovoditi višekriterijumsko odlučivanje. Razmotrimo primer izbora najisplativije (ispod 300 EUR) mobilne platforme s operativnim sistemom Android za našu kompaniju. U tabeli 6.1 navedeni su modeli koji su izabrani na osnovu upita među zaposlenima kako bi suzili naš asortiman. Za svaki od njih navedeni su parametri koji su odabrani kao najrelevantniji. U nastavku se podaci odabranih parametara normalizuju kako bi se dobile uporedive vrednosti, ponderišu se kako bi se naglasile značajnije vrednosti i sabiraju se kako bi se dobile ocene za odabrane uzorke.

**Tabela 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 % |

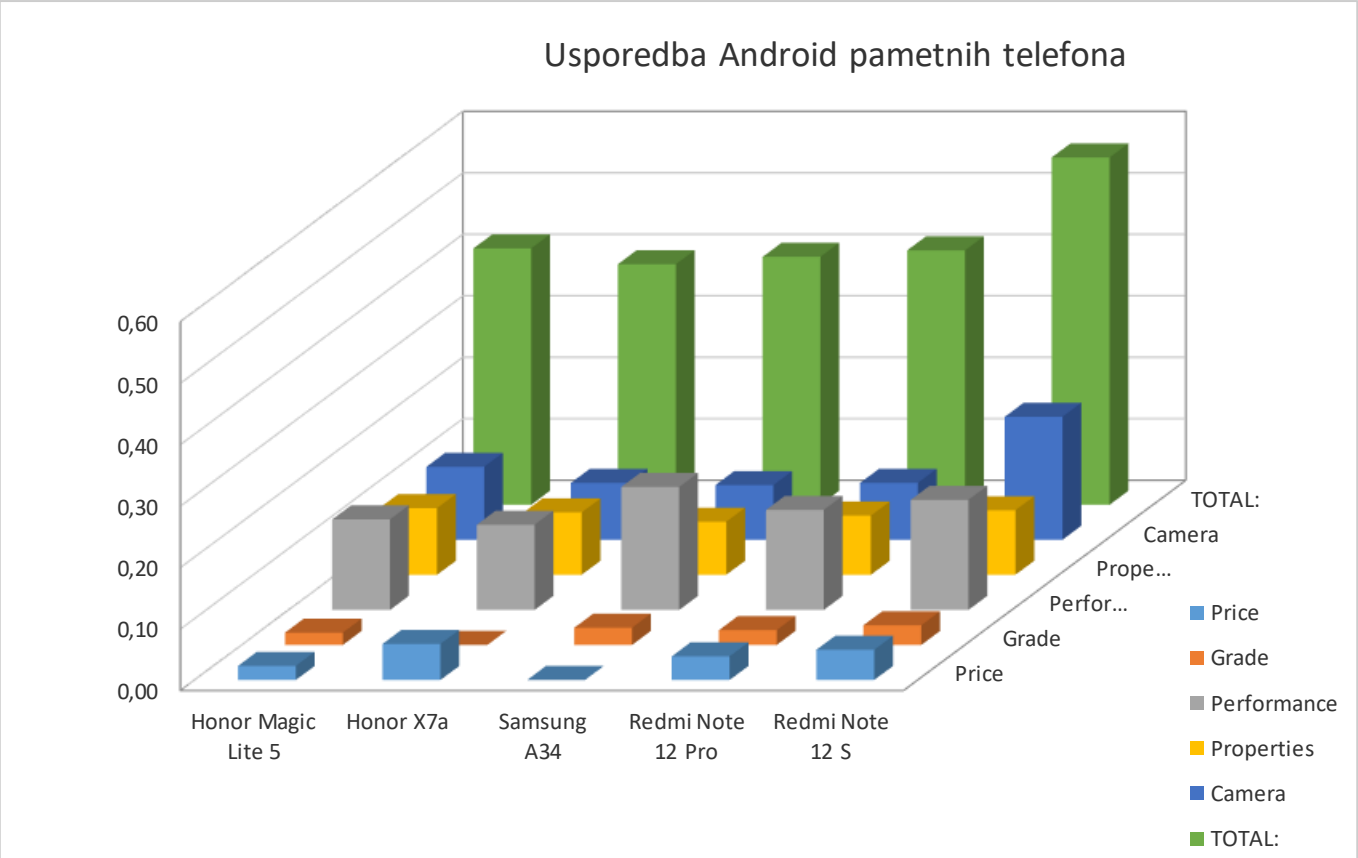
| NORMALIZED PARAMETERS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Price (€) | Grade* | Performance | | | Properties | | | Camera |
| Model | | (1-10) | proc.speed (GHz) | RAM (GB) | int.mem. (GB) | weight (g) | size (mm3) | bat.cap. (mAh) | (MP) |
| Honor Magic Lite 5 | 0,11 | 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 tabela (tabela 6.2) i eventualno grafikon (slika 6.8) koji ukratko prikazuju proces donošenja odluka i prezentuju najbolji izbor kao i prednosti i slabosti pojedinih varijanti. Često najbolje ocenjeni primerak nije onaj koji se najbolje

pokazao u svim kategorijama, već onaj koji u proseku najbolje odgovara našim kriterijumima izbora, kao i odmeravanju parametara. To je ujedno i glavna snaga metode MCDM, budući da bi nas izbor prema bilo kojem pojedinačnom parametru mogao dovesti u zabludu.

**Tabela 4.2MCDM sažetak za naš primer izbora mobilne platforme.**




**Slika 4.28 Izbor najbolje 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 |

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

## 6.5 Inženjerstvo zasnovano na znanju

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

### Životni ciklus upravljanja iskustvom

Potreba za prikupljanjem, upravljanjem i korišćenjem 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. Identifikovati: uočiti neusaglašenost sa željenim stanjem koja se pojavljuje u procesu proizvodnje zbog loše definisanog proizvoda ili procesa.
2. Uhvatite: iskustvo sa svojim svojstvima je uhvaćeno.
3. Analizirati: napravljena je analiza osnovnog uzroka snimljenog iskustva kako bi se identifikovala odgovarajuća strategija lijeka i njezina ponovna uporaba kako bi se spriječile ponovne anomalije.
4. Sačuvati: uvidi iz analize arhiviraju se s iskustvom.
5. Tražiti i pronaći: iskustvo se traži i pronalazi.
6. Upotrebiti: koristi se element iskustva.
7. Ponovno korišćenje: zaključuje se ciklus upravljanja znanjem i započinje novi.

KBE je uopšteno dopunjeno drugim disciplinama, čije bliže razmatranje prelazi opseg ovog poglavlja:

- Računarima potpomognuto upravljanje projektima (PS).
- Računarima potpomognuto projektovanje (CAD), proizvodnja (CAM) i robotika (CIM).
- Računarsko simulaciono modeliranje i analiza (SMA).

- Računarima potpomognuto detaljno planiranje proizvodnje (MPS/MRP).

## 6.6 Zaključak

Kao što je predstavljeno u ovom poglavlju, glavne primene BI-a u korporativnom upravljanju odnose se na poslovnu analitiku (BA) i sisteme za podršku odlučivanju (DSS). Obično se nazivaju operacionim istraživanjima (OR). Uz osnovne BI tehnike, opisane u ovom poglavlju, u poglavljima 3 i 4 o upravljanju podacima i simulacionom modeliranju i analizi (SMA) data su neka dodatna razmatranja o prikupljanju podataka, manipulisanju i prezentiranju, koja takođe podržavaju donošenje odluka. Ukratko, BI aplikacije nalaze se u sistemima upravljanja znanjem (KMS), koji se sastoje od:

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

Na osnovu BA, DSS i SMA rezultata osmišljena su iskustva koja unapređuju institucionalno znanje i konstituišu njihove ekspertske sisteme zasnovane na znanju (KBS). Kao što je pokazano u Gumzej i dr. (2023), KBE ih može upotrebiti za uvođenje načela „naučenih lekcija" u upravljanje poboljšanjima preduzeća 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.

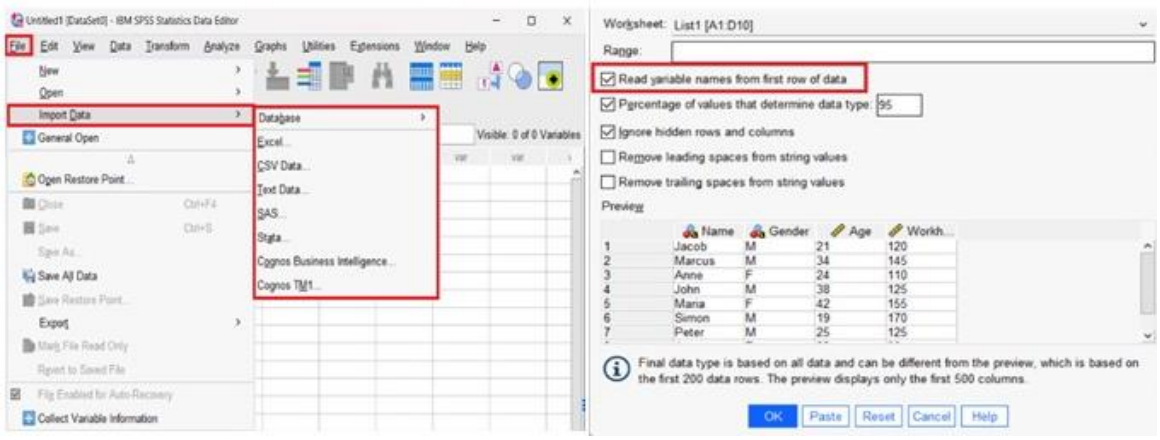
# 7. Statistička obrada podataka SPSS

Do sada ste već stekli osnovno razumevanje statistike, manipulacije podacima, simulacije, modeliranja i analize unutar logističkih lanaca snabdevanja, zajedno s direktnim metodama linearne regresije. Dok statistika nudi širok raspon modela i tehnika za poboljšanje vaših napora optimizacije, sprovođenje analize i prepoznavanje potencijalnih poboljšanja, možda ste primetili da kako složenost analiziranih podataka i proračuna raste, tradicionalni pristupi mogu postati sve zamršeniji i izazovniji za izračunavanje. Kako zamršenost podataka i izračunavanja raste, konvencionalne metode mogu biti nedovoljne i, u nekim slučajevima, ugroziti pouzdanost rezultata. Kako bi se to prepoznalo, statistika koristi različite softverske programe koji automatizuju 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 pružićemo sažeti uvod u primarnu upotrebu softvera SPSS, istražujući njegove funkcionalnosti i praktične primene. Nakon početnog uvoda sledi praktična primena programa kroz četiri bazna testa za izračunavanje rezultata: T-test, korelacije, Hi-kvadrat i ANOVA. Kako bismo vam olakšali učenje, predstavićemo jednostavne probleme i njihova rešenja kako biste se lakše upoznali s ovim testovima.

## 7.1 Osnove IBM-ovog SPSS-a

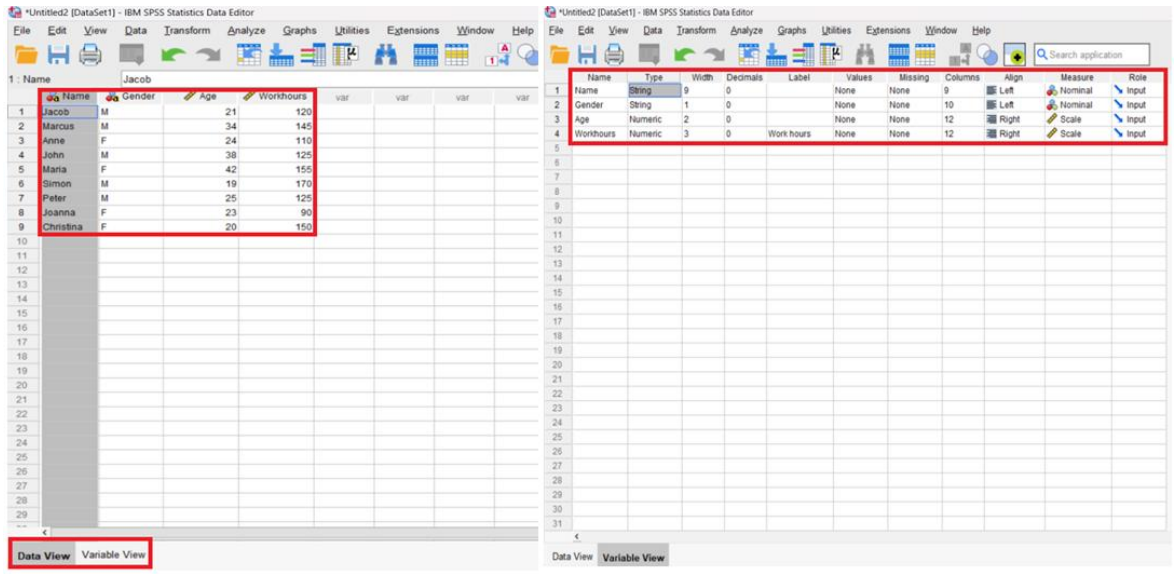
Možda ste već imali iskustva sa SPSS softverom. Međutim, ako vam ipak treba, dopustite da vam ponudimo kratki uvod. SPSS, poput svog opšte priznatijeg pandana Excela, olakšava manipulaciju podacima, analizu i vizualizaciju. Ipak, za razliku od Excela, koji ponekad može biti naporan i složen za programiranje funkcija, SPSS nudi korisnički interfejs 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 konfigurisanje 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 tabeli programa Excel. Srećom, SPSS može uvesti podatke iz različitih formata datoteka u svoje proračunske tabele. Nakon što pripremite finalne Excel proračunske tabele, otvorite softver SPSS. Na početnom ekranu idite na ikonu "Datoteka" i odaberite "Uvezi podatke". U sedećem prozoru možete odabrati format podataka koji nameravate uvesti (pogledajte sliku 7.1). Sledeći ovaj korak, pronađite pripremljenu datoteku, odaberite je i pređite na sedeći prozor. Ovaj prozor će od vas tražiti da konfigurišete dodatne postavke. Ako ste već uključili nazive kolona u prvi red vaših podataka, odaberite opciju "Pročitaj nazive varijabli iz prvog reda podataka" (pogledajte sliku 7.1), a zatim kliknite "Završi" da bi se podaci pojavili u proračunskoj tabeli (IBM, 2021).



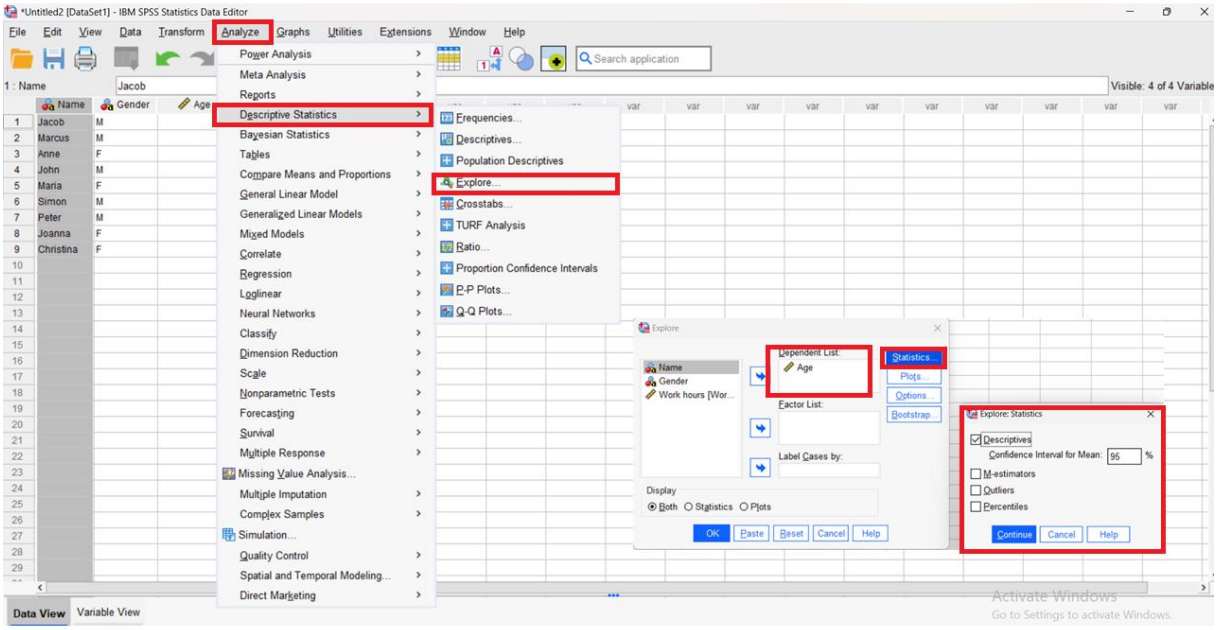

**Slika 4.29SPSS postavke uvoza podataka.**

Sada kada imamo podatke u našoj proračunskoj tabeli, primetićete jasnu razliku u prezentaciji u poređenju s Excelom. SPSS kategorizuje podatke u dve primarne vrste, svaka s dve dodatne podvrste. Kao što je prikazano na slici 7.2, podaci se mogu klasifikovati kao numerički ili kategorički. Numerički podaci sastoje se od brojeva i mogu se kategorizovati kao diskretni (s ograničenim opcijama) ili kontinuirani (nude beskonačne mogućnosti). S druge strane, kategorički podaci sastoje se od reči i mogu se dalje razlikovati kao redni (imaju hijerarhiju) ili nominalni (bez hijerarhije). Zavisno od prirode vaših podataka, možda ćete morati konfigurisati varijable kako bi se uskladile sa željenom analizom. U većini slučajeva, SPSS će automatski prikladno kategorizovati varijable. Pretpostavimo da želite izvršiti dalju manipulaciju tipovima podataka. U tom slučaju možete pristupiti opciji "Prikaz" i pod "Prikaz varijable" prilagoditi varijabilne informacije kao što su naziv, vrsta, širina, mera i više (IBM, 2021.).



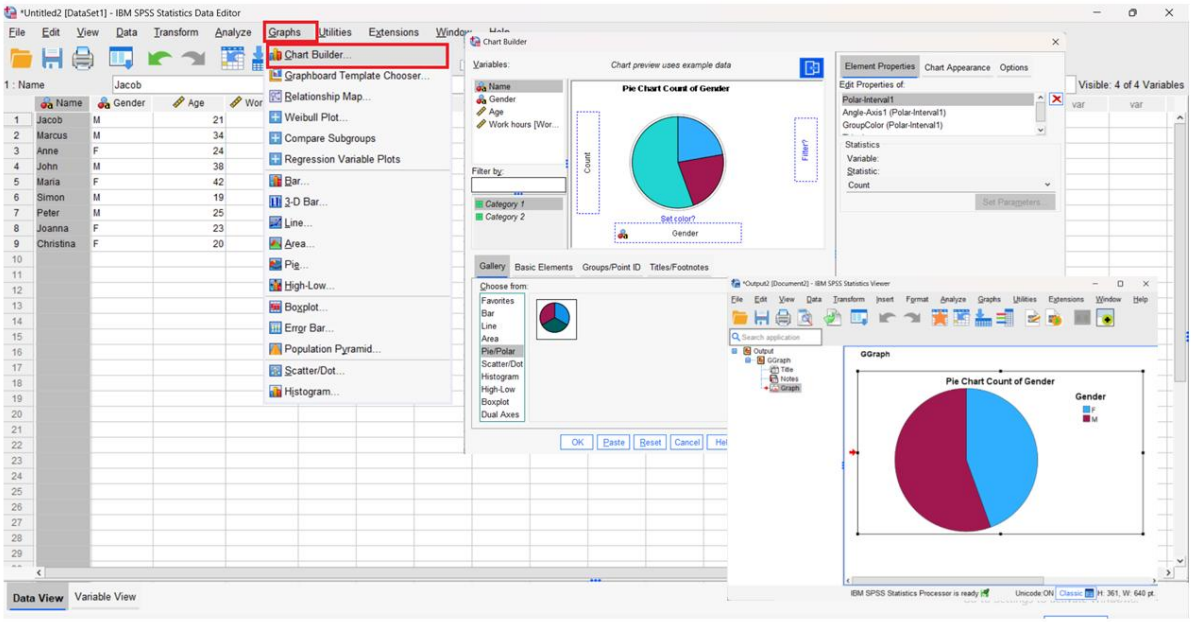

**Slika 4.30Prozori za prikaz podataka i varijabli.**

Nakon što ispravno postavite svoje podatke, možete ih istraživati unutar SPSS-a. SPSS omogućava korisnicima izvođenje detaljne statističke analize bez oslanjanja na unapred definisane funkcije. Na početnom ekranu (pogledajte sliku 7.3), idite na "Analiziraj", nakon čega sledi "Deskriptivna statistika", a zatim odaberite "Istraži". U odjeljku "Istraži" pronaći ćete različite opcije zavisno od karakteristika podataka koje ste uneli. U ovom načinu rada SPSS će vam pružiti informacije o "deskriptivnoj statistici" o vašim podacima. Iako je ovo važno za početnu analizu podataka, nudi samo osnovne uvide i ne ulazi u detaljniju statističku analizu, koja će biti obrađena u narednim poglavljima. Pre nego što nastavimo dalje, takođe ćemo istražiti još jednu funkciju u SPSS-u — vizualizaciju grafikona (IBM, 2021).




**Slika 4.31 Postavke deskriptivne statistike.**

SPSS nudi niz opcija za vizualizaciju podataka, uključujući histograme, kutijaste dijagrame, stubičaste grafikone, dijagrame rasipanja, linijske grafikone, pita grafikone i još mnogo toga. Do ove tačke trebali biste imati osnovno razumevanje o tome što svaka vrsta grafikona predstavlja i kako tumačiti rezultate koje oni pružaju. Stoga ćemo se fokusirati na to kako izraditi te grafikone unutar softvera SPSS. Da biste izradili grafikone, odaberite karticu "Grafikoni" na početnom ekranu, nakon čega sledi "Izrada grafikona". U novom prozoru možete odabrati vrstu grafikona koju želite izraditi i odabrati varijable koje želite uključiti. Nakon izbora "Završi", pojavi će se novi prozor s rezultatima vizualiziranim u odabranom formatu grafikona. U ovom novom prozoru možete aktivno komunicirati s grafikonom, što vam omogućava izmenu varijabilnih boja i fontova, istraživanje distribucija varijabli na grafikonu, i više (IBM, 2021).



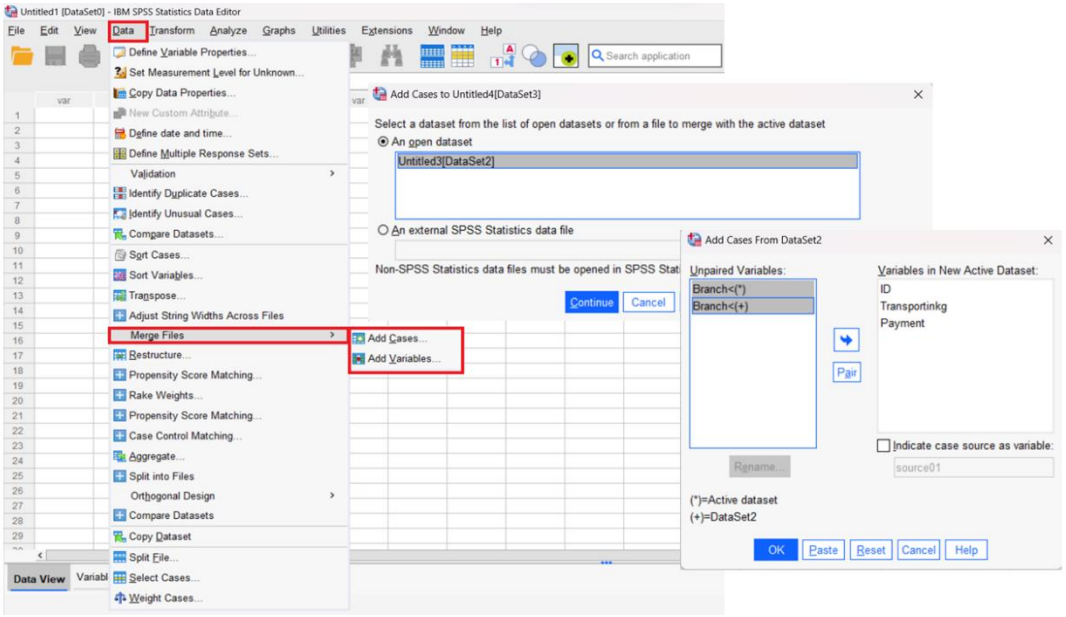

**Slika 4.32Postavke izrade grafikona u SPSS-u.**

Do ove tačke pokrili smo tri od četiri pravila za "Istraživanje podataka", koja uključuju posmatranje podataka (istraživanje neobrađenih podataka), identifikaciju podataka (određivanje tipova podataka) i, u određenoj meri, grafičko prikazivanje i opisivanje podataka putem deskriptivne statistike i izrade grafikona. Poslednje pravilo je "Formulacija pitanja", gde 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 primer, u našem posmatranom primeru, pitanje bi moglo biti: "Da li je naša analizirana populacija pretežno ženska?" Korišćenjem i grafikona i deskriptivne statistike možemo zaključiti da se naša populacija sastoji uglavnom od muških osoba. Kada formulišete svoja pitanja, uvek uzmite u obzir dostupne podatke i varijable koje ste identifikovali (Garth, 2008). Ovime je završen prvi deo SPSS analize podataka, a sada ćemo nastaviti s pripremom testa.

## 7.2 Upravljanje podacima

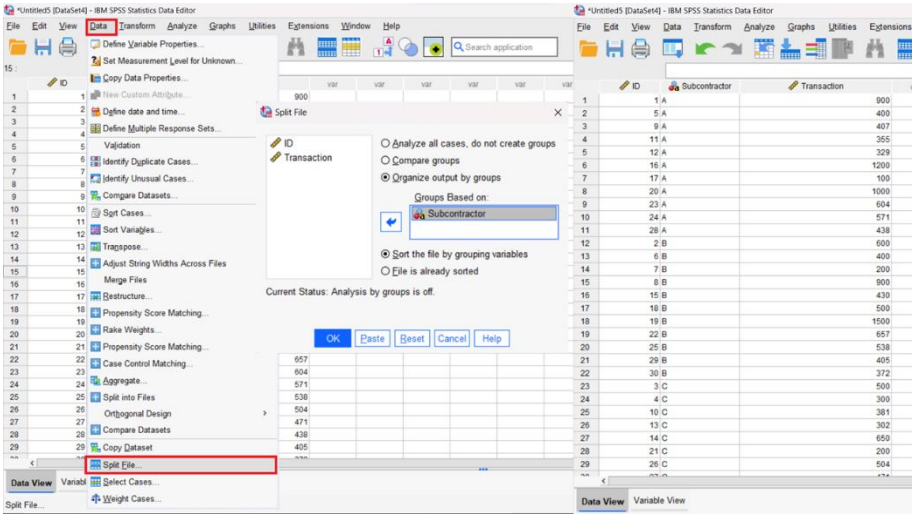
Kada se bavite ključnim podacima u SPSS softveru, postaje ključno razumeti tehnike za manipulisanje informacijama u pojedinačnim aktivnim skupovima podataka. SPSS pruža funkcionalnosti koje olakšavaju manipulaciju postojećim podacima sadržanim u aktivnim

skupovima podataka. Povremeno možete naići na dve baze podataka odvojeno uvezene u skupove podataka, no prednost je da se spoje radi poboljšane analize. Razmotrimo logističku kompaniju s dve podružnice, od kojih svaka daje podatke o troškovima i prevozu tereta u kilogramima. Cilj menadžera je analizirati ukupnu učinkovitost poduzeća. U SPSS-u to uključuje navigaciju "Podaci", izbor "Spoji datoteke" i dve različite opcije. Jedan uključuje izbor "Slučajevi" i određivanje varijable za spajanje, uklanjanje te varijable dok spaja ostale. Alternativno, izborom opcije "Varijabla" zadržava se varijabla u novom skupu podataka. Praktična primena očita je u našem scenariju logistike, gde spajanje skupova podataka pojednostavljuje sveobuhvatnu analizu učinka kompanije.



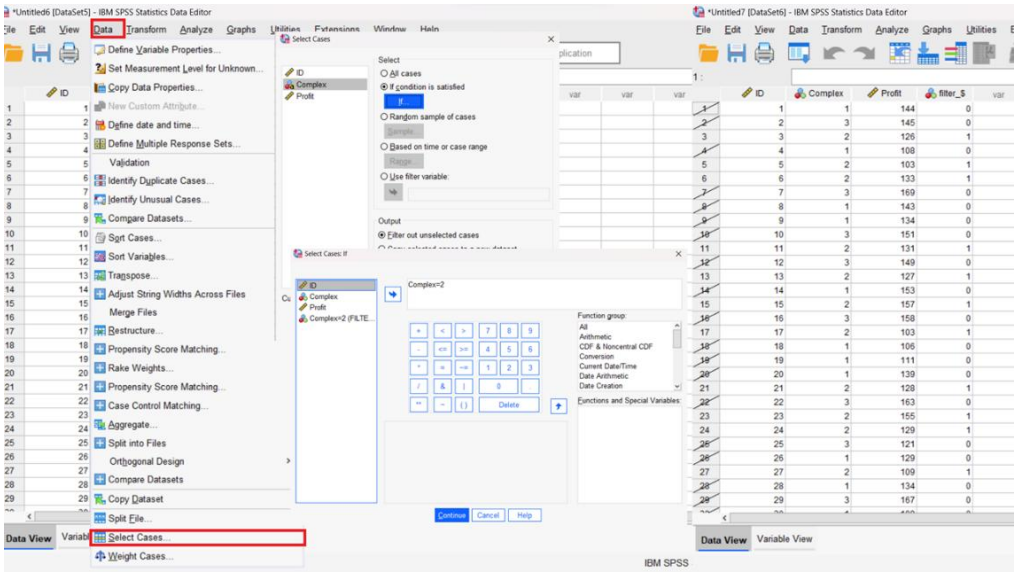

**Slika 4.33Prozor za spajanje datoteka.**

Dok funkcije spajanja i razdvajanja omogućavaju određenu manipulaciju podacima, opcija "Odaberi slučajeve" nudi različite prednosti. Zamislite da imate podatke za prodavnice B, C i D u jednoj bazi podataka, a fokus je isključivo na poređenju prodavnice A i prodavnice C. Izborom "Podataka" i "Odaberi slučajeve" možete odrediti varijable od interesa, uspešno filtrirajući izbaciti neželjene podatke. Na primer, postavljanje Prodavnice C kao 2 upućuje softver da se koncentriše isključivo na Prodavnicu C, generišući izlaz koji je zatim dostupan za naknadne analize, kao što je deskriptivna statistika, fokusirajući se isključivo na odabrane slučajeve. Takav pristup takođe omogućava komparativnu analizu samo između vrednosti Prodavnica A i Prodavnica C.



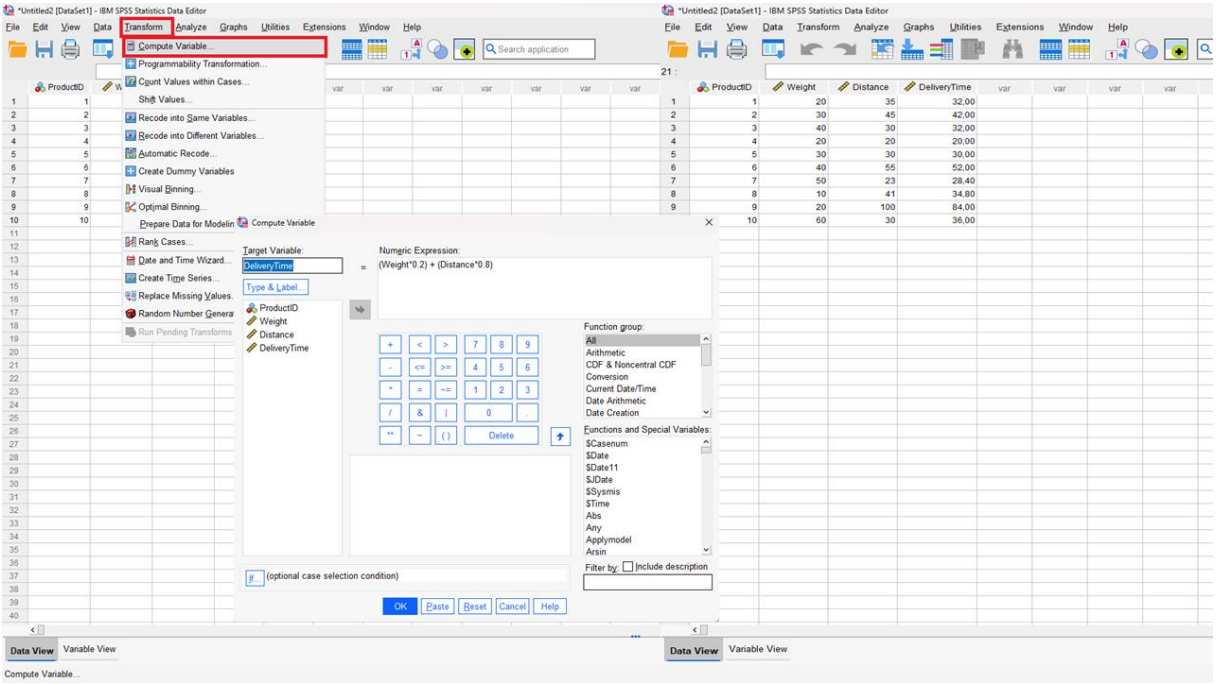
**Slika 4.34Prozor za deljenje datoteke.**

Dok funkcije spajanja i razdvajanja omogućavaju određene manipulacije podacima, postoji i opcija "Odaberi slučajeve". Zamislite da pouzdano znamo da prodavnica A ima u proseku 120 € dobiti i želimo to uporediti s prodavnicom C. Nažalost, u našoj bazi podataka imamo podatke za prodavnice B, C i D u jednoj bazi podataka i analiza bi uključivala podatke iz sve tri prodavnice. Klikom na "Podaci" i "Odaberi slučajeve" možemo odabrati varijablu na koju se želimo fokusirati. U našim slučajevima smo definisali da prodavnica C treba biti postavljena kao 2, a zatim smo kreirali funkciju za softver da se fokusira samo na prodavnicu C. Izlaz se zatim može koristiti za naknadnu analizu izborom ove nove kolone (npr. deskriptivna statistika).



**Slika 4.35Izbor slučaja.**

Povremeno skupovi podataka mogu već sadržavati varijable, ali ipak postoji potreba za uvođenjem novih varijabli na osnovu postojećih. Uzmimo, na primer, menadžera logističke kompanije koji poseduje podatke o težini i pređenoj udaljenosti za razne proizvode, ali zahteva vreme isporuke za optimizaciju ruta. U SPSS-u, da bi se to postiglo, potreban je klik na "Transform", a zatim na "Compute Variables". Nova varijabla, DeliveryTime, stvara se unutar novog prozora postavljanjem numeričkih izraza. U ovom slučaju, dodjeljivanje lestvice od 0,8 za udaljenost i 0,2 za težinu rezultuje novom varijablom koja predstavlja vreme isporuke, što je ključni dodatak skupu podataka. Postoji fleksibilnost izračunavanja dodatnih varijabli, kreiranih za potrebe statističkih testova.



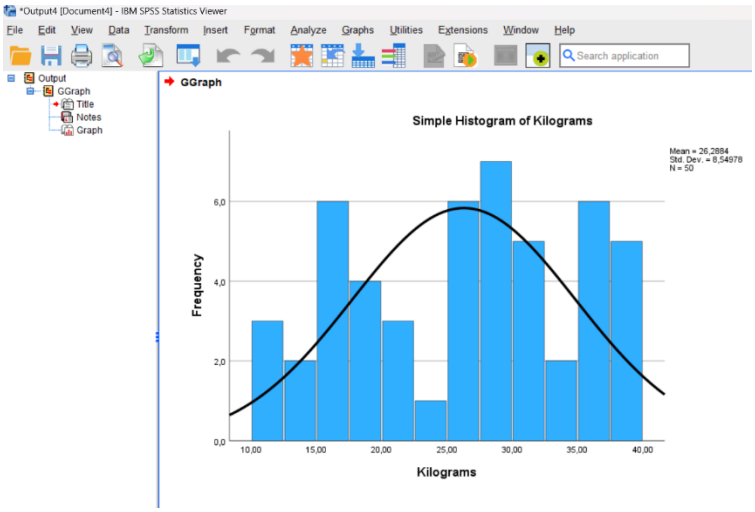

**Slika 4.36Postupak izračunavanja varijabli.**

Na ovaj način se zaključuje mali pregled funkcija upravljanja podacima koje pokriva SPSS, a koje bi mogle biti korisne tokom sledećih testova modela koji su obuhvaćeni u ovom poglavlju. Nastavićemo s fazama koje su potrebne pre nego što možemo sprovesti statistički test u softveru SPSS.

## 7.3 Priprema testa

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

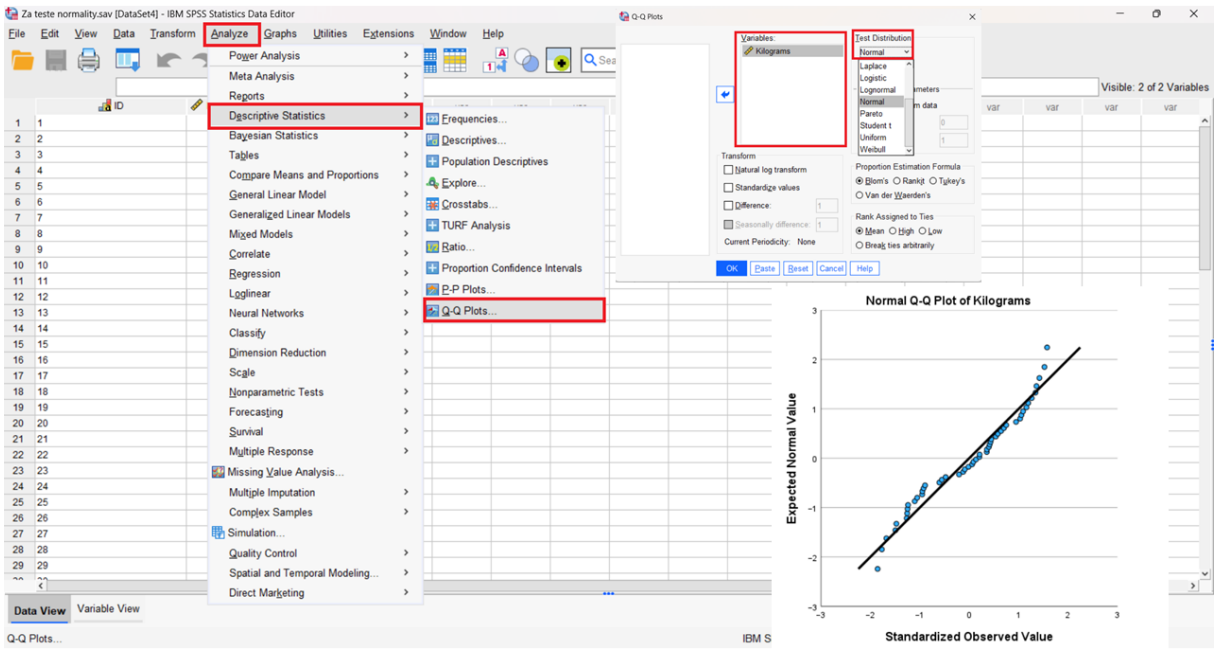
poglavljima 7.1 i 7.2), analizu podataka i interpretaciju rezultata (Garth, 2008; George i Mallery, 2022). U ovom poglavlju naš fokus je na analizi podataka pomoću softvera SPSS. Budući da smo hipoteze već obradili u prethodnim poglavljima, naš primarni fokus biće na sprovođenju testova normalnosti unutar SPSS-a. Postoje tri metode za procenu normalnosti: histogram, QQ-grafikon i test normalnosti. Preporučljivo je upotrebiti najmanje dve, 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 vrednosti varijable liče na zvonastu liniju. Ako su stupci više nagnuti u levu ili desnu stranu, to može značiti eksponencijalnu distribuciju. Na primer, generisali smo bazu podataka od 100 ID-ova, svaki s varijablom koja predstavlja težinu u kilogramima. Prateći uputstva, izradili smo histogram, kao što je prikazano na slici 7.9. Kao što se vidi sa slike, stupci su raspoređeni po grafikonu i iako možda ne odražavaju savršenu krivu liniju, ipak sugerišu normalnu distribuciju i pozitivan rezultat testa (George i Mallery, 2022; Goeman i Solari, 2021).




**Slika 4.37Histogram rezultata testa normalnosti.**

Još jedna opcija za sprovođenje testova normalnosti je QQ-grafikon, koji se može pokrenuti klikom na "Analyze" (hrv. Analiziraj), nakon čega sledi "Descriptive Statistics", a zatim izborom "Q-Q Plots". Prednost ovog pristupa je što omogućava procenu više varijabli istovremeno (Williamson, bd). Test se smatra uspešnim kada se tačke na dijagramu grupišu usko oko prave linije, što predstavlja normalnu distribuciju. Ako tačke formiraju "repove", to

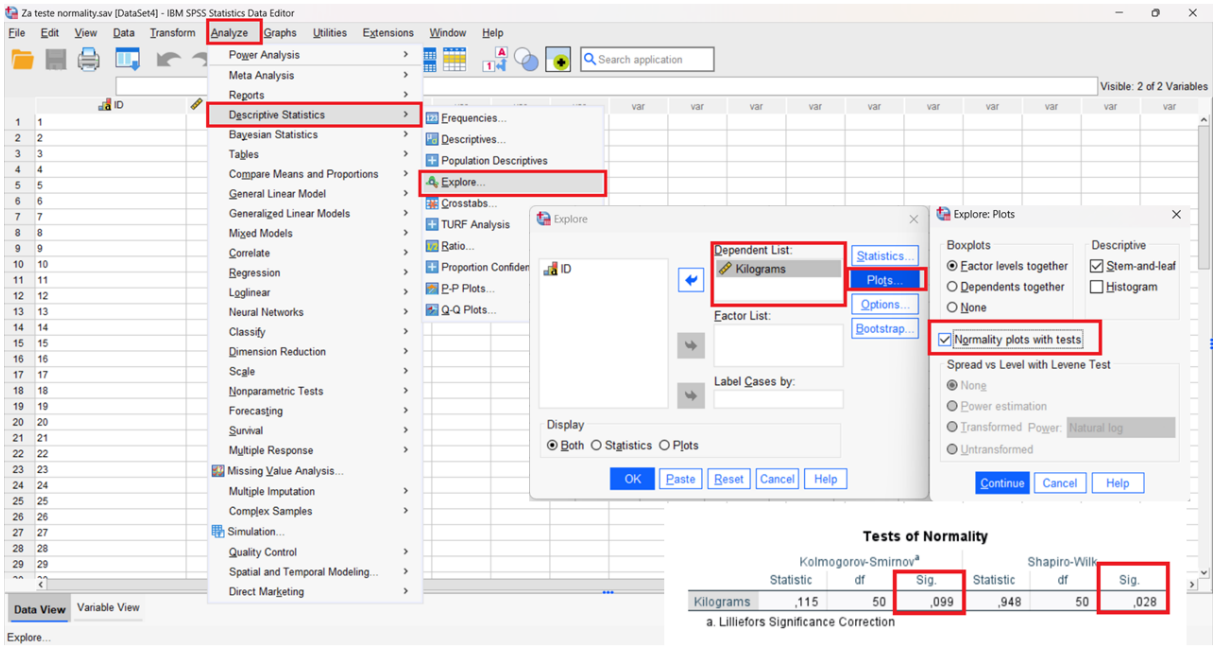
ukazuje na neuspešan test normalnosti (Andersen i Dennison, 2018). Koristeći istu bazu podataka iz testa histogramskog grafikona, sproveli smo QQ grafikon test. Na slici 7.10 u nastavku možete primetiti da je većina tačaka klastera za našu varijablu poravnata s pravom linijom, što ukazuje na normalnu 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 sprovođenje 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 sledi "Descriptive Statistics", a zatim "Explore". Morate postaviti varijable koje želite proveriti ispod okvira "Dependent List" (hrv. Zavisna lista). Zatim pod "Plots" odaberite "Normality Plots with Tests" (hrv. Grafike normalnosti s testovima). Test se smatra uspešnim ako je stupac Sig ($p$ -vrednost) u rezultatima veći od 0,05, što ukazuje na normalnu distribuciju. Ako je $p$ -vrednost manja od 0,05, to ukazuje da distribucija nije normalna i test se smatra neuspešnim. Ovaj test smo 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 $p$ -vrednost veća od 0,05. Međutim, za

Shapiro-Wilkov test, *p* -vrednost je niža, što ukazuje na negativan rezultat testa. Do ovih različitih rezultata dolazi jer oba pristupa imaju različite postavke osetljivosti i snagu u otkrivanju odstupanja (Ghasemi i Zahediasl, 2012.). Budući da smo već sproveli testove QQ dijagrama i histogramskog grafikona, Test normalnosti može se generalno smatrati pozitivnim. Uz potvrđene testove normalnosti, možemo sprovesti glavne testove, kao što je test jednog uzorka.




**Slika 4.39Postavke i rezultati testa normalnosti.**

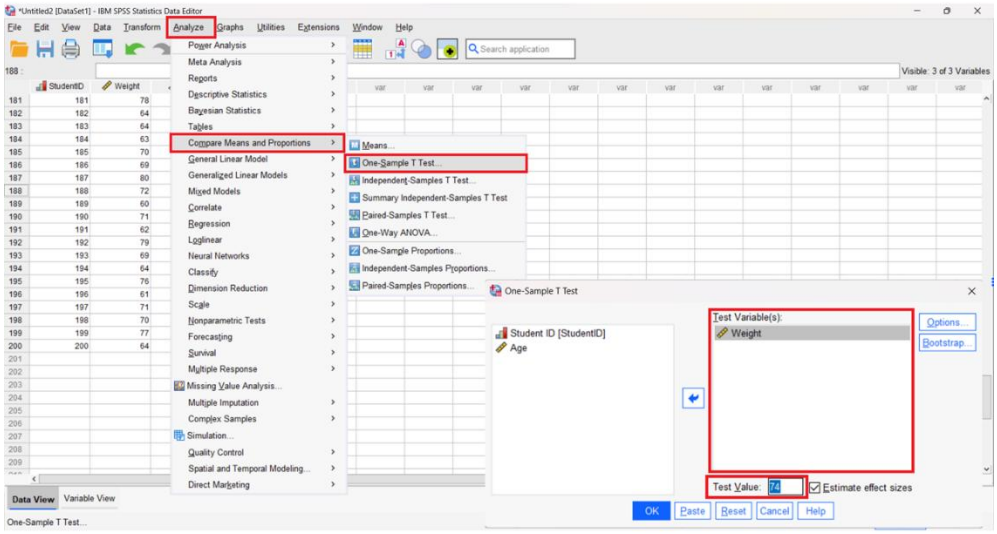
## 7.4 T-test jednog uzorka

Već je obrađena teorija za T-test jednog uzorka u prethodnim poglavljima, stoga ćemo se prvenstveno fokusirati na sprovođenje 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 starost) (Kim, 2015.). Prateći uputstva iz prethodnih potpoglavlja sprovodimo sledeće korake:

- Istražite podatke, tačnije naše **varijable** i **deskriptivnu statistiku** i postavite naše **pitanje**.
- Proverite **normalnost**, budući da bi samo jedan varijabilni histogram i QQ grafikon trebali biti dovoljni.

- Postavite hipotezu, gde se za **nultu -** varijabla ne razlikuje od određene vrednosti i **alternativu** gde je drugačija.
- Uradite **Studentov T-test** .
- Tumačite rezultate, fokusirajući se na to da li je **nulta hipoteza odbijena** ili **ne** , odgovorite na pitanje i napišite izveštaj o našem testu.

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




**Slika 4.40Postavke T-testa jednog uzorka.**

Nakon potvrde testa, pojaviće se drugi prozor s rezultatima naše analize (pogledajte sliku 7.13). Ovaj prozor pruža nekoliko informacija u vezi s našom analizom. U ovom slučaju su obe *p* -vrednosti niže od 0,05, što ukazuje na značajnost testa. Dodatno, proveravamo vrednosti *t* i *df* , koje su u našem slučaju -9,806 odnosno 199. Iz ovih rezultata možemo zaključiti da je naša nulta hipoteza odbačena. Stoga, izveštaj o rezultatima glasi: "Prosečna

težina studenta značajno je niža (srednja vrednost = 69,63) od vrednosti od 74 kg (t-test jednog uzorka, $t$ = -9,806, $df$ = 199, $p$ -vrednost < 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 | Upper |
|---|---|---|---|---|---|---|---|
| Weight | -9,806 | 199 | <,001 | <,001 | -4,370 | -5,25 | -3,49 |

**One-Sample Effect Sizes**

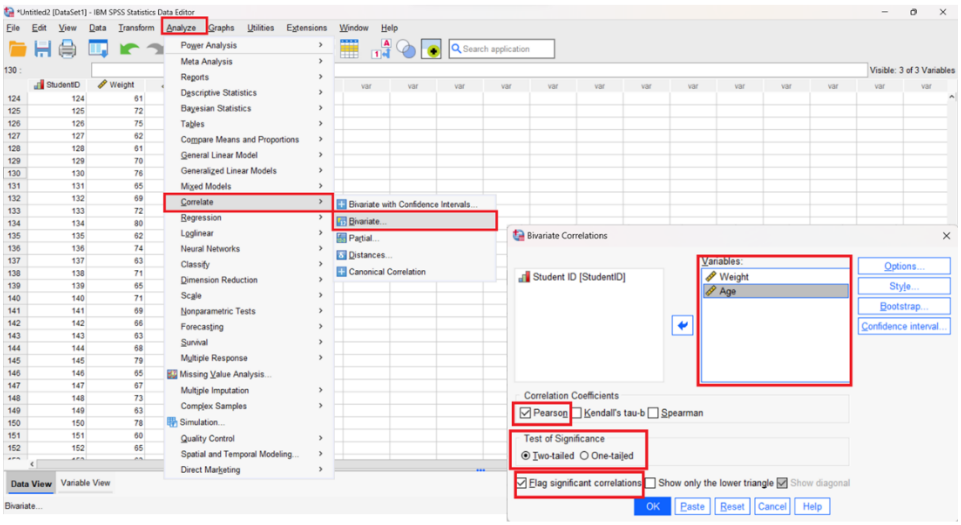
| | | Standardizer[a] | Point Estimate | 95% Confidence Interval Lower | Upper |
|---|---|---|---|---|---|
| Weight | Cohen's d | 6,303 | -,693 | -,847 | -,538 |
| | Hedges' correction | 6,326 | -,691 | -,844 | -,536 |

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

Pređimo sada na drugi test, a to je test korelacije. Uradićemo ga koristeći istu bazu podataka kao u primeru t-testa s jednim uzorkom. Slično t-testu s jednim uzorkom, pratićemo postupak uz nekoliko izmena. Kada se vrši korelacija između dve varijable, važno je odrediti koja je zavisna, a koja nezavisna varijabla (Janse i dr., 2021.; Mishra i dr., 2019). Ovaj izbor možete napraviti na osnovu vašeg istraživačkog pitanja. U našem slučaju želimo istražiti "Postoji li korelacija između starosti studenta i njegove težine?". Nakon pitanja, težinu smatramo zavisnom varijablom, a starost nezavisnom varijablom, jer želimo istražiti jesu li varijacije u godinama povezane s varijacijama u težini. Definišemo naše nulte i alternativne hipoteze (vidi 7.3 i 7.4), a zatim pokrećemo test klikom na "Analyze", nakon čega slede "Correlate" (hrv. Koreliraj) i "Bivariate" (hrv. Bivarijantno). Obe varijable treba staviti u polje "Variable". Proverite jesu li odabrani ili postavljeni "Pearson", "Two-Tailed" i "Flag Significant" (pogledajte sliku 7.14). U ovom slučaju smo 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 umesto Pearsona) (George i Mallery, 2022; McClure, 2005).




**Slika 4.42Postavke testa korelacije.**

Još jednom dobijamo rezultate u novom prozoru (pogledajte sliku 7.15). Iz rezultata možemo videti da je naša Pearsonova korelacija -0,038, a *p* -vrednost 0,596. U korelacionoj analizi, što je vrednost korelacije bliža nuli, to je korelacija između varijabli slabija. U našem slučaju, korelacija je vrlo blizu nule, što ukazuje da nema značajne korelacije između dve varijable (McClure, 2005). Dodatno, visoka *p* -vrednost (0,596) sugeriše da nema značajnih dokaza za zaključak da postoji značajna korelacija između dve odabrane varijable (Williamson, bd). Kao rezultat, naša nulta hipoteza nije odbačena. Na tom osnovu možemo izvestiti da "nije bilo korelacije između starosti i težine studenta".

**Correlations**

**Correlations**

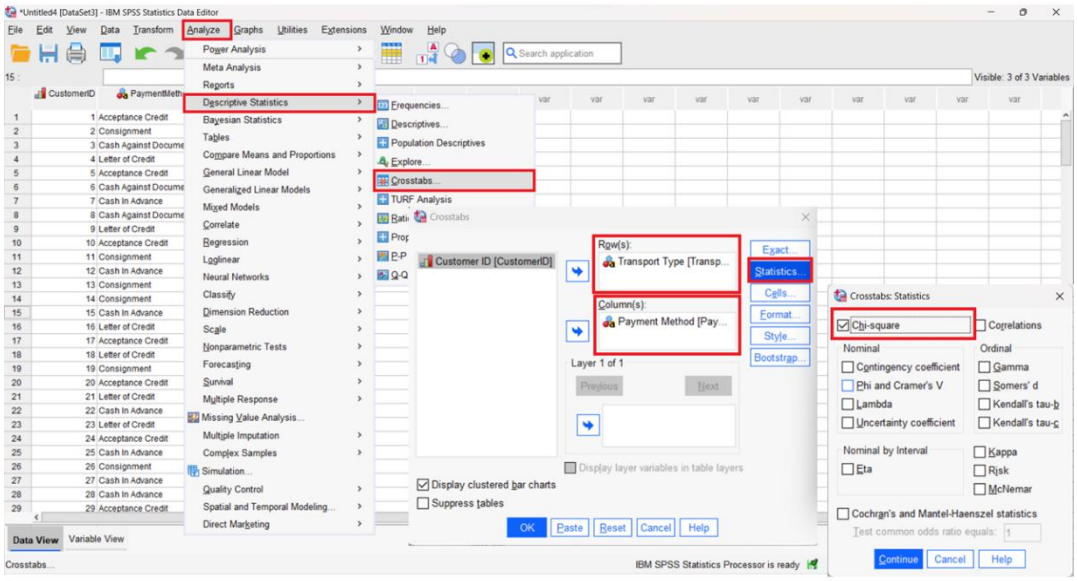
| | | Weight | Age |
|---|---|---|---|
| Weight | Pearson Correlation | 1 | -,038 |
| | Sig. (2-tailed) | | ,596 |
| | N | 200 | 200 |
| Age | Pearson Correlation | -,038 | 1 |
| | Sig. (2-tailed) | ,596 | |
| | N | 200 | 200 |

**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 upoređuje dve kategoričke varijable, a ne numeričke varijable (Turhan, 2020). Kao i postupak u odeljcima 7.4 i 7.5, počinjemo istraživanjem podataka i

formulisanjem istraživačkog pitanja. U našem primeru imamo logističku kompaniju s 200 kupaca, i podatke o vrsti plaćanja i vrsti prevoza koju je svaki kupac odabrao. Pitanje na koje želimo odgovoriti je: "Pokazuju li različite vrste plaćanja različite preferencije za vrste prevoza?" Budući da se radi samo o kategoričkim varijablama, nema potrebe za testom normalnosti. Postavljamo našu nultu hipotezu (preferencije za vrste prevoza iste su za sve vrste plaćanja) i alternativnu hipotezu. Za sprovođenje hi-kvadrat analize kliknite na "Analyze", nakon čega sledi "Descriptive Statistics", i odaberite "Crosstabs" (Unakrsne analize). Ključno je smestiti varijable na osnovu vašeg istraživačkog pitanja u kolone ili redove (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 vrednost iznosi 11,614, *df* vrednost 12, a *p* -vrednost (asimptotska značajnost) 0,477. Na osnovu ovih rezultata možemo zaključiti da ne postoji značajna povezanost između dve varijable, a nulta hipoteza nije odbačena. Stoga sledi izveštaj: "Nema otkrivenih značajnih preferencija između različitih vrsta plaćanja za različite vrste prevoza (dvostrani Chi-Square test, chi-sq = 11,614, *df* = 12, *p* -vrednost = 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**

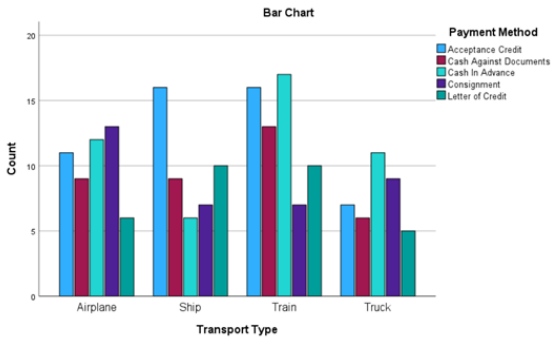
Count

| | | 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 | | |

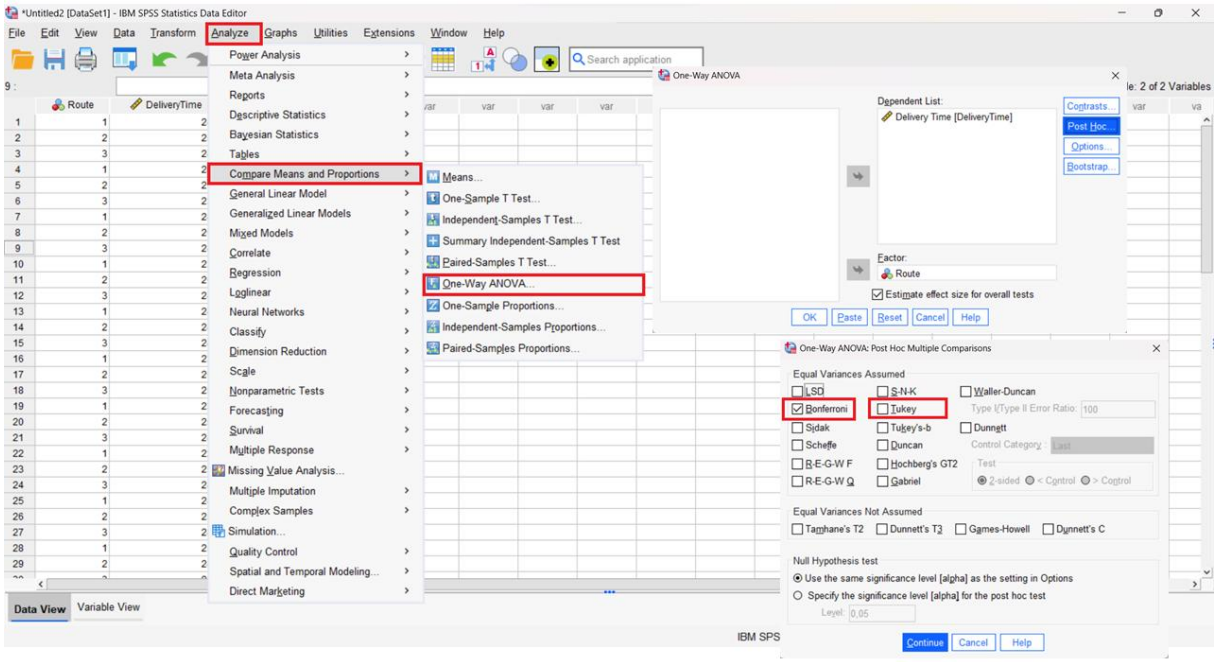
a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,89.




**Slika 4.45Rezultati Hi-kvadrat testa.**

## 7.7 ANOVA

Poslednji test koji ćemo pokriti je ANOVA test, koji se posebno fokusira na jednostavniji model poznat kao jednosmerna ANOVA, koji uključuje kategoričku varijablu i numeričku varijablu (Goeman i Solari, 2021). Kao i kod T-testa, sledićemo isti postupak: istražiti podatke, formulisati istraživačko pitanje, sprovesti test normalnosti i postaviti hipoteze. Razmotrimo studiju slučaja transportnog dispečera koji radi za logističku kompaniju. Dispečer blisko sarađuje s partnerskom kompanijom i redovno planira tri različite rute kamionima za isporuku robe. Zbog politike "Just-in-time" koja naglašava brže isporuke, postavlja se pitanje: "Utiče li izbor rute dostave na vreme isporuke za kompaniju?" Da biste pokrenuli ANOVA test u SPSS-u, idite na "Analyze" nakon čega sledi "Compare Means..." i zatim "One-way ANOVA". Postavite zavisnu varijablu u okvir "Dependent List" (Popis zavisnih), a varijablu Faktor u okvir "Factor" (pogledajte sliku 7.18). Za detaljnu analizu uključili smo i Post Hoc postavku. Važno je napomenuti da se Post Hoc analiza treba sprovesti samo ako je početni ANOVA test pozitivan. Primenom Post Hoc analize možemo identifikovati 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 vrednost 11,173 (više vrednosti ukazuju na više varijacija između grupa) i *p*-vrednost <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 u našem slučaju je takođe valjan (George & Mallery, 2022). Nakon sprovođenja Bonferronijevog testa korekcije, možemo videti da su najbolje *p*-vrednosti zabeležene u slučaju rute 2 (pogledajte sliku 7.19). U izveštaju možemo zaključiti da je „postojala značajna razlika u izboru rute isporuke u korelaciji s vremenom isporuke (1-way ANOVA, $F$ =11,173, $df$ = 47, $p$-vrednost = <0,001). Ruta 2 imala je najbolje rezultate vremena 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 razumevanje da smo u ovom poglavlju pokrili neke od uobičajenih testova. Postoje i drugi testovi, kao što je ANOVA ponovljenih merenja, testovi pouzdanosti i testovi osetljivosti, koji se takođe mogu modelirati i analizirati pomoću softvera SPSS. Ovi dodatni testovi pružaju širi raspon alata za analizu podataka i daju korisne uvide u različita istraživanja i praktične primene.

## 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. Osnove poslovne analitike uključujući R i SQL

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

Na ova i slična pitanja pokušaćemo dati odgovore u sledećem poglavlju.

## 8.1 Šta je poslovna analitika?

BA predstavlja holistički pristup analizi podataka i poslovnom odlučivanju. To je okruženje vođeno podacima s ciljem poboljšanja poslovnih performansi kompanije pružanjem temelja za informisanije donošenje odluka. To je sistemski proces razmišljanja koji primenjuje kvalitativne, kvantitativne i statističke računarske alate i metode za analizu podataka, sticanje uvida, informisanje i podršku donošenju odluka. Svaka određena analiza može da koristi različite tehnike uključujući dijagnostičke, prediktivne, preskriptivne i optimizacione modele (Power i dr., 2018). Mikalef i dr., (2019) daju plan za akademsko istraživanje i praktičnu primenu, ističući transformativni potencijal analitike kada se pravilno integriše u organizacione procese. U skladu s tim, autori navode da BA zahteva od organizacija da radikalno redizajniraju način na koji se takvim inicijativama pristupa, kako se dizajniraju i usavršavaju, kako se planiranje resursa i orkestriranje izvršava i strateški usklađuje, kao i da ponovno vrednuju svoje očekivane rezultate izvođenja, 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. Opšti cilj je poslovna učinkovitost kroz 'vertikalizaciju', upotrebljivost i integraciju s operativnim sistemima (Kohavi i dr., 2002). BA ima mnogo područja primene i povezanih izvedenica: finansijska analitika, analitika lanca snabdevanja, analitika krize, analitika znanja, marketinška analitika, analitika kupaca, analitika usluga,

analitika ljudskih resursa, analitika talenata, analitika procesa, analitika rizika (Holsapple i dr., 2014.) .

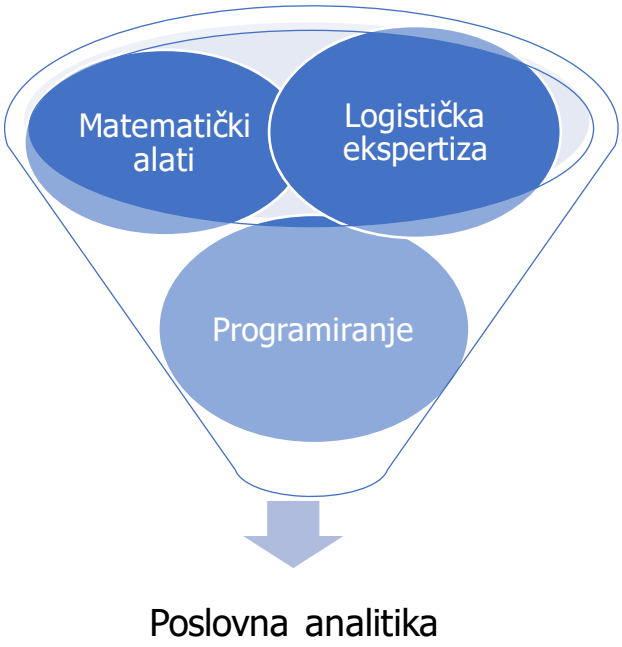
Postoje tri vrste platformi poslovne analitike:

- **Opisne** – posmatraju postojeće podatke i daju sažetak statistike i osnovnu vizualizaciju.
- **Prediktivne** – koriste postojeće podatke za procenu najverovatnijih budućih scenarija.
- **Preskriptivne** – automatski obrađuju veliku količinu podataka (engl. *big data*), poslovna pravila, tržišne uslove itd. Ove platforme koriste metode mašinskog učenja i veštačke inteligencije. Cilj je potpuno automatizovano donošenje odluka o tome koje akcije kompanija treba preduzeti s obzirom na trenutnu situaciju kako bi postigla željene poslovne ciljeve.

Zanimanje za *big data* i poslovnu analitiku eksponencijalno je poraslo tokom protekle decenije (Mikalef i dr., 2019). Savremeni BA ukorenjen je u stalnom napretku sistema za podršku odlučivanju. Ovaj napredak uključuje sve snažnije mehanizme za sticanje, generisanje, asimilaciju, izbor i emitovanje znanja relevantnog za donošenje odluka. S obzirom na nasleđe podrške odlučivanju, poslovna analitika nužno učestvuje u tim mehanizmima i iskorišćava ih. Znanje koje se mora obraditi kreće se od kvalitativnog do kvantitativnog, a BA uključuje obe vrste znanja, kako je prikladno za donošenje odluke (Kohavi i dr., 2002). Razlog zašto bi neka organizacija trebala primeniti BA je u problemima koje rešava. Problemi koje BA ističe su problemi smanjenja učinkovitog upravljanja preduzećem. Shodno tome, postoji nekoliko razloga za primenu BA (Holsapple i dr., 2014):

- Ostvarivanje konkurentske prednosti;
- Podrška strateškim i taktičkim ciljevima organizacije;
- Bolji organizacioni učinak;
- Bolji ishodi odluka;
- Bolji ili informisaniji procesi odlučivanja;
- Proizvodnja znanja;
- Dobijanje vrednosti iz podataka.

Bez obzira na vrstu platforme koja se koristi u BA za rešavanje i podršku procesu donošenja odluka u svakoj kompaniji, postoje tri ključna stuba svakog BA rešenja (slika 8.1). Generalno, BA zauzima mesto u spektru između računarske nauke/matematike/nauke o podacima (s jedne strane) i poslovanja i menadžmenta (s druge strane). Poslovna analitika zahteva i tehničko i poslovno znanje. Glavni problem u dizajniranju BA je to što granice nisu jasne (Power i dr., 2018). Shodno tome, za izvođenje BA potrebni su matematički alati za identifikaciju, izdvajanje i predstavljanje uvida na pravi način putem tabela, grafika, formula itd. Dodatno, alati za programiranje služe kao podrška ovoj vrsti aktivnosti i omogućavaju brze proračune bez grešaka, u poređenju s tradicionalnim pristupom papira i olovke. Poslednje, ali ne manje važno, potrebna je logistička ekspertiza u određenom području ili poslovnom problemu kako bi se odredili ključni uticajni faktori i povezani ekosistem.




**Slika 4.48Ključni stubovi BA u kontekstu lanca snabdevanja i logistike.**

Ključni potrošač je poslovni korisnik, čiji posao, verovatno u *merchandisingu*, marketingu ili prodaji, nije direktno 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 vreme 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 Šta je R?

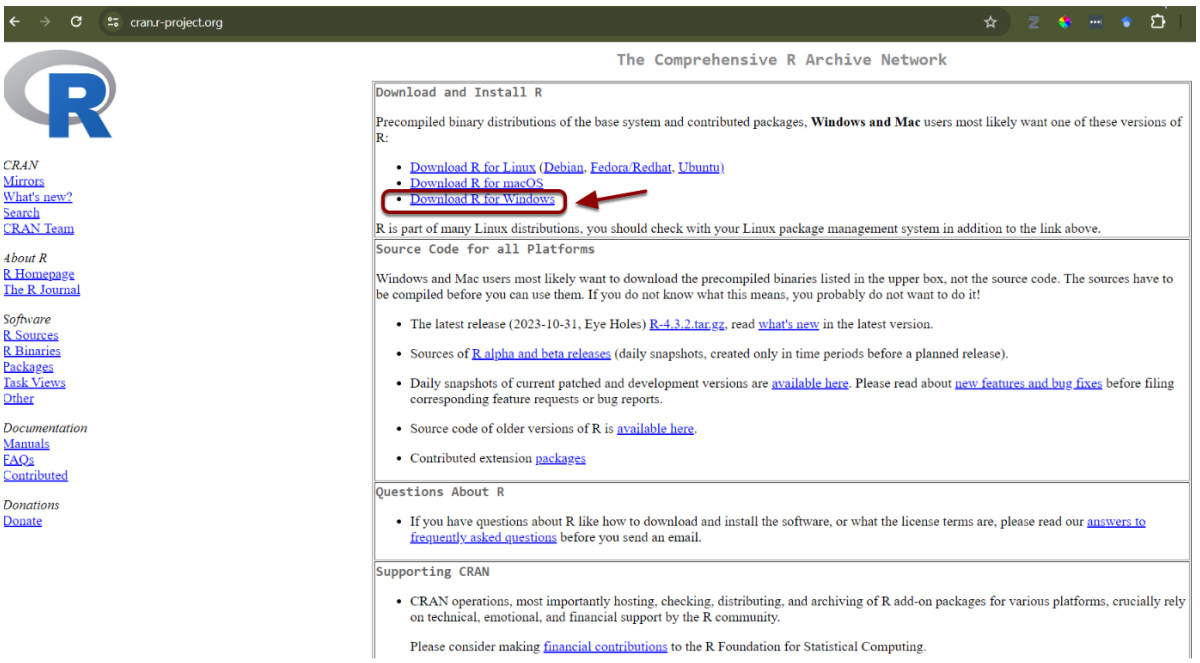
R je integrisani paket softverskih mogućnosti za manipulaciju podacima, proračun i grafički prikaz (R Core Team, 2019). Između ostalog, ima i sledeće:

- učinkovito rukovanje i skladištenje podataka,
- skup operatora za proračune na nizovima, posebno matricama,
- velika, koherentna, integrisana zbirka posrednih alata za analizu podataka,
- grafičke mogućnosti za analizu i prikaz podataka bilo direktno na računaru ili u štampanoj kopiji, te dobro razvijen, jednostavan i učinkovit programski jezik (nazvan 'S') koji uključuje uslove, petlje, korisnički definisane rekurzivne funkcije i mogućnosti unosa i izlaza (većina funkcija koje pruža sistem napisane su 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 korišćenje od programskih jezika kao što su C++ ili Fortran (Torfs i Brauer, 2014). R je u velikoj meri sredstvo za nove metode interaktivne analize podataka. Brzo se razvijao i proširio velikom kolekcijom paketa. Međutim, većina programa napisanih u R-u u suštini su prolazni, napisani za jednu analizu podataka (R Core Team, 2019).

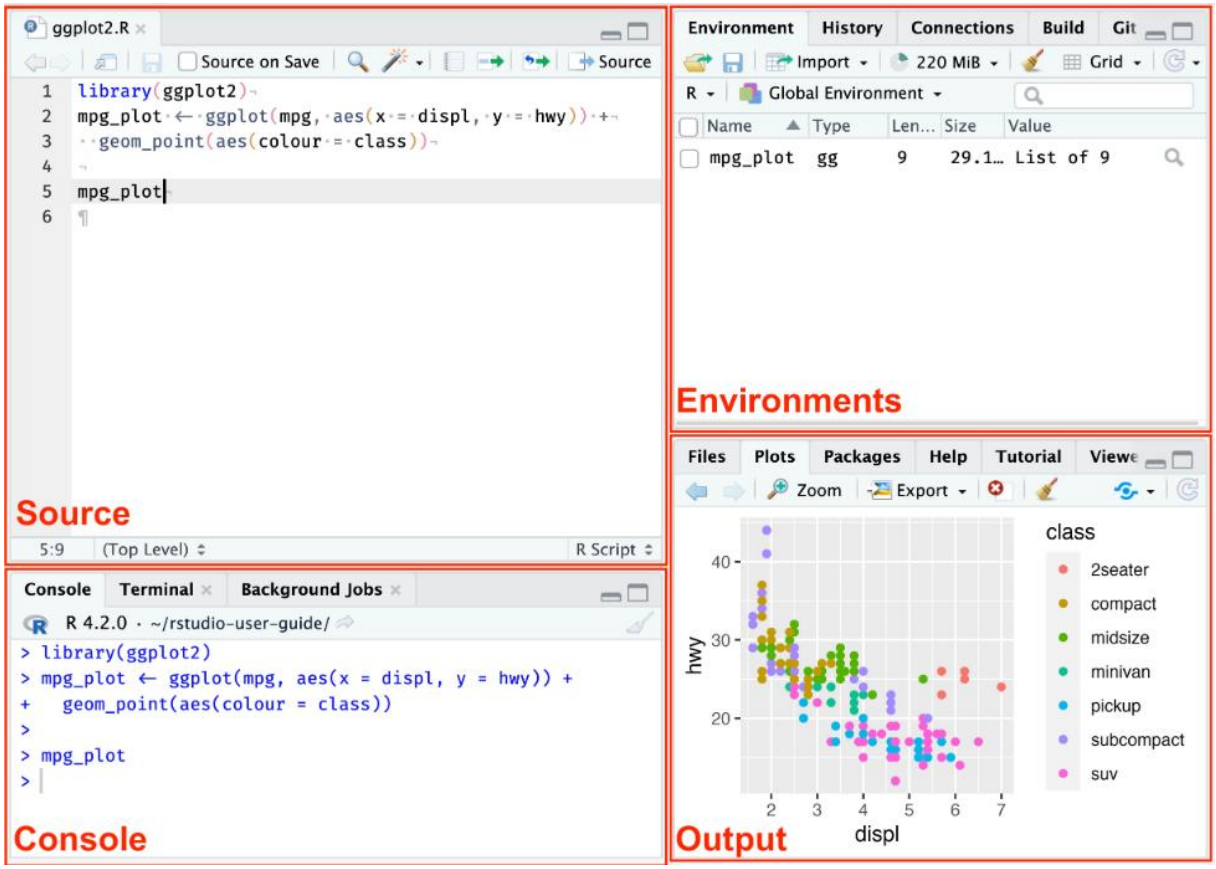
### Instalacija R-a i R Studija

Da biste instalirali R, idite na cran.r-project.org i kliknite na preuzimanje R za određeni operativni sistem na vašem računaru (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, biće bez naprednog integrisanog 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 sledeći ekran korisničkog interfejsa (slika 8.3).




**Slika 4.50Korisnički interfejs i R i Rstudio (RStudio, 2024).**

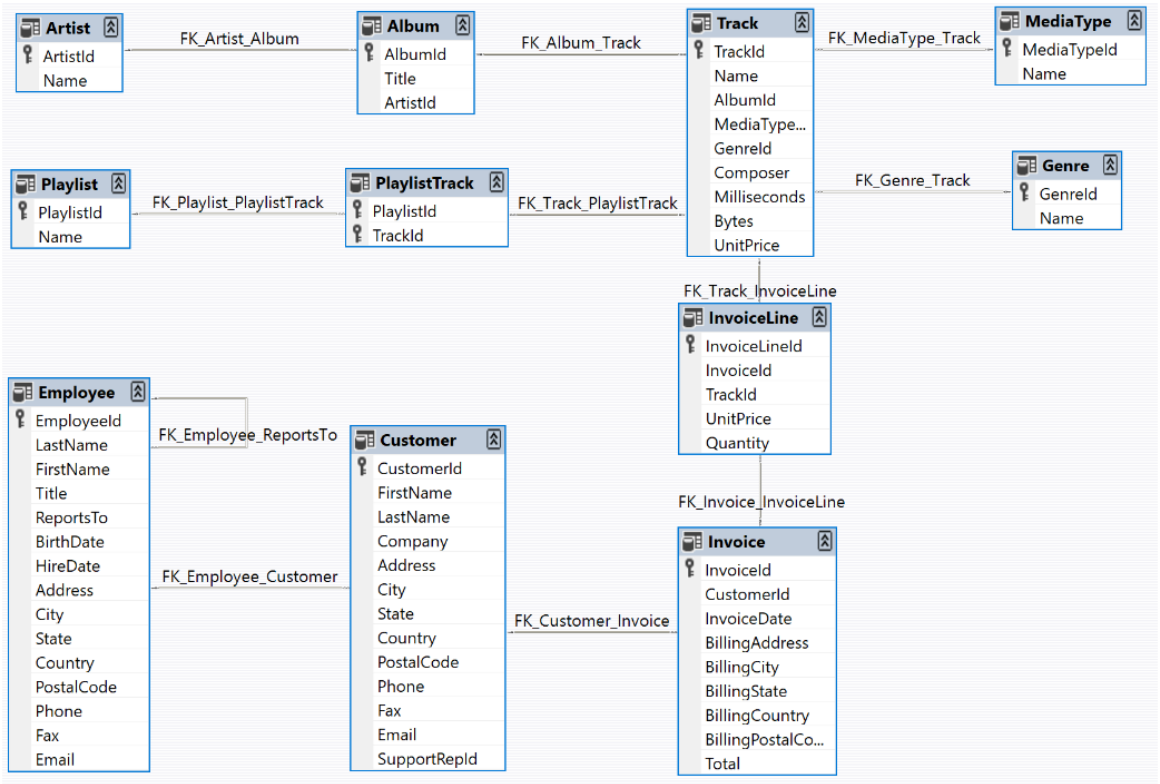
RStudio korisnički interfejs ima 4 primarna prozora (RStudio, 2024.):

- „Source" (srp. Izvor);
- „Console" (srp. Konzola);
- „Environments" (srp. Okruženje), koji sadrži kartice Okruženje, Istorija, Veze, Izrada, VCS i Vodič;
- „Ouput" (srp. Rezultat), koji sadrži kartice Datoteke, Ploče, Paketi, Pomoć, Pregledač i Prezentacija.

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

## 8.3 Šta je SQL i kako je povezan s BA i R?

Chinook baza podataka je ogledna baza podataka koja se koristi za učenje i demonstraciju sistema za upravljanje bazom podataka (engl. *database management systems* - DBMS) i SQL upita. Dizajnirana je kao digitalna multimedijalna prodavnica koja koristi stvarne podatke iz iTunes knjižnice; izmišljena imena/adrese kupaca i zaposlenih; i slučajne podatke za informacije o prodaji. Baza podataka sadrži različite tabele koje predstavljaju podatke muzičke prodavnice, uključujući informacije o izvođačima, albumima, pesmama, kupcima, fakturama i slično (slika 8.4).




**Slika 4.51Model podataka baze podataka Chinook.**

Slika 8.4 prikazuje model podataka koji okružuje Chinook bazu podataka s različitim tabelama podataka i njihovim ključevima (jedinstveni identifikatori) i zajedničkim tabelama poput tabele PlaylistTrack. Tabele prenose različite informacije o određenoj digitalnoj prodavnici (tabela 8.1).

**Tabela 4.3Informacije sadržane u svakoj od tabela baze podataka Chinook.**

| Naziv tabele | Opis |
|---|---|

| | |
|---|---|
| Umetnik | Sadrži informacije o muzičkim umetnicima. |
| Album | Sadrži informacije o muzičkim albumima, od kojih je svaki povezan s izvođačem. |
| Zapis | Sadrži informacije o pojedinačnim muzičkim zapisima, uključujući reference na albume, vrste medija i žanrove. |
| Žanr | Sadrži informacije o muzičkim 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. |
| Zaposleni | Sadrži informacije o zaposlenima, uključujući njihove uloge, veze s izveštajima 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 pesama | Sadrži informacije o popisima za reprodukciju. |
| Zapis popisa pesama | Povezuje pesme s popisima za reprodukciju, pokazujući koje su pesme uključene u koje popise za reprodukciju. |

## 8.4 Kako su povezani poslovna analitika, SQL i R?

Veza između BA, SQL i R je prirodna jer bi svi poslovni podaci trebali da se skladište u SQL bazama podataka. Ovo je još uvek idealistički cilj budući da još uviek postoji loše upravljanje podacima u nekim delovima malih i srednjih preduzeća koja još uvek ne razumeju u potpunosti snagu podataka. U velikim kompanijama to je davno prepoznato i podaci su pravilno strukturirani u bazama podataka (SQL ili nekoj drugoj, ali najčešće u SQL-u). S druge strane, analizu podataka moguće je izvršiti u SQL-u, ali je u tu svrhu bolje koristiti statistički orijentisani softver, gde je u fokusu R, kao jedna od najpopularnijih statističkih platformi za analizu podataka.

Shodno tome, SQL i R mogu se smatrati savršenim alatom za saradnju kada je problem u pitanju iz područja BA. Postoji nekoliko glavnih razloga, a jedan od njih je taj što se BA podaci svakodnevno menjaju i ažuriraju u skladu s realnim tržištem i aktivnostima kompanije: prodaja, zaposleni, prihodi itd. SQL baze podataka savršene su za beleženje tih promena i ažuriranje postojećih podataka, dok su R skripte vrlo dobre u automatizaciji zadataka kao i u dizajniranju novih paketa za analizu datih podataka. Razlog za to je što je R

više građen oko koncepta analize podataka, nego na opštem programiranju kao što je Python, na primer.

### 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 standardizovani interfejs za interakciju s različitim DBMS-ovima, dopuštajući korisnicima povezivanje, postavljanje upita i dosledno upravljanje transakcijama u različitim bazama podataka. Paket RSQLite, koji se pridržava DBI interfejsa, posebno olakšava interakciju s bazama podataka SQLite, omogućavajući korisnicima izvršavanje SQL upita, preuzimanje podataka i izvođenje operacija baze podataka direktno iz R-a. Zajedno, ovi paketi pojednostavljuju proces rada s bazama podataka u R-u, nudeći kohezivan i učinkovit tok rada.

Kako bismo učinkovito demonstrirali izvođenje SQL operacije iz R-a i generisanje željenih uvida iz podataka u vezi s postojećim problemom, dali smo nekoliko isečaka 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 4.52Isečak koda za uspostavljanje veze između SQL-a i R-a i istraživanje tabela podataka sadržanih u SQL-u.**

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

Nakon što se veza uspostavi, postoji niz mogućih analiza koje se mogu sprovesti, zavisno od poslovnog cilja i budućoj upotrebi datih rezultata. Ovde ćemo, zbog ograničenja prostora, pokazati samo delić moguće analize podataka, s malim isečkom 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). Tabela 8.2 predstavlja rezultate upita podataka putem isečka 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 4.53Isečak koda za postavljanje upita SQL-u putem R-a i određivanje 10 najprodavanijih albuma.**

**Tabela 4.410 najprodavanijih albuma u digitalnoj prodavnici Chinook.**

| Naslov albuma | Ime umetnika | Prodata 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 umetnika | Prodata 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 karakteristika vremenskih serija u lancima snabdevanja

Šta je predviđanje potražnje? Kako možemo efikasno vizualizovati podatke o kupcima i doneti zaključke o njima? Kako sprovesti inženjering karakteristika vremenskih serija?

Na ova i slična pitanja pokušaćemo dati odgovore u sledećem poglavlju.

## 9.1 Šta je potražnja kupaca i predviđanje potražnje?

Zahtev krajnjeg kupca pokreće ceo lanac snabdevanja (Syntetos i dr., 2016). Shodno tome, potražnja kupaca je ključna komponenta za planiranje svih logističkih procesa u lancu snabdevanja pa je određivanje nivoa potražnje kupaca od velikog interesa za menadžere lanca snabdevanja. Komplementarno, predviđanje potražnje bitna je aktivnost za planiranje i raspoređivanje logističkih aktivnosti unutar promatranog lanca snabdevanja (Mircetic i dr., 2017). Precizni modeli predviđanja potražnje direktno utiču na smanjenje logističkih troškova budući da daju procenu potražnje kupaca (Mircetic i dr., 2016). Predviđanje u lancima snabdevanja prevazilazi operativni zadatak ekstrapolacije zahteva potražnje na jednom nivou. Uključuje složena pitanja kao što su koordinacija lanca snabdevanja i dijeljenje informacija između višestrukih učesnika (Syntetos i dr., 2016).

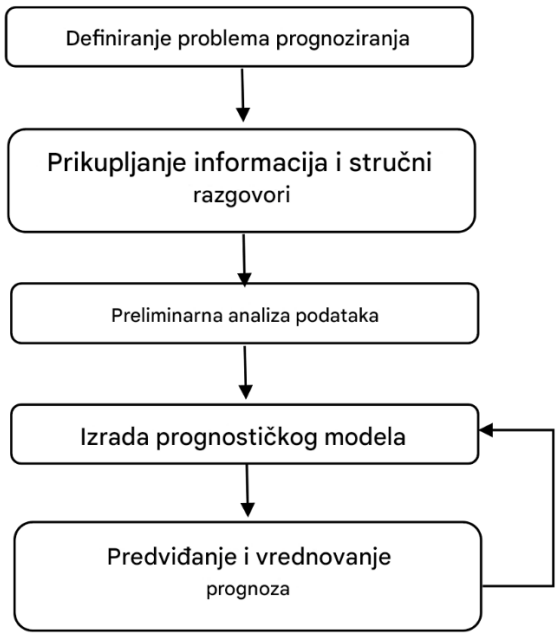
Potražnja kupaca i prateće prognoze su od vitalnog značaja za lance snabdevanja, budući da pružaju osnovne ulazne podatke za planiranje i kontrolu svih funkcionalnih područja, uključujući logistiku, marketing, proizvodnju itd. (Mircetic, 2018). Kad bi potražnja krajnjih kupaca bila stalna, ili poznata sa sigurnošću puno unapred, tada bi rad lanca snabdevanja bio direktna (unazad) vežba planiranja. Međutim, potražnja nije poznata i stoga je treba predvideti. Nesigurnost povezana s ovom potražnjom čini upravljanje lancem snabdevanja

vrlo teškim (Syntetos i dr., 2016). Na učinkovitost predviđanja potražnje utiču inherentne neizvesnosti u vremenskoj seriji potražnje koje lanci snabdevanja imaju (Rostami-Tabar, 2013.). Posledično, rešavanje i razumevanje ovih neizvesnosti je veliki izazov za menadžere prilikom koordinacije i planiranja operacija unutar lanaca snabdevanja (Mircetic, 2018).

Nesigurnost potražnje je jedan od najznačajnijih izazova za savremene lance snabdevanja. Nedavna pandemija COVID-19 dodatno je naglasila ovaj problem, uzrokujući široko rasprostranjene poremećaje koji su komplikovali planiranje i kontrolu lanca snabdevanja (Nikolopoulos i dr., 2020). Predviđanje potražnje u lancima snabdevanja često uključuje predviđanje potražnje za brojnim artiklima. Prognostičari u lancima snabdevanja obično ekstrapoliraju vremenske serije podataka za svaku jedinicu zaliha (engl. *stock-keeping unit* – SKU) pojedinačno. Na primer, trgovac na malo može koristiti podatke o prodajnom mestu za generisanje predviđanja na nivou pojedinačne prodavnice (Mircetic i dr., 2022).

## 9.2 Koraci predviđanja potražnje u lancu snabdevanja?

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




**Slika 4.54Osnovni koraci za pravilnu implementaciju predviđanja unutar preduzeća (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 orijentisanih) prognoza. Shodno tome, definisanje problema često je najteži deo predviđanja i zahteva razumevanje načina na koji će se predviđanja koristiti, kao i uloge funkcija predviđanja unutar promatranog preduzeća. Prognostičar bi trebao potrošiti dosta vremena na komunikaciju sa svima koji su uključeni u prikupljanje podataka, održavanje baze podataka i korišćenje prognoza za buduće planiranje. Jedan od glavnih otežavajućih činioca u definisanju problema je kako će se konačna prognoza koristiti u svakodnevnim logističkim procesima (koja platforma, dizajn softvera, korisnički interfejs itd.).

Za fazu prikupljanja informacija uvek su potrebne najmanje dve vrste informacija: statistički podaci i akumulirana stručnost ljudi koji prikupljaju podatke i koriste predviđanja. U praksi je često teško dobiti istorijske podatke za kreiranje dobrog statističkog modela. Takođe, postoji veliki nesporazum o tome šta su podaci o potražnji i što se može koristiti kao njihova zamena. Postoji loša praksa korišćenja podataka o otpremi i isporuci kao zamene za podatke o potražnji, što će samo pogoršati proces donošenja odluka na bazi predviđanja napravljenih na jednostavnoj vrsti podataka. Podaci o prodaji jedina su pouzdana zamena za podatke o potražnji (Syntetos i dr., 2016), iako ovo pojednostavljenje nije savršeno, posebno u lancima snabdevanja s puno situacija kada nedostaju zalihe (engl. *out-of-stock* – OOS).

Za fazu preliminarne analize podataka, preporučuje se uvek započeti analizu podataka s grafičkim prikazima kako bi se odgovorilo na sledeća pitanja. Postoje li dosledni obrasci? Postoji li značajan trend? Postoji li primetna 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 dalju analizu podataka sužavanjem fokusa na modele koje treba primeniti na otkrivena obležja potražnje. Uglavnom, jednostavni model, ovako određen, može pobediti one sofisticiranije i komplikovanije (Rostami-Tabar & Mircetic, 2023).

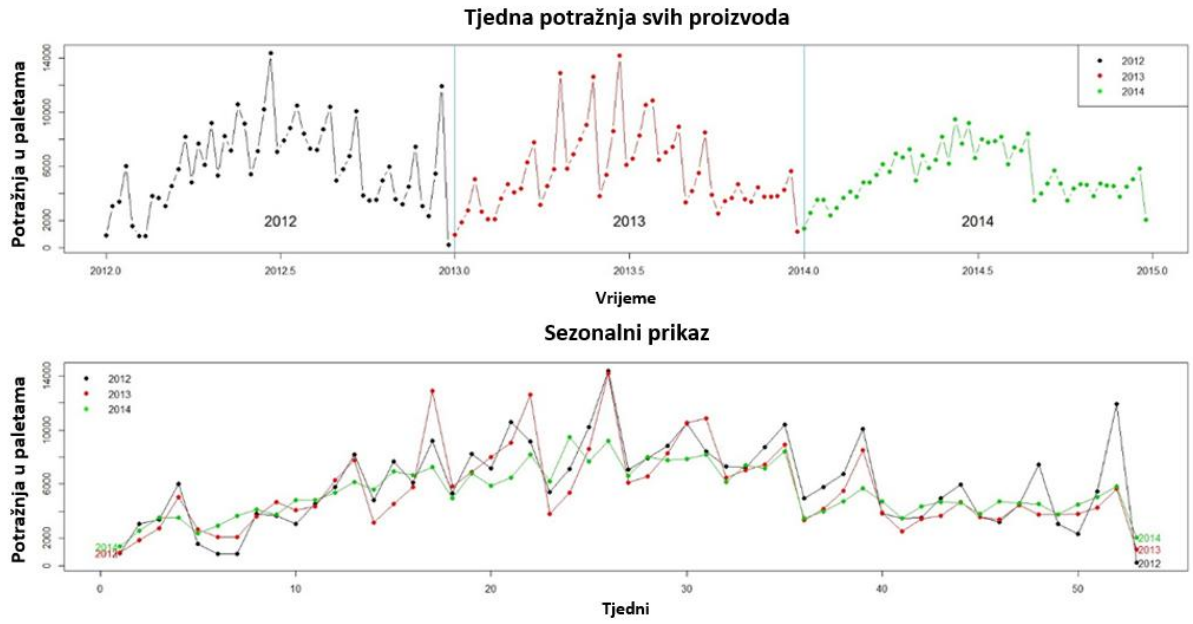
Izbor i izrada modela predviđanja je najvažniji korak pri izradi modela predviđanja. Koji model koristiti zavisi od nekoliko faktora, od kojih su najvažniji dostupnost istorijskih podataka i korelacija između zavisnih i nezavisnih varijabli. Uobičajeno je da se prilikom izbora modela upoređuju dva ili tri potencijalna modela. Svaki model je veštačka konstrukcija bazirana na skupu pretpostavki (eksplicitnih i implicitnih) i generalno uključuje jedan ili više parametara koji se moraju izraditi korišćenjem poznatih istorijskih podataka. U sledećem

poglavlju biće prikazan proces razvoja i primene modela ARIMA, kao jednog od najboljih i najpopularnijih modela.

Ocenjivanje modela predviđanja je korak kojim se meri upotrebljivost kreiranog modela. Nakon izbora prognostičkog modela i procene njegovih parametara, model se koristi za izradu prognoza. Tačnost modela procenjuje se korišćenjem različitih statistika, ali takođe je važno testirati predviđanja putem merila poslovnih implikacija (tj. metrike korisnosti).

## 9.3 Predviđanje potražnje u prehrambenoj industriji

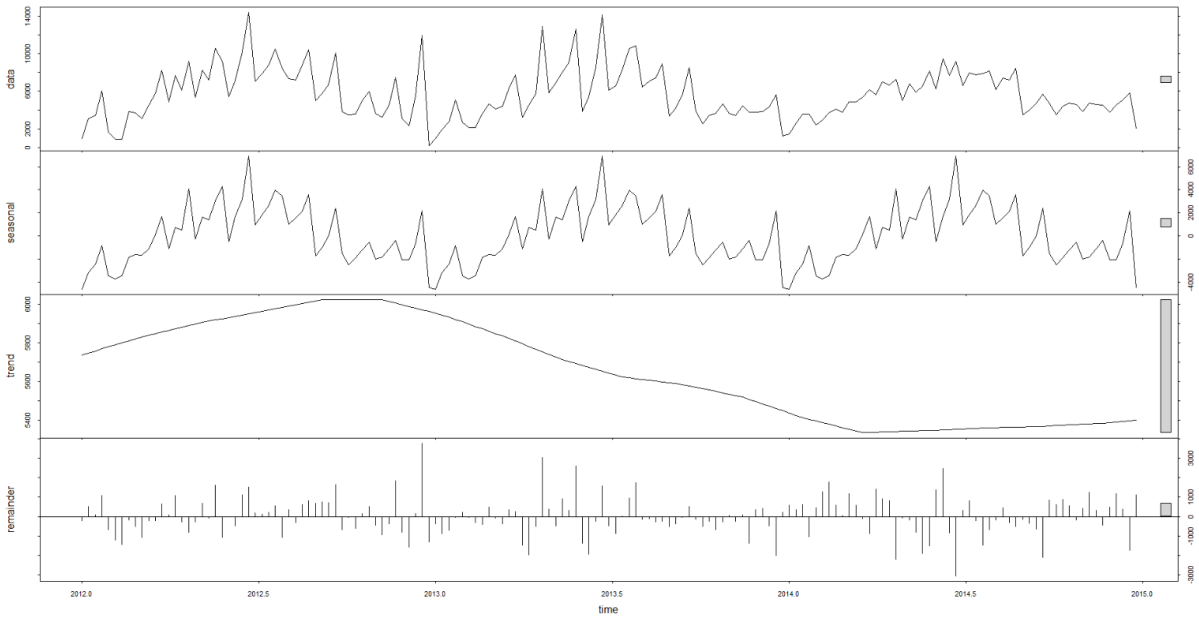
Podaci o potrošnji za sve proizvode promatranog preduzeća u prehrambenoj industriji prikazani su kao nedeljna potražnja u periodu od januara 2012. do decembra 2014. (slika 9.2). X osa predstavlja vreme, dok su vrednosti potražnje prikazane na y osi. Ovi podaci prikazani su u nedelnim intervalima, jer to odgovara periodu u kojem se vrši snabdevanje krajnjih prodajnih mesta. Shodno tome, menadžment kompanije fokusiran je na predviđanje nedeljne potrošnje tržišta.




**Slika 4.55Podaci o potražnji za posmatranu prehrambenu kompaniju.**

Slika 9.2 pokazuje nekoliko važnih obeležja i karakteristika datih podataka. Prvo, podaci imaju snažan sezonski karakter s vršnom prodajom koja se događa sredinom godine (letnji meseci). Drugo, podsezonski grafikon (donji grafikon na slici 9.2) prikazuje značajnu

promenu uzorka pada trenda u 2014.! Ovo je vrlo značajna karakteristika za izbor pravog modela predviđanja i za menadžere u preduzeću budući da otkriva značajan pad potrošnje i gubitak tržišta. Za dalje istraživanje posmatranih karakteristika koristi se metodologija dekompozicije sezonskog trenda (engl. *seasonal trend decomposition* - STL) (slika 9.3). STL dekompozicija deli izvorne obrasce potražnje na tri komponente: sezonsku, trend i ostatak.




**Slika 4.56STL dekompozicija podataka o potražnji.**

STL je otkrio da posmatrane vremenske serije pokazuju aditivnu prirodu, što znači da se fluktuacije oko krive linije trend-ciklus ne povećavaju značajno tokom vremena. Kao rezultat, Box-Coxove transformacije nisu bile potrebne za neobrađene vremenske serije. Dekompozicija je otkrila da je sezonska komponenta dominantna u posmatranom nizu, pokazujući visoke fluktuacije unutar jedne godine. S obzirom na ograničeni broj godina posmatranja, teško je identifikovati poslovni ciklus. Dekompozicija je takođe pokazala da je trend u seriji minimalan, s opadajućim uzorkom počevši od sredine 2013.

Ove identifikovane karakteristike predstavljale su važan input tokom 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 kombinuje i autoregresivnu komponentu i komponentu pokretnog proseka, zajedno s razlikom kako bi podaci bili stacionarni. Ovaj model je posebno uspešan 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, omogućava preciznije i pouzdanije prognoze, koje su ključne za uspešno planiranje lanca snabdevanja i upravljanje zalihama. S-ARIMA strukturni oblik predstavljen je u jednačini (1).

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

gde su $\Phi(z)$ i $\Theta(z)$ polinomi reda *P* odnosno *Q*, od kojih svaki ne sadrži korene 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 korišćenju S-ARIMA modela je izbor odgovarajućeg redosleda modela, što uključuje određivanje vrednosti *p, q, P, Q, D* i *d*. Ako su *d* i *D* poznati, poredak *p, q, P* i *Q* može se odabrati korišćenjem informacionog kriterijuma kao što je Akaikeov informacioni kriterijum (AIC) ili Bayesov informacioni kriterijum (BIC). Formule za AIC i BIC date su prema:

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

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

gde *je* $k=p+q+P+Q+1$ ako je uključen konstantni izraz i 0, inače, *L* je maksimizirana verovatnoća modela prilagođenog različitim podacima, *SSE* je zbir kvadrata grešaka, *N* je broj opažanja koji se koristi za procenu, a *k* je broj prediktora u modelu.

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

- Upotrebite Canova-Hansen test za određivanje *D* u okviru ARIMA.
- Odaberite *d* primenom uzastopnog KPSS testa jediničnog korena na sezonski diferencirane podatke (ako je $D = 1$) ili na izvorne podatke (ako je $D = 0$).
- Odaberite optimalne vrednosti 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 njihovi rezultati prikazani su u tabeli 9.1. Za testiranje performansi različitih modela koristi se nekoliko mera: srednja apsolutna procentualna greška (engl. *mean absolute percentage error* - MAPE), koren srednje kvadratne greške (engl. *root mean square error* - RMSE), srednja apsolutna skalirana greška (engl. *mean absolute scaled error* - MASE), AIC i BIC (jednačine 2 i 3)

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

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

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

gde 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 mere za menadžere u lancima snabdevanja su RMSE i MAPE jer daju "osećaj" koliko je model dobar u realnim brojevima (RMSE) i procentima (MAPE).

**Tabela 4.5Rezultati 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. | 21,42 % | 1,02 % | 36.52 | 40,46 0 |

godine

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

Tabela 9.1 pokazuje da je model S-ARIMA (5,0,1)(1,0,0) $_{52}$ nadmašio konkurentne modele postižući najniže greške RMSE, MAPE i MASE. Oblik modela S-ARIMA (5,0,1)(1,0,0) $_{52}$ predstavljen je u jednačini (4), gde 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 .

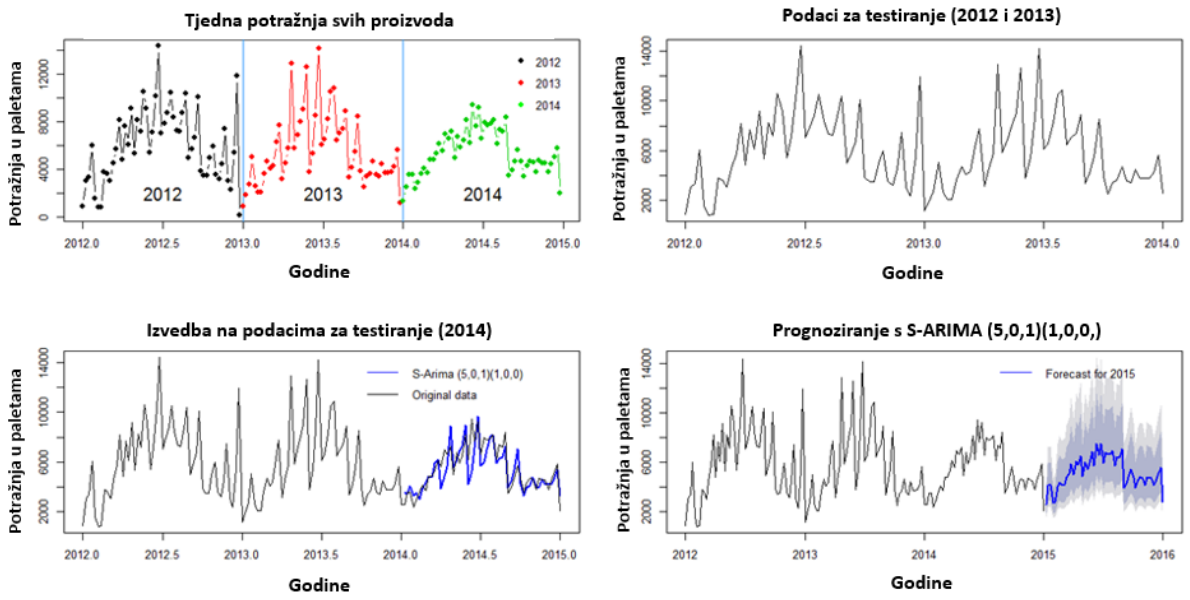
$$(1-\phi_1 B-\phi_2 B-\phi_3 B-\phi_4 B-\phi_5 B)(1-\Phi_1 B^{52})y_t = c+(1+\theta_1 B)e_t \text{ , } (4)$$

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 predviđanja, model S-ARIMA (5,0,1)(1,0,0) $_{52}$ u proseku proizvodi grešku od 1023 proizvoda, što znači 14,18 %. Ovi rezultati naglašavaju važnost pažljivog izbora uslova za uključivanje u S-ARIMA model budući da nema značajnih razlika između performansi modela s različitim parametrima. Procena je pokazala da su modeli koji uključuju autoregresivne (*p, P*) i komponente pokretnog proseka (*q, Q*) učinkovitiji u predviđanju potrošnje pića od onih koji uključuju sezonske ili nesezonske razlike. Osim toga, uporedni 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 lošije rezultate od jednostavnog prosečnog modela naivne prognoze, što dokazuje njihova MASE greška veća od jedinice.

## 9.5 Predviđanja buduće potražnje

Slika 9.4 prikazuje rezultate S-ARIMA (5,0,1)(1,0,0) $_{52.}$ Gornji levi 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, kao ulaz postoje podaci za dve godine i treba predvideti buduću potražnju godinu dana unapred! Ovo je prilično čest scenariio u lancima snabdevanja i logistici jer postoji nepisano pravilo da kompanije čuvaju istoriju 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 levi prikaz na slici 9.4 predstavlja performanse promatranog modela. Učinkovitost modela već je prikazana u tabeli 9.1 putem različitih statističkih mera, ali menadžerima je obično teško da steknu utisak koliko je model dobar ili loš. U tu svrhu grafički prikaz prikazuje performanse modela. Moglo bi se tvrditi da model prilično dobro prati testne podatke i za većini perioda pokazuje izvrsne performanse. Kako bi se generisale 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 nedelje unapred 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 vrednosti. Model predviđa kontinuirani pad potražnje koji je započeo 2014. Uzroci ovog pada mogu biti različiti i menadžeri na strateškom nivou preduzeća trebali bi ih dodatno istražiti.

## 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. Veštačka inteligencija i mašinsko učenje u lancima snabdevanja

Šta je veštačka inteligencija (AI)? Da li je to stvarno "živo" stvorenje sposobno razmišljati i donositi sopstvene odluke na osnovu svog uma, prošlih iskustava, etike, uverenja itd.? Kako je to povezano sa mašinskim 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 procesima i optimizacijama? Da li postoje specifične arhitekture i primeri primene AI-a i ML-a na lance snabdevanja?

Na ova i slična pitanja pokušaćemo dati odgovore u sledećem poglavlju, završavajući stvarnim primerom studije slučaja primene AI & ML algoritama u distributivnom skladištu.

## 10.1 Šta je veštačka inteligencija?

Područje veštačke inteligencije započelo je 1950-ih kada su se stručnjaci za računare zapitali: "Da li računari mogu razmišljati kao ljudi"? Istraživači su u to vreme bili oduševljeni mogućnošću učenja računara 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 automatizuju intelektualni zadaci koje obično obavljaju ljudi (Chollet, 2021). Algoritmi u području veštačke 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. Štaviše, u ranoj fazi razvoja veštačke inteligencije, ovi su algoritmi bili dominantniji. Prema tome, taj deo veštačke inteligencije poznat je kao simbolička veštačka inteligencija koja se bazira na ideji da se može postići ljudski nivo performansi i inteligencije programiranjem računara s velikim skupom eksplicitnih pravila za rešavanje posmatranog problema. Ovaj pristup dao je izvrsne rezultate u logičkom problemu koji je bio dobro definisan, poput računara koji igra šah, ali se pokazao komplikovanim za složenije probleme. Stvarni svet se pokazao puno komplikovanijim nego što bi se sva eksplicitna pravila programiranja mogla uneti u računar.

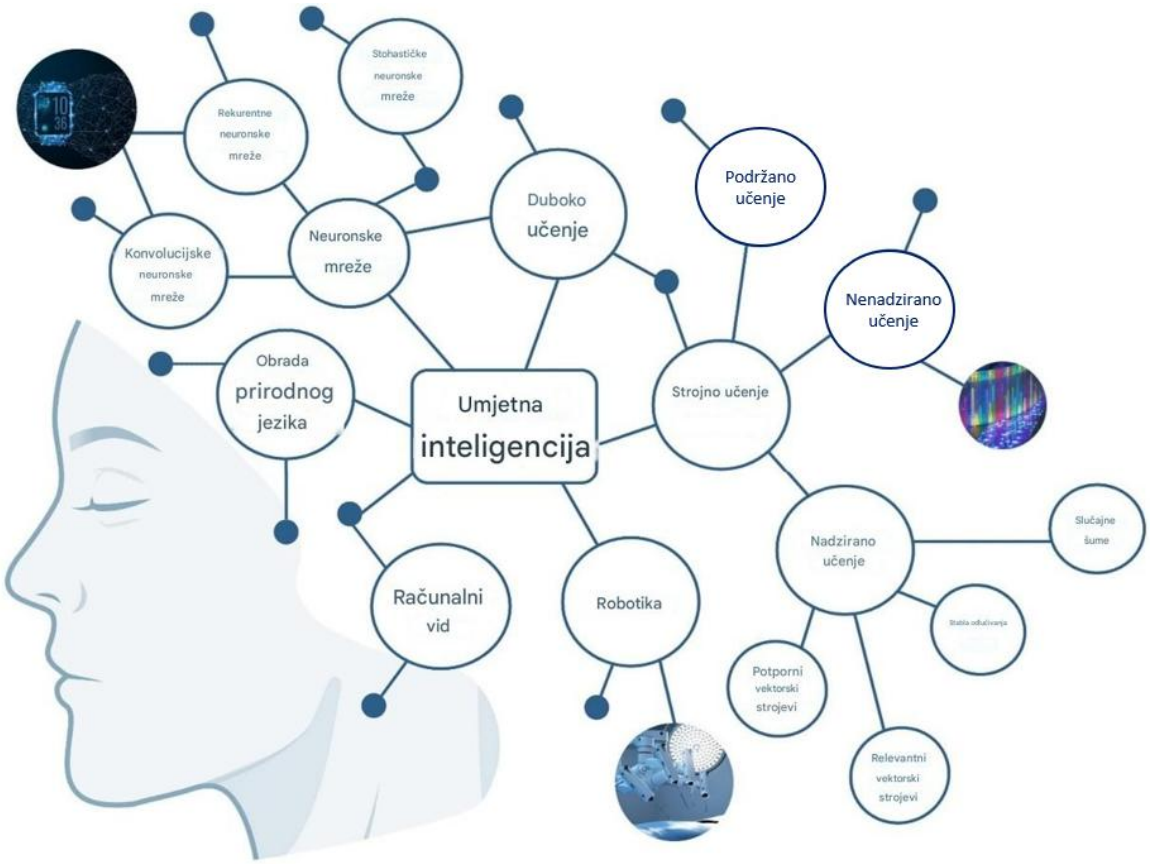
Ovaj pristup je fokusiran na ideju za datu situaciju - učini ovo ili ono (ako-onda pravila). Ovo je lako razumljiv pristup, ali s druge strane vrlo dugotrajan i ponekad je jako teško odrediti sve moguće scenarije koje je potrebno uneti u program. Kao smešan primer, ali dobra ilustracija ove teme, pogledajte sliku 10.1 koja prikazuje nekoliko scenarija za određivanje prognoze na osnovu statusa kamena.


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

Područje simboličke veštačke inteligencije najveću je popularnost steklo 1980-ih s pojavom ekspertskih sistema (ES). ES predstavljaju podskup sistema za podršku odlučivanju (DSS) (Turban, 1998), usmerenih na pružanje sposobnosti računarima da donose odluke sličnih onima koje donose stručnjaci unutar određenog područja. Ovi su sistemi izrađeni za rešavanje zamršenih problema upotrebom niza pravila ili algoritama koji simuliraju procese ljudskog zaključivanja. Olson i Courtney (1992) opisuju ekspertske sisteme (ES) kao računarske programe koji simuliraju ljudske misaone procese za donošenje odluka unutar određenog domena, uključujući stepen veštačke inteligencije koji odgovara zaključcima do kojih bi ljudski stručnjak došao. ES komponenta je posebno korisna za podršku donosiocima odluka u područjima koja zahtevaju specijalizovana znanja (Turban i dr., 2005). U osnovi, ES prikuplja ekspertizu ljudskog stručnjaka (ili drugog izvora) i prenosi je na računar. Ova tehnologija može ili pomoći donosiocima odluka ili ih u potpunosti zameniti, što je čini jednim od najšire primenjivanih i komercijalno najuspešnijih oblika veštačke inteligencije (Turban i dr., 2005). Jedan od ključnih razloga za razvoj ES-a je distribucija stručnog znanja široj publici (Jackson, 1999). U sledećim potpoglavljima, demonstriraće se primena ES-a

zasnovanog na AI i ML algoritmima, menadžerima u centralnom skladištu kao delu celokupnog DSS-a.

Danas se područje veštačke inteligencije sastoji od različitih pristupa i algoritama, no oni koji se najviše koriste opisani su na slici 10.2. Razišlo se od ES sistema, 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, 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 lekova i obrada prirodnog jezika. Duboko učenje ima sposobnost automatskog otkrivanja relevantnih obeležja, smanjuje potrebu za ljudskom intervencijom u dizajnu obeležja, što ga čini vrlo uspešnim u iskorišćavanju velikih skupova podataka i snage računara (LeCun i dr., 2015).




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

Veliki korak prema današnjoj situaciji bilo je istraživanje klasifikacije vrednosti koje su sproveli Hinton i dr. (2006), koje je uspelo postići više od 98% tačnosti na klasifikaciji baze podataka Modifikovanog nacionalnog instituta za standarde i tehnologiju (engl. *Modified National Institute of Standards and Technology* - MNIST). Jedan od načina razmišljanja o tome kako je veštačka inteligencija prešla iz simboličke veštačke inteligencije u ML i šta je suština ML-a jeste zamisliti ML algoritme kao amorfnu masu koja se sama oblikuje prema željenim ishodima. Sistem pravila koji uzima input u output menja se od problema do problema i prilagođava postojećem stanju. Cilj mu je pronaći pravila koja će automatizovati zadatak traženjem statističkih obrazaca unutar podataka. Ovakav pristup rešavanju različitih problema značajno smanjuje vreme postavljanja sistema (u poređenju sa ako – onda pravilima) i čini ga univerzalnijim pristupom za rešavanje različitih problema.

Bengio i dr. (2021) naglasili su da je budućnost veštačke inteligencije u dubokom učenju i revolucionarnom uticaju meke pažnje i transformatorskih arhitektura u veštačkoj inteligenciji. Ove inovacije omogućavaju neuronskim mrežama da se dinamički fokusiraju na važne ulaze i skladište informacije u diferencijalnim memorijama, značajno poboljšavajući sekvencijalnu obradu.

## 10.2 Šta je ekosistem AI-a i ML-a?

Ekosistem AI i ML algoritama sastoji se od tri ključna stuba:

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

Ulazni podaci predstavljaju zapise podataka određenih karakteristika ili karakteristike (zavisno od uočenog problema). Tačnost ulaznih podataka ključna je za izgradnju tačnih algoritama. To često nije slučaj u stvarnim aplikacijama i obično se značajno vreme i trud posvećuju prikupljanju podataka, čišćenju, argumentaciji, objedinjavanju, proveri lažnih unosa itd. Osim tačnih podataka, drugi važan aspekt karakteristika ulaznih podataka je njihova reprezentacija i kodiranje. Različiti načini kodiranja podataka mogu otkriti različite karakteristike podataka i značajno "pomoći" ML modelima u otkrivanju skrivenih obrazaca u podacima. Ovde vidimo Ahilovu petu ML algoritama. Često se previše pažnje posvećuje

kreiranju metoda za izvlačenje informacija i inteligencije iz podataka (tj. samih algoritama), dok se premalo pažnje posvećuje ulaznim podacima i njihovom odnosu s izlaznim podacima. Generalno se podrazumeva da su ulazni podaci u uzročno-posledičnoj vezi s izlaznim podacima, što ponekad uopšte nije slučaj. Stoga bi sledeći veliki korak u razvoju ML-a trebao biti pronalaženje boljih načina za prikupljanje, predstavljanje i kodiranje podataka.

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

Troškovna funkcija predstavlja način merenja učinka AI i ML. U osnovi, idealno bismo želeli da odgovori algoritama odgovaraju izlaznim podacima, za date ulazne podatke. Troškovna funkcija takođe je povratni signal skupu parametara koji usmeravaju rad algoritma, tj. omogućuje optimizaciju realizacije algoritma kroz proces učenja (pronalaženje optimalnog skupa parametara). Proces učenja obično uključuje nadzirano učenje, gde se model trenira na označenim podacima kako bi se smanjile greške predviđanja pomoću tehnika kao što su stohastički gradijentni spuštanje i širenje unazad. To omogućava modelu da uspešno prilagodi svoje unutašnje parametre, što dovodi do poboljšane realizacije zadataka kao što su otkrivanje i klasifikacija objekata (LeCun i dr. 2015).

## 10.3 Koji se alati koriste u ML-u?

Generalno, ML algoritmi mogu se klasifikovati u dvije glavne kategorije: nadzirano i nenadzirano učenje.

Nadzirano učenje uključuje uvežbavanje algoritama na označenom skupu podataka, gde je svaka ulazna tačka podataka uparena s tačnim izlazom. Ova jasna "slika" o tome kakav bi trebao biti tačan odgovor za dati ulaz omogućava algoritmu da nauči funkciju mapiranja od ulaza do izlaza. Shodno tome, poznati su i ulazni i izlazni podaci (Athanasopoulou i dr., 2022). Uobičajene primene nadziranog učenja uključuju zadatke klasifikacije (npr. određivanje da li je e-pošta spam ili ne) i zadatke regresije (npr. predviđanje cena kuća na osnovu različitih karakteristika). Neki od najpopularnijih algoritama koji su dokazani brojnim primenama su generalizovani aditivni modeli, slučajne šume, boosting, klasifikaciona i

regresiona stabla, potporni vektorski uređaji, proširena linearna regresija, logistička regresija, k-najbliži susedi, linearna diskriminantna analiza, laso, neuronske mreže, adaptivni neuro-fuzzy sistem zaključivanja itd. (Rostami-Tabar i Mircetic, 2023). Nadzirano učenje je moćno jer iskorišćava podatke koje su komentarisali ljudi za postizanje visoke tačnosti u predviđanjima. Međutim, njegova učinkovitost uveliko zavisi od kvaliteta i količine označenih podataka.

Nasuprot tome, nenadzirano učenje bavi se skupovima podataka koji nemaju označene odgovore. Shodno tome, algoritmi nenadziranog mašinskog učenja koriste neoznačene skupove podataka koji uključuju samo ulaze (Athanasopoulou i dr., 2022). Ovde se algoritam daje samo s ulaznim podacima, a cilj mu je pronaći osnovne obrasce, strukture ili odnose unutar podataka. Uobičajene tehnike učenja bez nadzora uključuju grupisanje (npr. grupisanje 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 retki ili nedostupni.

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

Jedan od najpopularnijih i najuspešnijih algoritama za ML dolazi iz oblasti neuronskih mreža. Neuronske mreže postoje od 1950-ih, ali su svoju popularnost stekle 1980-ih i u poslednjih 12 godina. Izgrađene su na aproksimaciji bioloških neurona i načina na koji dele delove informacija u mozgu, ali osim toga, nema značajnih mešusobnih veza. Danas se najčešće koristi oblik neuronskih mreža u obliku dubokog učenja, koji predstavlja nekoliko skrivenih slojeva između ulaznih i izlaznih karakteristika, koji izvode nekoliko nelinearnih transformacija ulaznih karakteristika. Budući da se ovo pokazalo vrlo uspešnim, duboko učenje danas je jedno od najistaknutijih potpodručja mašinskog učenja (Chollet, 2021).

## 10.4 Studija slučaja?

Nalazi Wenzela i dr. (2019) o ML-u u upravljanju lancem snabdevanja ukazuju na rastuću integraciju ML aplikacija u različitim zadacima lanca snabdevanja. Shodno tome, posmatrana studija slučaja predstavlja primenu AI-a i ML-a, izvedenu u centralnom skladištu fabrike hrane (Mirčetić i dr., 2016; Mirčetić i dr., 2014). U krugu fabrike nalazi se 30 viljuškara. Viljuškari su angažovani 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 mesta i godišnju proizvodnju od 300.000 do 350.000 paleta. Trenutno fabrika direktnom dostavom snabdeva oko 20.000 prodavnica.

Problem angažovanja viljuškara povezan je s činjenicom da preveliko ili premalo angažovanje viljuškara u različitim fabričkim procesima dovodi do značajnih finansijskih i tržišnih gubitaka. Trenutno se proces donošenja odluka gde će i šta će koji viljuškar raditi, bazira na odlukama stručnjaka (menadžera). Stručne odluke zasnovane su na njihovom iskustvu, bez pomoći bilo kakvog sistema za podršku odlučivanju (DSS). Brojni empirijski dokazi sugerišu da ljudska intuitivna procena i donošenje odluka često nisu optimalni, posebno u uslovima složenosti i stresa (Druzdzel i Flynn, 2002). Ovo naglašava važnost uključivanja sistema za podršku odlučivanju (DSS) koji pomažu stručnjacima u procesu donošenja odluka.

U ovoj aplikaciji odabrali smo nekoliko ML algoritama za pomoć u optimizaciji operacija otpremnog skladišta. ML algoritmi sastavljeni su u jedinstveni okvir za donošenje odluka koji služi kao DSS za menadžere i stručnjake u određenoj kompaniji. Štaviše, celi DSS za donošenje odluka može se posmatrati kao AI platforma, budući da stalno preračunava predloge iz nekoliko ML modela (koliko i koje viljuškare koristiti) i automatski bira one najbolje, s obzirom na dostavljene unose operatera.

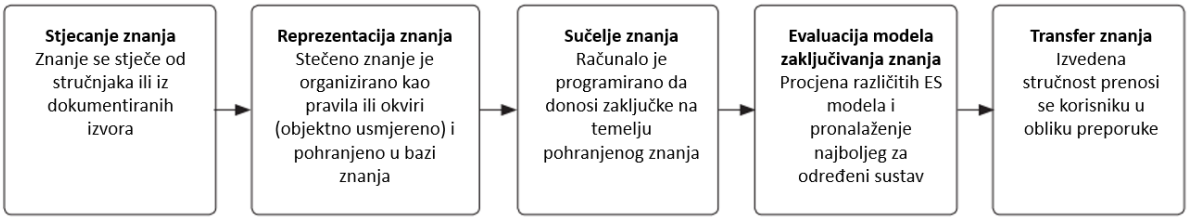
### Opis problema

Proces utovara ključan je za skladišnu logistiku i utiče na nivo tržišne usluge. Tokom otpreme, skladišni stručnjak određuje broj i vrši izbor viljuškara za utovar, vođen sa tri faktora: (1) dovršetak utovara unutar navedenog vremenskog okvira, (2) minimiziranje ometanja drugih zadataka viljuškara i (3) usklađivanje upotrebe viljuškara s mogućnostima održavanja, koje mogu podneti dva remonta istovremeno. Svaki viljuškar prolazi četiri do pet remonta godišnje.

Viljuškari su vitalni za procese utovara, koji moraju podržati marketinšku strategiju kompanije, a istovremeno osigurati nesmetanu realizaciju ostalih aktivnosti. Pogrešna raspodela viljuškara može dovesti do neiskorišćenosti resursa ili naštetiti ugledu kompanije i nivou usluge. Kašnjenja u utovaru povlače kazne. Menadžer mora koordinirati korišćenje viljuškara u svim aktivnostima kako bi izbegao istovremene remonte i upravljao različitim potrebama održavanja. Iako menadžeri obično donose tačne odluke, okruženje visokog stresa može dovesti do grešaka. Stoga je DSS potreban za povećanje poverenja i pouzdanosti donošenja odluka.

## AI i ML kao DSS za centralno skladište

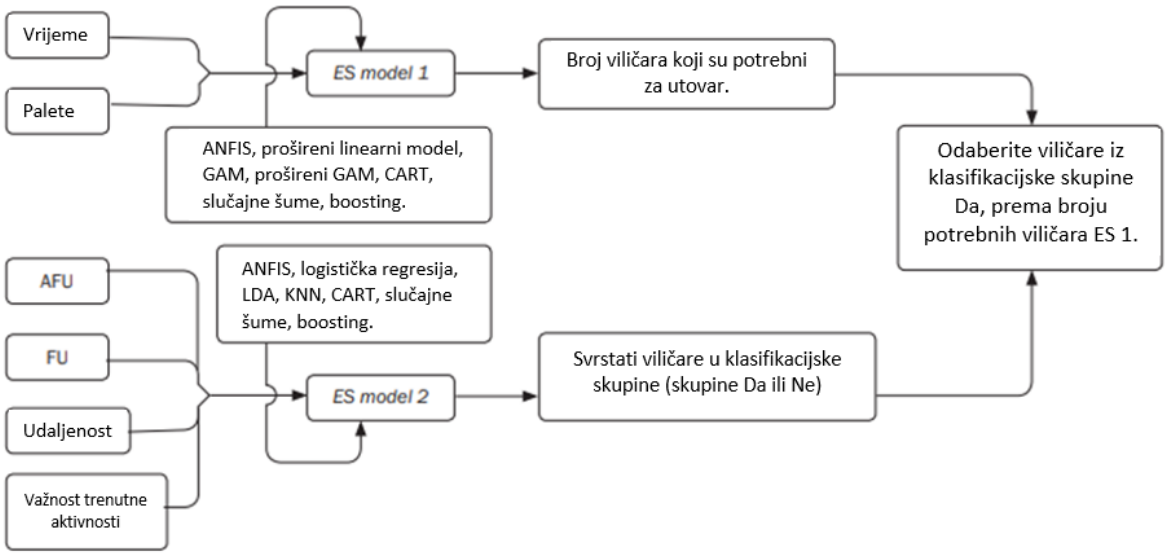
Prvi korak za generisanje AI i ML sistema je sticanje stabilnog i ispravnog izvora znanja (baza podataka) i rasvetljavanje poslovnih uloga koje treba podržavati. Stoga slika 10.3 predstavlja metodologiju za izgradnju AI i ML DSS sistema.




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

Sticanje znanja ostvareno je razgovorima s menadžerima, posmatranjem njihovih procesa donošenja odluka i pregledom skladišne dokumentacije. Za razvoj DSS-a za navedeni problem uspostavljene su dve baze znanja. Prva baza znanja uključuje odluke o broju viljuškara raspoređenih u zoni utovara (434 stručne odluke), dok druga pokriva koji su viljuškari korišćeni (368 stručnih odluka) u različitim operativnim scenarijima. Tokom faze zaključivanja znanja primenjeno je nekoliko ML algoritama pomoću Matlab softvera: Adaptivni neuro-fuzzy sistem zaključivanja (ANFIS), generalizovani aditivni modeli (GAM), slučajne šume, boosting, klasifikaciona i regresiona stabla (CART), proširena linearna regresija, logistička regresija, k-najbliži susedi (KNN) i linearna diskriminantna analiza (LDA). Procenjeni su različiti ML modeli i identifikovani su oni s najboljim performansama. ANFIS i CART pokazali su vrhunske rezultate i odabrani su kao konačni DSS-ovi za praktičnu primenu

u kompaniji. Prenos znanja je omogućen kroz korisnički interfejs finalnih DSS modela. Struktura i logika DSS-a ilustrovana je na slici 10.4.




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

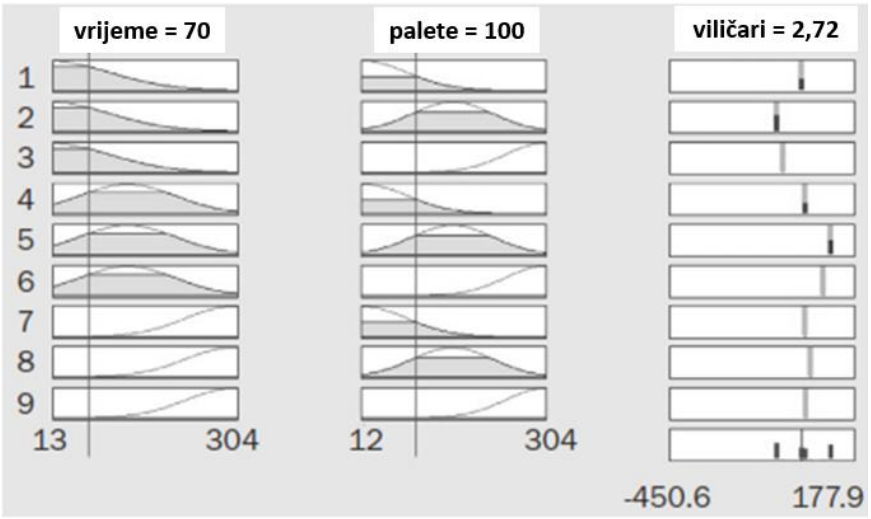
DSS okvir sastoji se od ulaznog sloja koji se sastoji od nekoliko ključnih faktora koji utiču na uključivanje viljuškara. ML sloj ima ML modele koji ponovno izračunavaju predlog o tome koliko i koje viljuškare koristiti u datom scenariju unosa. Modeli s najboljim učinkom biraju se kao modeli ekspertskih sistema (ES modeli) budući da se baza znanja na kojoj se formiraju ML modeli izdvaja od stručnjaka. Prvi model fokusiran je na određivanje broja viljuškara potrebnih u zoni utovara (ES model 1). Drugi model bavi se problemom izbora pojedinačnih viljuškara koji će se angažovati (ES model 2). Oba modela razvijena su korišćenjem nadziranih tehnika mašinskog učenja. Prema Turbanu i dr. (2005.), mašinsko učenje je pokazalo izvrsne rezultate u dizajniranju inteligentnih sistema za podršku odlučivanju (DSS). ES modeli šalju signale (ML sugestije i predloge) dalje u operaciju sortiranja, gde svaki viljuškar koji je klasifikovan u grupi sortiranja „Da", može biti angažovan u određenom procesu utovara.

Konsultacijama su identifikovani faktori koji utiču na odluke menadžera. Za određivanje broja viljuškara u zoni utovara, ključni faktori su navedeno vreme utovara (15 do 135 minuta) i

količina tereta (15 do 225 paleta). Prilikom izbora viljuškara koje će angažovati, menadžer uzima u obzir važnost trenutne aktivnosti (ocenjuje se od 1 do 9 prema politici kompanije), stepen iskorišćenja viljuškara, njegovu udaljenost od zone za utovar i prosečni stepen iskorišćenja svih viljuškara. Svaki viljuškar ima određeni broj radnih sati pre nego što je potreban remont, a njegova je uporaba ograničena izvan tog ograničenja. Iskorišćenje viljuškara (engl. *Forklift Utilization* - FU) je procenat radnih sati koje koristi pojedinačni viljuškar, dok je prosečna iskorišćenost viljuškara (engl. *Average Forklift Utilization* - AFU) prosečno radno vreme svih viljuškara. Veći AFU sugeriše da će većini viljuškara uskoro trebati remont.

### Korisnički interfejs DSS-a

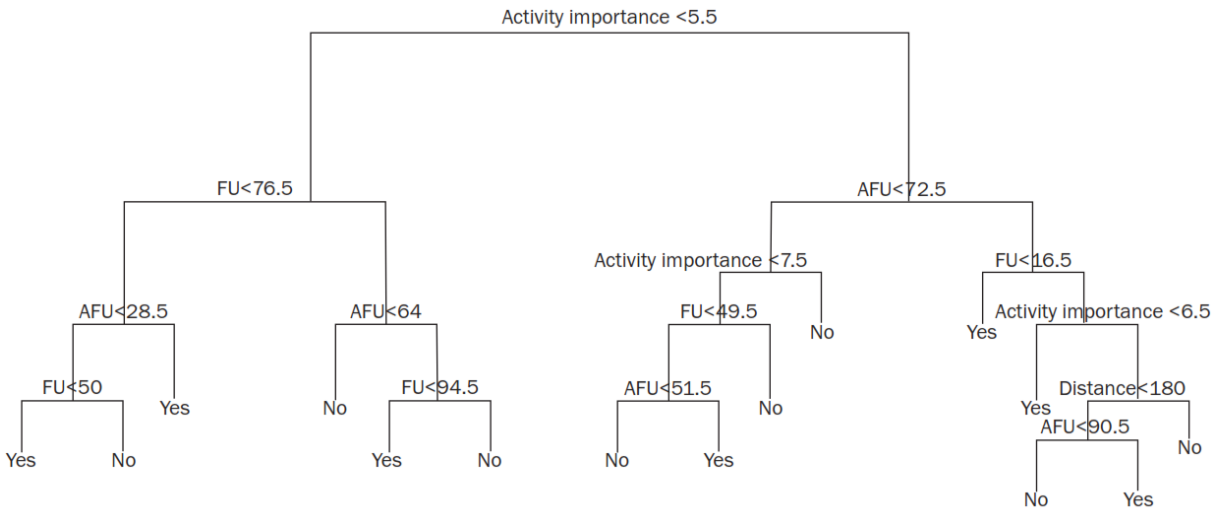
U većini situacija unosa najbolja izvedba je pokazana putem ANFIS-a i CART-a. Shodno tome, oni su odabrani kao pokretači datog DSS-a i njegovih ES-ova. Korisnički interfejs ES modela 1 prikazan je na slici 10.5, a operaterima omogućava brzo i jednostavno donošenje odluka o broju viljuškara za postavljanje jednostavnim pomeranjem vertikalne linije kroz domen ulaznih varijabli, na osnovu navedenog vremena utovara i količine tereta.




**Slika 4.62Sistem neizrazitog zaključivanja ES modela 1.**

ES Model 2 služi kao dopunski alat ES Modelu 1, poboljšavajući donošenje odluka dajući informacije o tome da li treba određeni viljuškar biti postavljen u zoni utovara (slika 10.6). Uzimajući u obzir položaj viljuškara (udaljenost od zone utovara), njegovu trenutnu aktivnost (važnost aktivnosti), njegovu iskorišćenost radnih sati (FU) i prosečnu iskorišćenost svih viljuškara (AFU), korisnici mogu lako utvrditi da li je određeni viljuškar prikladan za utovar ili

treba odabrati neki drugi. CART stablo odlučivanja je jednostavno za tumačenje, eliminišući potrebu za unosom vrednosti u softver. Umesto 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 viljuškara.**

Menadžeri mogu svakodnevno koristiti predstavljeni DSS, koji pomaže u postizanju boljeg odgovora lanca snabdevanja na zahteve kupaca i osiguravajući visoku verovatnoću isporuka na vreme. Predloženi AI & ML DSS pokazao je uspešne rezultate u sticanju "know-how" stručnog znanja i hvatanju "logike zaključivanja". Korišćenjem ove metode, stručnost menadžera može se izdvojiti i primeniti na druge skladišne procese. Ovo je posebno važno i vredno za praktičare jer je angažovanje stručnjaka za skladištenje često skupo. Osim toga, DSS takođe može poslužiti kao alat za obuku menadžera početnika, pomažući im da steknu iskustvo i poboljšaju svoje veštine donošenja odluka tokom vremena. Stoga su AI i ML sistemi 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 TABELA

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 -