STATISTIČNE METODE ZA ANALIZO LOGISTIČNIH PODATKOV

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

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

Sofinancira Evropska unija

Financira Evropska unija.
Izražena stališča in mnenja so izključno stališča in mnenja avtorjev in ne odražajo nujno stališč Evropske unije ali Izvajalske agencije za izobraževanje in kulturo (EACEA).
Evropska unija in agencija EACEA ne moreta biti odgovorni zanje.

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

STATISTIČNE METODE ZA ANALIZO LOGISTIČNIH PODATKOV

Poznan 2025

Založnik:
Wyższa Szkoła Logistyki
Estkowskiego 6
61-755 Poznan, Poland
www.wsl.com.pl

ISBN 978-83-62285-73-0 (online)

Poznan 2025, Issue I

Prevodi:
- prof. Dr. Rebeka Kovačič Lukman, University of Maribor, Faculty of Logistics
- asist. Kristijan Brglez, University of Maribor, Faculty of Logistics
- prof. Dr. Maja Fošner, University of Maribor, Faculty of Logistics
- doc. Dr. Benjamin Marcen, University of Maribor, Faculty of Logistics
- prof. Dr. Roman Gumzej, University of Maribor, Faculty of Logistics

Recenzenti:
- prof. Dr. Ajda Fošner, University of Primorska, Faculty of Management
- prof. Dr. Nikša Alfirević, University of Split, Faculty of Economics

Tehnični urednik: Kristijan Brglez, Fakulteta za logistiko Univerze v Mariboru, Slovenija

Oblikovanje naslovnice: Michał Adamczak, Poznan School of Logistics, Poznan, Poland

Monografija je bila napisana v sklopu projekta Business Analytics Skills for the Future-proof Supply Chains (BAS4SC) [2022-1-PL01-KA220-HED-000088856], financiranega s strani Erasmus+ Programme.

Financirano s strani Evropske unije. Izražena stališča in mnenja so zgolj stališča in mnenja avtorja(-ev) in ni nujno, da odražajo stališča in mnenja Evropske unije ali Evropske izvajalske agencije za izobraževanje in kulturo (EACEA). Zanje ne moreta biti odgovorna niti Evropska unija niti EACEA.

Monografija je prosto dostopna na spletni strani: enauka.put.poznan.pl

Predgovor

Področje logistike je vse bolj odvisno od vpogledov, ki temeljijo na podatkih, da bi optimizirali delovanje, zmanjšali stroške in zagotovili učinkovitost v oskrbovalnih verigah. Ta učbenik, *Statistične metode za analizo logističnih podatkov,* služi kot ključni vir za študente in strokovnjake, saj jih opremlja s potrebnimi znanji za obvladovanje kompleksnosti sodobnega upravljanja oskrbovalnih verig. Knjiga, ki je nastala v okviru projekta BAS4SC (Business Analytics Skills for Future-proof Supply Chains), ponuja celovit pregled statističnih metod, upravljanja podatkov in naprednih analitičnih tehnik, ki so izrecno prilagojene logistični industriji.

Ta učbenik s skrbno raziskavo odpravlja vrzeli v trenutni izobraževalni ponudbi, saj združuje teoretično znanje s praktično uporabo. V desetih poglavjih so podrobno obravnavane ključne teme, kot so napovedovanje povpraševanja, simulacijsko modeliranje, regresijska analiza ter vključevanje umetne inteligence in strojnega učenja v logistične operacije. Z uporabo splošno priznanih orodij, kot so SPSS, R in SQL, je vsebina tega učbenika zasnovana tako, da premošča vrzel med akademskim učenjem in potrebami industrije.

Upamo, da bo ta učbenik zagotovil temeljno znanje o poslovni analitiki za logistiko in spodbudil inovacije na tem področju ter spodbudil bodoč vodstven kader, ki se bo znal spopasti z izzivi hitro razvijajoče se oskrbovalne verige.

Prof. Dr. Sanja Bojić
Kristijan Brglez
Prof. Dr. Maja Fošner
Prof. Dr. Roman Gumzej
Prof. Dr. Rebeka Kovačič Lukman
Assist. prof. Dr. Benjamin Marcen
Prof. Dr. Marinko Maslarić
Assist. prof. Dr. Boško Matović
Dr. Dejan Mirčetić

KAZALO VSEBINE

UVOD

Ta učbenik z naslovom *Statistične metode za analizo logističnih podatkov* je tretji v seriji, ki je nastala v okviru projekta BAS4SC (Business Analytics Skills for Future-proof Supply Chains). Za določitev vsebine tega učbenika je bilo izvedenih več predhodnih raziskovalnih dejavnosti. Najprej je bila izvedena obsežna raziskava, ki je preučila predmete poslovne analitike, njihovo vsebino in veščine, ki jih posredujejo študentom logistike v Evropski uniji, Združenih državah Amerike in Združenem kraljestvu. Ta analiza je razkrila vrzel med znanjem in statističnimi spretnostmi, ki se zahtevajo na področju logistike, in tistimi, ki so trenutno na voljo študentom. Na podlagi poglobljenih intervjujev z univerzitetnim pedagoškim osebjem, študenti in strokovnjaki iz panoge je bilo kot bistvenih prepoznanih več kot 100 veščin poslovne analitike. Z metodo razvrščanja ABC je bilo za vključitev v to knjigo izbranih 33 veščin, ki se osredotočajo predvsem na matematiko, računalništvo, upravljanje, uporabno matematiko in statistiko. Združevanje teh opredeljenih potreb in veščin je privedlo do oblikovanja desetih vsebinskih poglavij, ki obravnavajo najpomembnejše veščine, potrebne na tem področju.

Prvo poglavje zajema uvodno statistiko in zagotavlja celovit pregled statističnih konceptov in njihove uporabe, zlasti pri analizi oskrbovalnih verig. Na začetku je poudarjena ključna vloga statistike pri optimizaciji oskrbovalnih verig. Uporablja opisno statistiko, kot so povprečje, mediana in standardni odklon, za analizo dobavnih rokov, ravni zalog in stroškov. V poglavju so predstavljene napovedne tehnike, kot sta regresija in analiza časovnih vrst za napovedovanje povpraševanja in zalog. Nadalje raziskuje pomen spremenljivk, razlikuje med kvalitativnimi in kvantitativnimi vrstami ter se poglobi v osnovne statistične mere, kot so povprečje, mediana, način, variance in standardni odklon.

Poleg tega obravnava grafične metode predstavitve podatkov, kot so histogrami in razpršeni diagrami, ter poudarja razliko med opisno in sklepno statistiko. Na koncu so predstavljeni ključni koncepti korelacije, regresije in verjetnostnih porazdelitev, ki ponujajo orodja za razumevanje odnosov med spremenljivkami in modeliranje naključnih pojavov v podatkih. Te statistične tehnike pomagajo izboljšati odločanje v oskrbovalni verigi, učinkovitost in obvladovanje tveganj.

Drugo poglavje, Statistika za poslovno analitiko, obravnava bistvene statistične koncepte in tehnike za pridobivanje vpogleda v poslovne podatke. Na začetku predstavi pomen analize podatkov pri poslovnem odločanju. Razloži temeljno vlogo normalne porazdelitve, ki služi kot osnova za številne statistične metode. Poglavje obravnava standardni odklon in poudarja njegov pomen pri merjenju

variabilnosti podatkov. Obravnava tudi vzorčne porazdelitve in osrednji mejni stavek ter pojasnjuje, kako iz vzorčnih podatkov sklepati na parametre populacije. Obravnavane so teme, kot so preverjanje hipotez, Z-skupine in t-skupine, ki so v pomoč pri sprejemanju odločitev in izračunavanju verjetnosti. Poglavje se zaključi z obravnavo standardne napake in intervalov zaupanja, ki pomagajo kvantificirati negotovost ocen. Na koncu poglavje bralce opremi s statističnimi orodji, potrebnimi za poslovno analitiko, kar jim omogoča sprejemanje informiranih in na podatkih temelječih odločitev.

Poglavje o upravljanju podatkov obravnava različne vidike upravljanja podatkov v logistiki, pri čemer se osredotoča na oblike, organizacijo in tehnologije podatkov. Začne se z vlogo elektronske izmenjave podatkov (EDI) pri izmenjavi informacij v oskrbovalnih verigah s standardiziranimi alfanumeričnimi oblikami. Pojasnjuje koncept informacij, podatkov in znanja ter obravnava, kako se podatki digitalizirajo in organizirajo v podatkovne zbirke, skladišča in baze znanja. Poglavje obravnava logistične podatke, zlasti uporabo črtnih kod in oznak RFID za identifikacijo in sledenje v logistiki. Predstavlja tudi tehnike organizacije podatkov, od preglednic do relacijskih podatkovnih zbirk (RDB), in razlaga ključne pojme, kot so primarni in tuji ključi, normalizacija in poizvedovalni jeziki, kot je SQL. Poleg tega poglavje obravnava najboljše prakse za filtriranje podatkov in preprečevanje napak pri vnosu podatkov. Na koncu so obravnavane razlike med podatkovnimi skladišči in bazami znanja ter poudarjena njihova vloga pri poslovni analizi in odločanju.

Poglavje Simulacijsko modeliranje in analiza (SMA) se osredotoča na ustvarjanje digitalnih modelov za simulacijo realnih sistemov za optimizacijo in odločanje. Začne se z razlago Conant-Ashbyjevega teorema, ki nakazuje, da mora simulacijski model ustrezati kompleksnosti svojega ustreznika v realnem svetu, da ga lahko učinkovito regulira. SMA optimizira oskrbovalne verige in prometna omrežja v logistiki, pri čemer omogoča menedžerjem, da simulirajo in ovrednotijo različne scenarije. V poglavju so opisane ključne simulacijske metodologije, kot so diskretna simulacija dogodkov (DES) za procesno usmerjeno analizo, sistemska dinamika (SD) za delovanje sistema na visoki ravni, simulacija na podlagi agentov (ABS) za modeliranje vedenja posameznih subjektov in omrežna simulacija (NS) za analizo omrežnih tokov. Vsaka metoda omogoča vpogled v različne vidike logistike, od proizvodnih ciklov do optimizacije prometa. Poglavje se zaključi s pregledom logističnih simulacijskih projektov, ki so strukturirani v skladu z načrtovanjem za šest sigme in Demingovim ciklom izboljšav ter pomagajo pri načrtovanju, izvajanju in izpopolnjevanju kompleksnih logističnih sistemov.

Poglavje Linearna regresija z enim in več regresorji predstavlja regresijsko analizo za razumevanje razmerja med odvisnimi in neodvisnimi spremenljivkami. Začne se z enostavno linearno regresijo,

kjer ena neodvisna spremenljivka, kot so izdatki za oglaševanje, napoveduje rezultat, kot je prodaja. Poglavje pojasnjuje sestavo regresijskega modela in regresijsko enačbo, ki se uporablja za napovedovanje odvisne spremenljivke na podlagi vzorčnih podatkov. Obravnava tudi metodo najmanjših kvadratov, ki je bistvena tehnika za ocenjevanje regresijske premice z zmanjševanjem napak pri napovedovanju. Nato je predstavljen koeficient determinacije (R^2) za merjenje, kako dobro se regresijski model ujema s podatki. Poglavje se nato ukvarja z multiplo regresijo, pri kateri dve ali več neodvisnih spremenljivk napoveduje odvisno spremenljivko, kar omogoča celovitejšo analizo. Primeri vključujejo napovedovanje časa potovanja na podlagi razdalje in števila dostav.

Poglavje Uvod v operacijske raziskave se osredotoča na uporabo analitičnih metod za izboljšanje odločanja, zlasti na področju logistike in upravljanja oskrbovalne verige. Operativne raziskave uporabljajo modeliranje, statistiko in tehnike optimizacije za iskanje optimalnih rešitev zapletenih problemov, kar omogoča učinkovito upravljanje virov, nadzor zalog in optimizacijo procesov. Poglavje poudarja strateško logistično načrtovanje, ki vključuje metode, kot sta Six Sigma in proizvodnja Just-in-Time, za izboljšanje operativne učinkovitosti. Poslovna inteligenca (BI) in poslovna analitika (BA) sta ključnega pomena pri analizi podatkov, saj podjetjem omogočata sprejemanje informiranih odločitev z uporabo napovedovanja, napovedne analitike in vizualizacije podatkov. V poglavju je predstavljeno tudi večkriterijsko odločanje (MCDM), ki pomaga pri ocenjevanju in izbiri optimalnih rešitev na podlagi različnih meril. Sistemi za podporo odločanju (DSS) in na znanju temelječi inženiring (KBE) pomagajo pri vključevanju znanja in podatkov v procese odločanja, kar dodatno izboljša operativno učinkovitost in strateško načrtovanje.

Poglavje o statistični obdelavi podatkov s SPSS predstavlja IBM-ovo programsko opremo SPSS kot zmogljivo orodje za avtomatizacijo zapletenih statističnih analiz, povečanje zanesljivosti in olajšanje odločanja. Pojasnjuje, kako SPSS omogoča uvoz, manipulacijo in pripravo podatkov prek uporabniku prijaznega vmesnika. Poglavje zajema ključne funkcionalnosti, kot so opisna statistika, izdelava grafov in vizualizacija podatkov. Predstavlja tudi temeljne statistične teste - teste, korelacijo, hi-kvadrat in ANOVA - in bralce vodi skozi nastavitev in razlago vsakega testa. Poleg tega raziskuje orodja za upravljanje podatkov, kot so združevanje, delitev in izračunavanje spremenljivk v podatkovnih nizih, ter prikazuje, kako SPSS izboljšuje statistično analizo v logistiki in na drugih področjih.

Poglavje 8 obravnava poslovno analitiko (BA) in njeno uporabo z orodji, kot sta R, in SQL, za reševanje poslovnih problemov. Cilj BA je izboljšati odločanje in uspešnost podjetja z uporabo metod, ki temeljijo na podatkih. Vključuje deskriptivne, prediktivne in preskriptivne platforme za analiziranje podatkov in sprejemanje utemeljenih odločitev. R je predstavljen kot robustno

odprtokodno orodje za statistično analizo in vizualizacijo, medtem ko je SQL bistven za upravljanje in poizvedovanje po velikih podatkovnih bazah. Poglavje podrobno opisuje integracijo R in SQL za učinkovito poslovno analitiko in poudarja, kako je mogoče podatke, shranjene v SQL, analizirati z uporabo skript R za avtomatizacijo nalog. Praktični primeri, kot je poizvedovanje po podatkovni zbirki Chinook, ponazarjajo, kako R in SQL sodelujeta pri ustvarjanju vpogledov, na primer pri prepoznavanju najbolje prodajanih albumov. Ta sinergija med BA, R in SQL izboljša sposobnost upravljanja in analiziranja dinamičnih poslovnih podatkov.

Poglavje 9 se osredotoča na napovedovanje povpraševanja v oskrbovalni verigi, vizualizacijo in inženiring funkcij. Napovedovanje povpraševanja napoveduje potrebe strank, spreminja celotno oskrbovalno verigo in zmanjšuje logistične stroške. V poglavju so opisani ključni koraki za napovedovanje povpraševanja, vključno z opredelitvijo problema, zbiranjem podatkov, analizo trendov, izbiro modelov in njihovim vrednotenjem. Vizualizacija pomaga prepoznati vzorce, kot so sezonskost in trendi, ki so lahko podlaga za izbiro modela. S-ARIMA (Seasonal Autoregressive Integrated Moving Average) je poudarjen kot učinkovit model za obdelavo kompleksnih podatkov časovnih vrst, zlasti s sezonskimi vzorci povpraševanja. Primer iz živilske industrije prikazuje sposobnost modela S-ARIMA za napovedovanje povpraševanja in usmerjanje odločanja. Na koncu poglavje obravnava, kako preizkusiti in potrditi modele napovedovanja, da se zagotovi njihova učinkovitost v realnih aplikacijah, pri čemer se za oceno uspešnosti uporabljajo metrike, kot sta RMSE in MAPE.

Zadnje poglavje obravnava vlogo umetne inteligence in strojnega učenja v oskrbovalnih verigah, začenši s pregledom razvoja umetne inteligence od simbolnih sistemov do sodobnih pristopov strojnega učenja. Umetna inteligenca se nanaša na avtomatizacijo nalog, ki običajno zahtevajo človeško inteligenco, pri čemer je umetna inteligenca podmnožica umetne inteligence, ki se osredotoča na učenje vzorcev iz podatkov. Poglavje poudarja, kako se modeli AI/ML, kot sta nadzorovano in nenadzorovano učenje, uporabljajo za reševanje poslovnih problemov, vključno z napovedovanjem povpraševanja, upravljanjem zalog in optimizacijo. Bistvena študija primera vključuje uporabo algoritmov AI in ML v osrednjem skladišču tovarne hrane za optimizacijo uporabe viličarjev. Z vključitvijo sistemov za podporo odločanju (DSS) modeli AI/ML pomagajo vodjem pri izbiri optimalnega števila viličarjev, kar izboljša operativno učinkovitost. Poglavje poudarja, kako lahko umetna inteligenca/ML zajame strokovno znanje, zmanjša stroške in izboljša procese odločanja pri upravljanju oskrbovalne verige.

1. Uvodna statistika

1.1 Vloga in pomen statistike pri analizi podatkov v oskrbovalnih verigah

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

Opišimo nekaj pomembnih vlog statistike pri analizi oskrbovalne verige.

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

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

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

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

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

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

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

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

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

1.2 Osnovni pojmi statistike

Spremenljivke

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

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

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

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

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

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

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

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

1.3 Osnovni statistični koncepti s primeri

Povprečje (srednja vrednost)

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

Izračun povprečja:

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

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

Primer:

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

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

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

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

Mediana

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

Kako izračunati mediano:

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

Primer:

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

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

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

Modus

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

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

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

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

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

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

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

Razpon variance (VR, Razpon, razpon)

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

Izračun meje variacije je preprost:

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

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

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

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

Varianca in standardni odklon

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

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

Izračun standardnega odklona:

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

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

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

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

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

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

Varianca je povprečje teh vrednosti:

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

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

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

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

Kvantili

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

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

1.4 Prikaz statističnih podatkov

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

Tukaj je nekaj pogostih načinov za prikaz statističnih podatkov:

Tabele

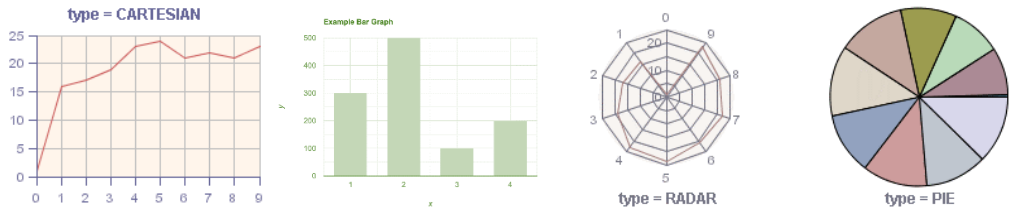
Tabele so osnovna metoda za prikazovanje podatkov. Primeri so tabele pogostosti, ki prikazujejo število pojavitev različnih vrednosti, in podatkovne tabele, ki prikazujejo več informacij o podatkih.

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

Slika 1.1 Primer tabele.

Grafični prikazi

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



Slika 1.2 Primeri grafičnih predstavitev podatkov.

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

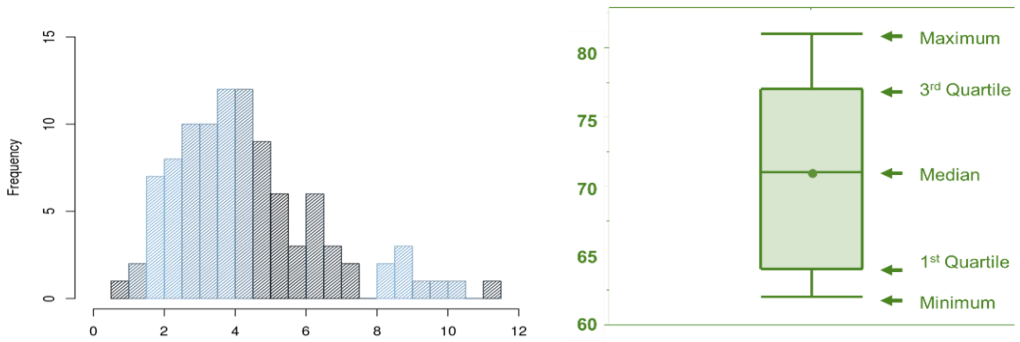
Stolpčni diagrami se uporabljajo za primerjavo količin v različnih kategorijah, zato so idealni za predstavitev diskretnih podatkov. Posebej učinkoviti so za poudarjanje razlik, podobnosti in trendov med skupinami. Stolpčni diagrami se pogosto uporabljajo, kadar je treba jasno in vizualno preprosto prikazati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

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

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

Histogrami

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



Slika 1.3 Histogram in kvantilni graf (škatlasti graf).

Kvantilni grafikon (okvirni grafikon)

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

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

1.5 Porazdelitev pogostosti

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

Postopek ustvarjanja frekvenčne porazdelitve vključuje naslednje korake:

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

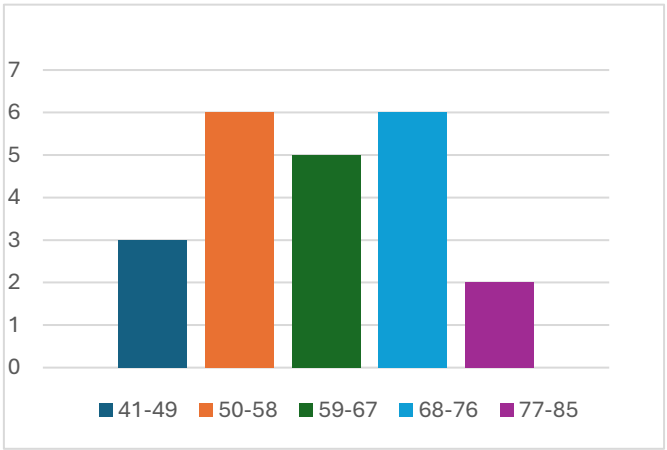
- Ustvarite frekvenčno tabelo: ustvarite tabelo, ki prikazuje vse različne vrednosti spremenljivke in število pojavitev za vsako vrednost.
- Risanje histograma: če imate veliko število različnih vrednosti, lahko ustvarite histogram, ki prikazuje frekvenčno porazdelitev. To je grafični prikaz, ki prikazuje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

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

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

Slika 1.4 Tabela porazdelitve pogostosti.

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



Slika 1.5 Graf porazdelitve pogostosti.

1.6 Opisna in sklepalna statistika

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

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

Inferenčna statistika

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

Glavni cilji inferenčne statistike so:

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

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

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

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

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

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

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

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

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

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

1.7 Korelacija in regresija

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

Korelacija

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

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

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

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

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

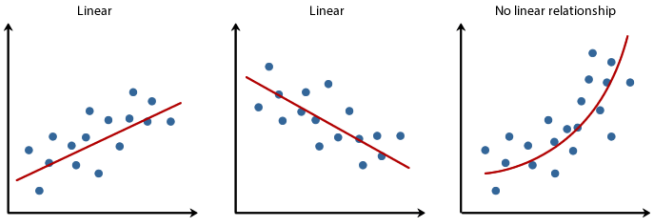
Regresija

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

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

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

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

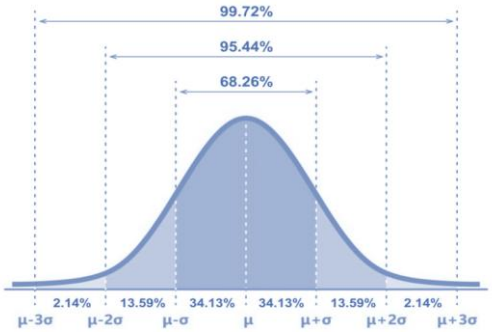


Slika 1.6 Grafi preproste linearne regresije.

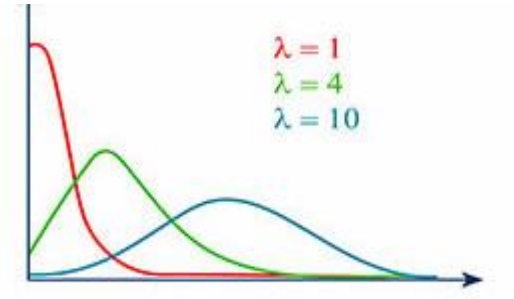
1.8 Verjetnostne porazdelitve

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

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



Slika 1.8 Graf normalne porazdelitve.



Slika 1.7 Graf Poissonove porazdelitve.

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

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

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

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

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

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

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

Literatura poglavje 1

- Introductory Statistics. Bentham Science Publishers, Kahl, A. (Publish 2023). DOI:10.2174/9789815123135123010l
- 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

Dodatne povezave do literature in videoposnetkov na YouTubu Poglavje 1

- https://open.umn.edu/opentextbooks/textbooks/196
- https://www.scribbr.com/category/statistics/
- https://stats.libretexts.org/Bookshelves/Introductory_Statistics
- https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
- https://saylordotorg.github.io/text_introductory-statistics/

- https://drive.uqu.edu.sa/_/mskhayat/files/MySubjects/20178FS%20Elementary%20Statistics/Introductory%20Statistics%20(7th%20Ed).pdf
- https://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 poslovno analitiko

Dobrodošli v svetu poslovne statistike, kjer se podatki spremenijo v pomembne vpoglede, ki usmerjajo odločanje in odkrivajo skrite resnice. V tem izčrpnem raziskovanju se podajamo na pot, na kateri razkrivamo bistvene statistične koncepte in tehnike, ki so osnova za temeljito analizo poslovnih podatkov. Od razumevanja zapletenosti porazdelitev do uporabe testiranja hipotez in oblikovanja intervalov zaupanja - vsako poglavje razkriva nov vidik statistične pismenosti.

V središču statistične analize je normalna porazdelitev, krivulja v obliki zvona, ki je značilna za številne pojave v naravi in človeškem vedenju. V tem delu se bomo poglobili v bistvo normalne porazdelitve, razkrili njene lastnosti in pomen pri statističnem sklepanju. Z vizualizacijami in primeri iz resničnega sveta osvetlimo vseprisotnost te temeljne porazdelitve in njeno vlogo temeljnega kamna statistične teorije.

Standardni odklon služi kot kompas v statistični pokrajini, ki nas vodi skozi spremenljivost, značilno za zbirke podatkov. V tem poglavju razčlenjujemo koncept standardnega odklona in razkrivamo njegov pomen pri količinskem opredeljevanju razpršenosti in ocenjevanju razširjenosti podatkovnih točk. Oboroženi z globljim razumevanjem standardnih odklonov boste samozavestno krmarili po podatkih ter natančno prepoznavali vzorce in odstopanja.

Spremenljivke so gradniki statistične analize, vsaka od njih pa ima posebne značilnosti in posledice. To poglavje pojasnjuje dihotomijo med zveznimi in diskretnimi spremenljivkami ter osvetljuje njihovo vlogo pri modeliranju in razlagi podatkov. Z razumevanjem nians vrst spremenljivk boste izkoristili celoten potencial statističnih tehnik, prilagojenih različnim podatkovnim strukturam.

Vzorčna porazdelitev je temelj statističnega sklepanja, saj zapolnjuje vrzel med vzorčnimi opazovanji in populacijskimi parametri. V tem poglavju razkrivamo koncept vzorčne porazdelitve in pojasnjujemo njen pomen pri podajanju verjetnostnih izjav o značilnostih populacije. S konkretnimi primeri boste razvili intuitivno razumevanje vloge porazdelitve vzorčenja pri zanesljivi statistični analizi.

Centralna mejna trditev je ključni koncept v statistiki, ki nam pomaga razumeti negotovost. V tem poglavju je na preprost način razložen Centralni mejni teorem, ki pokaže, kako naredi vzorčne sredine bolj predvidljive in pomaga pri preverjanju hipotez. Z razumevanjem tega koncepta boste lahko iz podatkov potegnili smiselne zaključke.

Razumevanje testiranja hipotez je bistveno za sprejemanje odločitev, ki temeljijo na podatkih. Z njim lahko ugotovimo, ali so opaženi vzorci v podatkih smiselni ali zgolj posledica naključja. S testiranjem hipotez lahko ocenimo predpostavke, primerjamo skupine in ocenimo statistično pomembnost rezultatov, zato je to pomembno orodje v znanstvenih raziskavah, poslovnih analizah in na številnih drugih področjih.

Rezultati Z in tabele Z služijo kot navigacijski pripomočki v morju standardne normalne porazdelitve ter olajšajo standardizirane primerjave in izračune verjetnosti. V tem poglavju so pojasnjene zapletene značilnosti Z-skupin, kar vam omogoča interpretacijo standardiziranih rezultatov in uporabo Z-tabel za statistično analizo. Z znanjem o Z-skupinah boste z gotovostjo in natančnostjo krmarili po prostranstvih normalne porazdelitve.

Kadar so vzorci majhni ali standardni odkloni populacije neznani, so t-skupine in t-tabele nepogrešljiva orodja za statistično analizo. To poglavje razkriva skrivnosti t-skupin in vas vodi skozi njihov izračun in interpretacijo s pomočjo t-tabel. Oboroženi s tem znanjem boste spretno krmarili po niansah t-distribucij in si zagotovili zanesljivo sklepanje v različnih statističnih scenarijih.

Normalna in t-distribucija sta stebra teorije verjetnosti, vsaka od njiju pa ima edinstvene značilnosti in uporabo. V tem poglavju pojasnjujemo razlike med tema porazdelitvama in vam omogočamo, da prepoznate, kdaj uporabiti vsako od njiju pri statistični analizi. S praktičnimi primeri in primerjalnimi analizami boste razvili podrobno razumevanje normalnih in t-razdelitev ter tako obogatili svoje statistično orodje.

Intervali zaupanja omogočajo vpogled v negotovost, ki spremlja populacijske parametre, in nam omogočajo količinsko opredelitev natančnosti ocen. V tem poglavju raziskujemo oblikovanje intervalov zaupanja za sredine in deleže ter razkrivamo metodologijo in razlago teh bistvenih statističnih orodij. Z obvladovanjem intervalov zaupanja boste pregledno in strogo izrazili negotovost, ki je neločljivo povezana z vašimi ugotovitvami.

Čeprav p-vrednosti omogočajo statistično sklepanje, lahko njihova napačna razlaga privede do napačnih sklepov in napačnih odločitev. V tem poglavju so obravnavane morebitne pasti prevelikega zanašanja na p-vrednosti, pri čemer je poudarjen pomen konteksta in velikosti učinka pri statistični analizi. S kritičnim pregledom in praktičnimi spoznanji boste previdno krmarili po zapletenosti p-vrednosti in tako zagotovili celovitost svojih statističnih zaključkov.

Na teh straneh se skrivajo ključi za razvozlavanje skrivnosti statistične analize, ki vam omogočajo, da se samozavestno in natančno orientirate po zapletenih podatkih. Ko se bomo skupaj podali na

to potovanje, naj bo radovednost naš kompas, poizvedovanje pa naša vodilna luč, ki nam bo osvetljevala pot do globljega razumevanja in uporabnih vpogledov.

2.1 Normalna porazdelitev

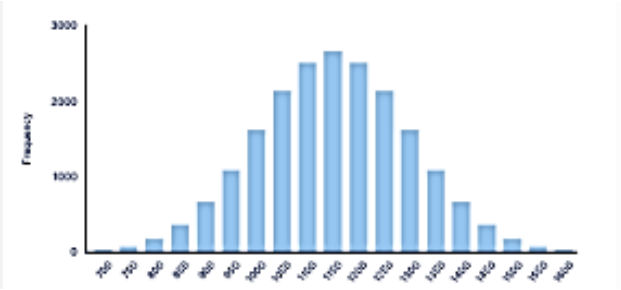
V središču statistične analize je normalna porazdelitev, vseprisotna verjetnostna porazdelitev, ki služi kot merilo za številne statistične tehnike. Spoznali bomo njene značilnosti, simetrično krivuljo v obliki zvona in njen pomen za razumevanje porazdelitve podatkov.

Normalna porazdelitev se uporablja na različnih področjih, vključno s financami, psihologijo, inženiringom in biologijo. Normalna porazdelitev je vsestransko orodje za analizo in razlago podatkov, od modeliranja borznih tečajev do razumevanja porazdelitve človeške višine.

V tem poglavju se bomo poglobili v matematične lastnosti normalne porazdelitve in raziskali, kako izračunati verjetnosti, percentile in z-rezultate. Poleg tega bomo obravnavali praktične tehnike za vizualizacijo in interpretacijo normalne porazdelitve z uporabo histogramov, diagramov gostote in kumulativnih porazdelitvenih funkcij.

Ob koncu tega poglavja boste dobro spoznali normalno porazdelitev in njen pomen v statistični analizi. Oboroženi s tem znanjem boste dobro opremljeni za reševanje naprednejših statističnih konceptov in njihovo uporabo v realnih podatkovnih nizih. Skupaj se podajmo na to potovanje, da razvozlamo skrivnosti normalne porazdelitve.

Normalna porazdelitev, znana tudi kot Gaussova porazdelitev ali krivulja zvona, kaže simetrično porazdelitev podatkov brez poševnosti. Če podatke narišemo na graf, tvorijo krivuljo v obliki zvona, pri čemer je večina vrednosti zbranih okoli središča in se z oddaljevanjem od njega zmanjšuje.



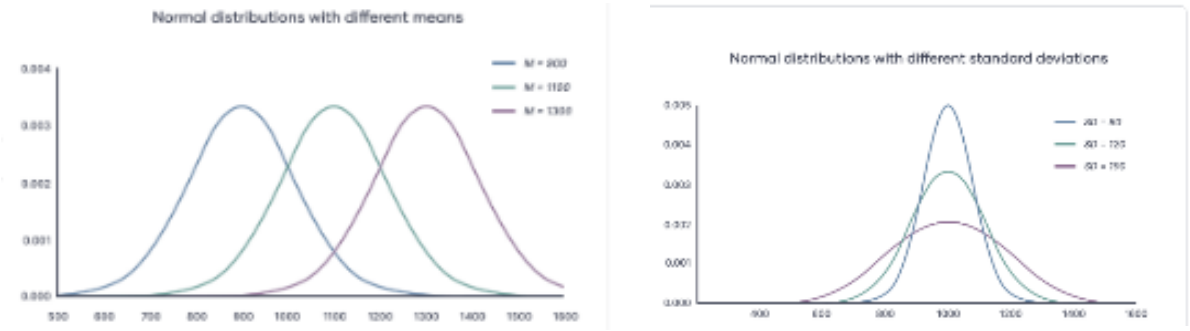
Slika 2.1 Primer Gaussove porazdelitve ali zvonaste krivulje.

Različne spremenljivke v naravoslovju in družboslovju imajo običajno normalno porazdelitev ali njen približek. Primeri so višina, porodna teža, bralne sposobnosti, zadovoljstvo na delovnem mestu in rezultati mature. Zaradi razširjenosti normalno porazdeljenih spremenljivk so številni statistični

testi prilagojeni takšnim populacijam. Spretno razumevanje značilnosti normalnih porazdelitev posameznikom omogoča uporabo inferenčne statistike za primerjavo skupin in izdelavo ocen populacije iz vzorcev.

Normalne porazdelitve imajo ključne značilnosti, ki jih zlahka opazimo v grafih:

- Povprečje, mediana in način so popolnoma enaki.
- Porazdelitev je simetrična glede na srednjo vrednost - polovica vrednosti je pod srednjo vrednostjo, polovica pa nad srednjo vrednostjo.
- Porazdelitev lahko opišemo z dvema vrednostma: srednjo vrednostjo in standardnim odklonom.



Slika 2.2 Normalna porazdelitev z različnimi srednjimi vrednostmi in z različnimi odstopanji.

Srednja vrednost služi kot lokacijski parameter, ki določa središče vrha krivulje. S prilagajanjem sredine se krivulja ustrezno premakne: z njenim povečanjem se krivulja premakne v desno, z njenim zmanjšanjem pa v levo. Medtem standardni odklon deluje kot parameter obsega in vpliva na razpon ali širino krivulje.

Standardni odklon raztegne ali stisne krivuljo. Majhen standardni odklon povzroči ozko krivuljo, velik standardni odklon pa široko krivuljo.

2.2 Empirično pravilo

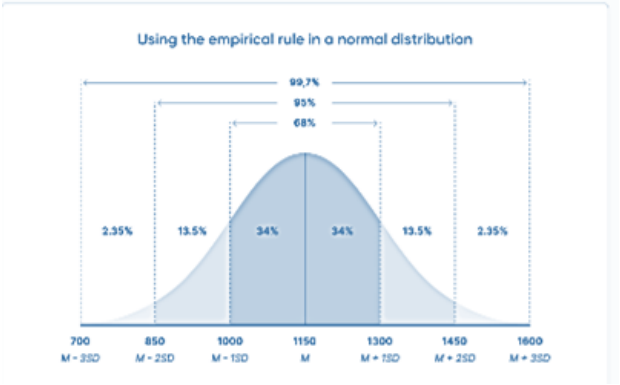
Empirično pravilo, znano tudi kot pravilo 68-95-99,7, omogoča vpogled v porazdelitev vrednosti znotraj normalne porazdelitve:

- Približno 68 % vrednosti je znotraj 1 standardnega odklona od povprečja.
- Približno 95 % vrednosti je znotraj 2 standardnih odklonov od povprečja.
- Približno 99,7 % vrednosti je znotraj 3 standardnih odklonov od povprečja.

Razmislite na primer o scenariju, v katerem so zbrani rezultati izpitov SAT, ki so jih dosegli učenci na novem tečaju za pripravo na izpite, in podatki ustrezajo normalni porazdelitvi s povprečnim rezultatom (M) 1150 in standardnim odklonom (SD) 150.

Z uporabo empiričnega pravila dobimo naslednja spoznanja:

- Približno 68 % rezultatov je v razponu od 1000 do 1300, kar ustreza enemu standardnemu odklonu nad in pod povprečjem.
- Približno 95 % rezultatov je v razponu od 850 do 1450, kar pomeni 2 standardna odklona nad in pod povprečjem.
- Skoraj vsi rezultati, približno 99,7 % so v razponu od 700 do 1600 točk, kar vključuje 3 standardne odklone nad in pod povprečjem.

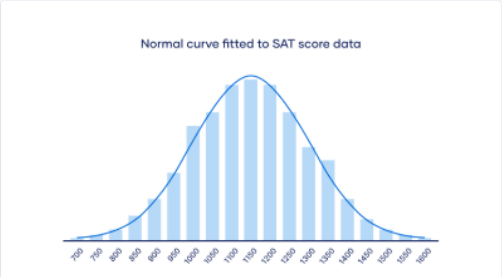


Slika 2.3 Empirično pravilo v normalni porazdelitvi.

Empirično pravilo je hitra metoda za ocenjevanje podatkov, ki omogoča odkrivanje odstopanj ali izjemnih vrednosti, ki odstopajo od pričakovanega vzorca. V primerih, ko podatki iz majhnih vzorcev znatno odstopajo od tega vzorca, so lahko primernejše alternativne porazdelitve, kot je t-razdelitev. Določitev porazdelitve spremenljivke omogoča uporabo ustreznih statističnih testov.

2.3 Formula normalne krivulje

Za konstruiranje normalne krivulje na podlagi dane srednje vrednosti in standardnega odklona lahko uporabimo funkcijo gostote verjetnosti, s čimer natančno predstavimo porazdelitev podatkov.



Slika 2.4 Normalna krivulja, prilagojena podatkom o rezultatih

V funkciji gostote verjetnosti območje pod krivuljo predstavlja verjetnost. Glede na to, da normalna porazdelitev služi kot verjetnostna porazdelitev, je kumulativna površina pod krivuljo vedno enaka 1 ali 100 %. Čeprav se zdi formula za normalno funkcijo gostote verjetnosti zapletena, je za njeno uporabo potrebno le poznavanje populacijske sredine in standardnega odklona. Z nadomestitvijo teh parametrov v formulo lahko določimo gostoto verjetnosti, povezano s katero koli vrednostjo x.

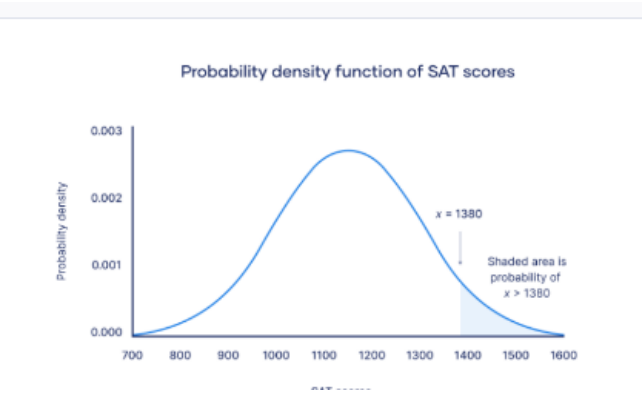
- *f(x)* = verjetnost
- *x* = vrednost spremenljivke
- *μ* = povprečje
- *σ* = standardni *odklon*
- σ^2 = varianta

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

Primer:

S pomočjo funkcije gostote verjetnosti želite izvedeti, kolikšna je verjetnost, da bodo rezultati mature SAT v vašem vzorcu presegli 1380 točk.

Na grafu funkcije gostote verjetnosti je verjetnost osenčeno območje pod krivuljo, ki leži desno od točke, kjer je vaš rezultat na testu SAT enak 1380.



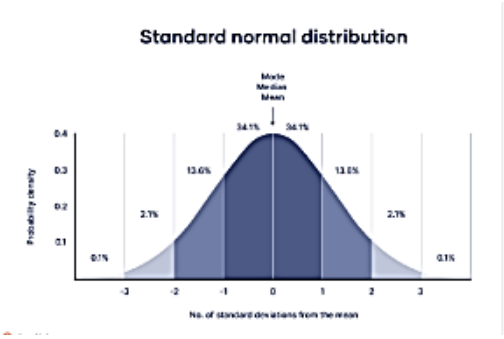
Slika 2.5 Graf funkcije gostote verjetnosti rezultatov na testu

Verjetnostno vrednost tega rezultata lahko določite s standardno normalno porazdelitvijo.

2.4 Standardna normalna porazdelitev

Standardna normalna porazdelitev, znana kot **z-razdelitev**, se razlikuje po tem, da ima srednjo vrednost 0 in standardni odklon 1. Vsako normalno porazdelitev lahko obravnavamo kot transformacijo standardne normalne porazdelitve s prilagoditvami v merilu, položaju ali oboje.

V kontekstu porazdelitve z se posamezna opazovanja, ki so v normalni porazdelitvi običajno označena z x, imenujejo z-vrednost. Te vrednosti z predstavljajo število standardnih odklonov, za katere posamezna vrednost odstopa od povprečja. Zato pretvorba vrednosti iz katere koli normalne porazdelitve v z-lestvice olajša primerjavo in analizo v okviru standardne normalne porazdelitve.



Slika 2.6 Graf standardne normalne porazdelitve.

Za določitev *z-rezultata vrednosti* morate poznati le srednjo vrednost in standardni odklon porazdelitve.

Razlaga formule Z-rezultata

- x = individualna vrednost
- μ = povprečje
- σ = standardni odklon

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

Normalne porazdelitve pretvorimo v standardno normalno porazdelitev iz več razlogov:

- Iskanje verjetnosti, da bodo opazovanja v porazdelitvi padla nad ali pod določeno vrednostjo.
- ugotavljanje verjetnosti, da se povprečje vzorca pomembno razlikuje od znanega povprečja populacije.

- Primerjava rezultatov na različnih porazdelitvah z različnimi srednjimi vrednostmi in standardnimi odkloni.

2.5 Ugotavljanje verjetnosti s pomočjo z-razdelitve

Vsakemu z-rezultatu ustreza verjetnost, pogosto imenovana p-vrednost, ki označuje verjetnost opazovanja vrednosti pod določenim z-rezultatom. S pretvorbo posamezne vrednosti v z-rezultatu lahko določimo verjetnost vseh vrednosti do te točke v okviru normalne porazdelitve.

Na primer, upoštevajte scenarij, v katerem želite ugotoviti, kolikšna je verjetnost, da bodo rezultati mature SAT v vašem vzorcu presegli 1380 točk. Na začetku izračunate z-vrednost s pomočjo povprečja in standardnega odklona porazdelitve. Pri srednji vrednosti 1150 in standardnem odklonu 150 z-vrednost pokaže število standardnih odklonov, za katere 1380 odstopa od povprečja.

Izračun formule

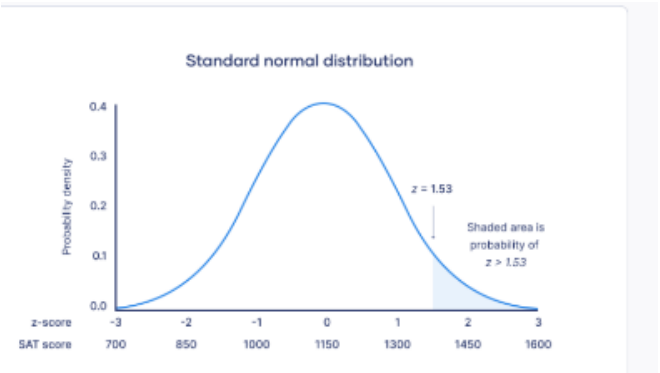
$$z = \frac{x - \mu}{\sigma} = \frac{1380 - 1150}{150} = 1.53$$

Za *z-vrednost* 1,53 je *p-vrednost* 0,937. To je verjetnost, da je rezultat SAT 1380 ali manj (93,7 %), in je površina pod krivuljo levo od osenčenega območja.

Če želite ugotoviti osenčeno površino, od 1 odštejete 0,937, kar je skupna površina pod krivuljo.

Verjetnost x > 1380 = 1 - 0,937 = **0,063**

To pomeni, da je verjetno le 6,3 % rezultatov SAT v vašem vzorcu več kot 1380.



Slika 2.7 Standardna normalna porazdelitev z označenim rezultatom SAT.

2.6 Porazdelitev vzorčenja

Vzorčne porazdelitve so temelj statističnega sklepanja, saj nam omogočajo, da na podlagi vzorčnih podatkov sklepamo o populacijah. Poglobili se bomo v zapletenost vzorčnih porazdelitev, razumeli, kako odražajo variabilnost vzorčnih statistik in njihovo ključno vlogo pri preverjanju hipotez.

Porazdelitev vzorca se nanaša na porazdelitev statistike, kot je vzorčna sredina ali vzorčni delež, pridobljene iz več vzorcev enake velikosti, ki so bili odvzeti iz populacije. Omogoča vpogled v obnašanje vzorčnih statistik in njihovo spremenljivost med različnimi vzorci.

2.7 Centralni mejni teorem in porazdelitev vzorca

Centralni mejni stavek (CLT) je temeljni koncept v statistiki, ki pojasnjuje obnašanje vzorčnih porazdelitev. Trdi, da se vzorčna porazdelitev vzorčne sredine z večanjem velikosti vzorca približuje normalni porazdelitvi, ne glede na obliko populacijske porazdelitve. Ta teorem nam omogoča, da na podlagi vzorčnih podatkov zanesljivo sklepamo o populacijskih parametrih.

Osrednji mejni teorem je temelj razumevanja normalnih porazdelitev v statistiki. V raziskovalnih okoljih je za pridobitev natančne ocene populacijske sredine pogosto treba zbrati podatke iz številnih naključnih vzorcev znotraj populacije. Te posamezne vzorčne sredine skupaj tvorijo tako imenovano vzorčno porazdelitev sredine.

Centralni limitni teorem opredeljuje dve ključni načeli:

1. **Zakon velikih števil:** S povečevanjem velikosti vzorca ali števila vzorcev se povprečje vzorca približuje povprečju populacije.
2. **Normalnost porazdelitve vzorca:** Pri delu z več velikimi vzorci se kljub prvotni porazdelitvi spremenljivke vzorčna porazdelitev povprečja približa normalni porazdelitvi.

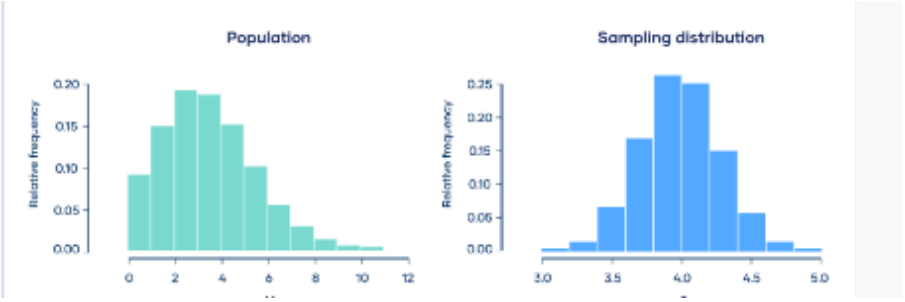
Parametrični statistični testi običajno predpostavljajo, da vzorci izhajajo iz normalno porazdeljenih populacij. Vendar teorem o centralni meji odpravlja potrebo po tej predpostavki za dovolj velike vzorce. Pri velikih vzorcih se parametrični testi lahko uporabljajo ne glede na porazdelitev populacije, če so izpolnjene druge ustrezne predpostavke. Za dovolj velik vzorec se običajno šteje vzorec velikosti 30 ali več.

Nasprotno pa je pri majhnih vzorcih zagotavljanje predpostavke o normalnosti ključnega pomena zaradi negotovosti, povezane z vzorčno porazdelitvijo povprečja. Natančni rezultati zahtevajo

potrditev, da populacija ustreza normalni porazdelitvi, preden se uporabijo parametrični testi z majhnimi vzorci.

Centralni mejni teorem nazorno pravi, da bodo sredine dovolj velikih vzorcev iz populacije pokazale normalno porazdelitev, tudi če osnovna populacijska porazdelitev odstopa od normalnosti.

Primer: Razmislite o populaciji, ki ima Poissonovo porazdelitev (prikazano na levi sliki). Ko iz te populacije vzamemo 10.000 vzorcev, od katerih je vsak sestavljen iz 50 opazovanj, se porazdelitev srednjih vrednosti vzorcev v skladu s centralnim limitnim teoremom natančno ujema z normalno porazdelitvijo (kot je prikazano na desni sliki).



Slika 2.8 Primer populacije v Poissonovi in normalni porazdelitvi.

Osrednji mejni teorem temelji na pojmu vzorčne porazdelitve, ki predstavlja verjetnostno porazdelitev statistike, izračunane iz številnih vzorcev, odvzetih iz populacije.

Konceptualizacija poskusa lahko pomaga pri razumevanju porazdelitev vzorčenja:

- Predstavljajmo si, da izberemo naključni vzorec iz populacije in izračunamo statistiko, na primer srednjo vrednost.
- Nato se izbere še en naključni vzorec enake velikosti in ponovno se izračuna povprečje.
- Ta postopek se večkrat ponovi, tako da dobimo množico srednjih vrednosti, ki ustrezajo posameznemu vzorcu.

Skupek teh vzorčnih povprečij je primer vzorčne porazdelitve. V skladu s centralnim limitnim teoremom se vzorčna porazdelitev povprečja nagiba k normalni porazdelitvi, če je velikost vzorca dovolj velika. Ne glede na porazdelitev populacije - naj bo normalna, Poissonova, binomska ali druga - vzorčna porazdelitev povprečja kaže normalnost.

Na srečo nam ni treba večkrat vzorčiti populacije, da bi ugotovili obliko vzorčne porazdelitve. Namesto tega so parametri vzorčne porazdelitve povprečja odvisni od parametrov same populacije.

- Srednja vrednost vzorčne porazdelitve je srednja vrednost populacije.

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

- Standardni odklon vzorčne porazdelitve je standardni odklon populacije, deljen s kvadratnim korenom velikosti vzorca.

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

S tem zapisom lahko opišemo porazdelitev vzorčenja povprečja:

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

Kje:

- $\bar{X}$ je vzorčna porazdelitev vzorčnih sredin
- $\sim$ pomeni "sledi porazdelitvi"
- N je normalna porazdelitev
- μ je srednja vrednost populacije
- σ je standardni odklon populacije
- n je velikost vzorca

Velikost vzorca, označena z n, predstavlja število opazovanj iz populacije za vsak vzorec, pri čemer se ohrani enakomernost vseh vzorcev. Velikost vzorca pomembno vpliva na vzorčno porazdelitev povprečja v dveh ključnih vidikih.

1. Velikost vzorca in normalnost:

- Pri večjih vzorcih so porazdelitve vzorčenja običajno podobne normalni porazdelitvi.
- Nasprotno pa lahko pri majhnih vzorcih vzorčna porazdelitev povprečja odstopa od normalnosti. To odstopanje nastane, ker je veljavnost centralnega limitnega teorema odvisna od "dovolj velike" velikosti vzorca.
- Običajno se za "dovolj velik" vzorec šteje vzorec 30 ali več oseb.
- Kadar je n<30, osrednji mejni teorem ne velja in porazdelitev vzorca odraža porazdelitev populacije. Zato je porazdelitev vzorčenja normalna le, če je populacijska porazdelitev normalna.
- Nasprotno pa pri n≥30 velja osrednji limitni teorem in porazdelitev vzorčenja se približa normalni porazdelitvi.

2. Velikost vzorca in standardni odkloni:

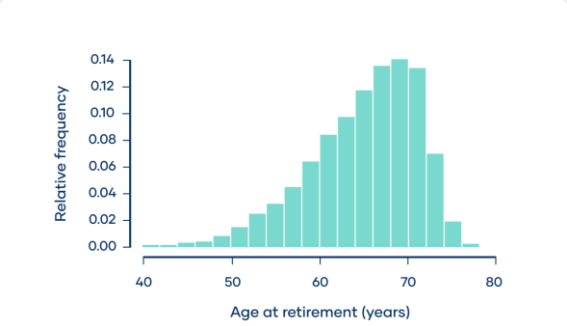
- Velikost vzorca neposredno vpliva na standardni odklon vzorčne porazdelitve, ki odraža variabilnost ali razpršenost porazdelitve.
- Pri manjših vzorcih je standardni odklon običajno večji, kar kaže na večjo variabilnost vzorčnih povprečij zaradi njihove nenatančne ocene populacijskega povprečja.
- Nasprotno pa večje velikosti vzorcev pomenijo manjše standardne odklone, kar kaže na manjšo variabilnost vzorčnih povprečij zaradi natančnejše ocene populacijskega povprečja.

Pomen centralnega limitnega teorema:

Parametrični testi, kot so t-testi, ANOVA in linearna regresija, imajo večjo statistično moč v primerjavi z večino neparametričnih testov. Ta večja statistična moč izhaja iz predpostavk o porazdelitvi populacij, ki temeljijo na osrednjem mejnem teoremu.

Neprekinjena porazdelitev

Poglejmo upokojitveno starost posameznikov v Združenih državah Amerike. Populacijo sestavljajo vsi upokojeni Američani, porazdelitev te populacije pa bi lahko predstavili na naslednji način:



Slika 2.9 Graf zvezne porazdelitve.

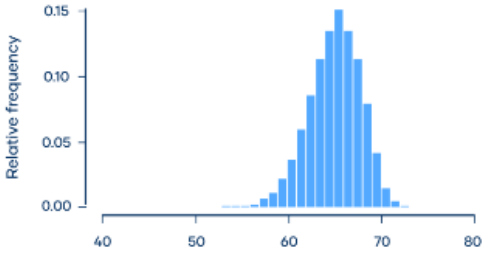
Porazdelitev upokojitvene starosti je nagnjena v levo, saj se večina upokojuje približno pet let pred povprečno upokojitveno starostjo 65 let. Vendar pa obstaja daljši rep posameznikov, ki se upokojijo veliko prej, na primer pri 50 ali celo 40 letih. Standardni odklon populacije je 6 let.

Predstavljajte si, da bi v tej populaciji izvedli manjše vzorčenje. Naključno izberemo pet upokojencev in zabeležimo njihovo upokojitveno starost. Na primer: 68, 73, 70, 62, 63

Povprečje tega vzorca služi kot približek povprečja populacije, čeprav z omejeno natančnostjo zaradi majhnega vzorca 5. Na primer: Povprečje = (68 + 73 + 70 + 62 + 63) / 5 je rezultat povprečje = 67,2 let.

Predpostavimo, da se ta postopek vzorčenja ponovi desetkrat, pri čemer vsak vzorec vključuje pet upokojencev. Izračuna se srednja vrednost vsakega vzorca, s čimer dobimo porazdelitev, znano kot vzorčna porazdelitev srednje vrednosti. Na primer: 60,8; 57,8; 62,2; 68,6; 67,4; 67,8; 68,3; 65,6; 66,5; 62,1.

Ker se ta postopek večkrat ponovi, se histogram, ki prikazuje sredine teh vzorcev, približa normalni porazdelitvi.



Slika 2.10 Normalna porazdelitev srednjih

Kljub temu da je porazdelitev vzorčenja v primerjavi s populacijo nekoliko bolj normalne oblike, je še vedno rahlo nagnjena v levo. Poleg tega je očitno, da je variabilnost vzorčne porazdelitve manjša od variabilnosti populacije.

V skladu s centralnim limitnim teoremom se vzorčna porazdelitev povprečja z večanjem velikosti vzorca približuje normalni porazdelitvi. Vendar trenutna vzorčna porazdelitev povprečja odstopa od normalnosti zaradi relativno majhnega vzorca.

2.8 Testna statistika

Testna statistika je številčna vrednost, pridobljena s testom statistične hipoteze, ki označuje stopnjo usklajenosti med opazovanimi podatki in porazdelitvijo, pričakovano pri ničelni hipotezi tega testa.

Ta statistika ima ključno vlogo pri izračunu p-vrednosti vaših ugotovitev, kar olajša odločitev o sprejetju ali zavrnitvi vaše ničelne hipoteze.

Toda kaj točno je testna statistika?

Testna statistika izraža podobnost med porazdelitvijo vaših podatkov in porazdelitvijo, predvideno pri ničelni hipotezi uporabljenega statističnega testa. Porazdelitev podatkov pojasnjuje pogostost vsakega opazovanja, za katero sta značilni njegova osrednja tendenca in variabilnost okoli nje. Ker različni statistični testi predvidevajo različne tipe porazdelitve, se izbira ustreznega testa uskladi z vašo hipotezo.

Testna statistika strne vaše opazovane podatke v eno samo številko, pri tem pa uporablja merila, kot so centralna tendenca, variabilnost, velikost vzorca in število napovednih spremenljivk v vašem statističnem modelu.

Običajno testna statistika izhaja iz opaznih vzorcev v vaših podatkih (npr. korelacij med spremenljivkami ali razlik med skupinami), deljenih z varianco podatkov (tj. standardnim odklonom).

Oglejte si naslednji primer:

Raziskujete povezavo med temperaturo in datumom cvetenja pri določeni vrsti jablane. Analizirajte obsežen nabor podatkov, ki zajema 25 let, in spremljajte temperaturo in datume cvetenja z naključnim letnim vzorčenjem 100 dreves na poskusnem polju.

- Ničelna hipoteza (H_0): Med temperaturo in datumom cvetenja ni korelacije.

- Nadomestna hipoteza (H_A ali H_1): Obstaja povezava med temperaturo in datumom cvetenja.

Da bi preverili to hipotezo, izvedete regresijski test, pri čemer je testna statistika t-vrednost. Ta t-vrednost primerja ugotovljeno korelacijo med spremenljivkama z ničelno hipotezo o ničelni korelaciji.

2.9 Vrste testne statistike

V nadaljevanju je opisan povzetek prevladujočih statističnih testov skupaj z ustreznimi hipotezami in kategorijami statističnih testov, v katerih se uporabljajo. Čeprav lahko različni statistični testi uporabljajo različne metodologije za izračun teh statistik, temeljne hipoteze in razlage testne statistike ostajajo enake.

Testna statistika	**Ničelna in alternativna hipoteza**	**Statistični testi, ki ga uporabljajo**
vrednost *t*	**Nič:** Srednji vrednosti dveh skupin sta enaki **Druga možnost:** Srednji vrednosti dveh skupin nista enaki	• T-test • Regresijski testi
vrednost *z*	**Nič:** Srednji vrednosti dveh skupin sta enaki **Druga možnost:** Srednji vrednosti dveh skupin nista enaki	• Test *Z*

Testna statistika	Ničelna in alternativna hipoteza	Statistični testi, ki ga uporabljajo
Vrednost *F*	**Nič:** Razlika med dvema ali več skupinami je večja ali enaka razliki med skupinami **Druga možnost:** Razlike med dvema ali več skupinami so manjše od razlik med skupinami	• ANOVA • ANCOVA • MANOVA
***X* vrednost²**	**-Nič:** Dva vzorca sta neodvisna **Druga možnost:** Dva vzorca nista neodvisna (tj. sta korelirana)	• Test Chi-kvadrat • Neparametrični korelacijski testi

V resničnem svetu boste testno statistiko običajno izračunali s statističnim programskim paketom, kot so R, SPSS ali Excel, ki bo tudi podal p-vrednost, povezano s testno statistiko. Kljub temu lahko formule za ročno izračunavanje teh statistik poiščete na spletu.

Pri preverjanju hipoteze o temperaturi in datumu cvetenja na primer izvedete regresijsko analizo. Rezultat regresijskega testa je: - t-vrednost, ki primerja ta koeficient s pričakovanim razponom regresijskih koeficientov pri ničelni hipotezi o odsotnosti povezave.

Rezultat t-vrednosti regresijskega testa, 2,36, predstavlja vašo testno statistiko.

2.10 Standardna napaka

Standardna napaka povprečja (SE ali SEM) služi kot kazalnik verjetne razlike med povprečjem populacije in povprečjem vzorca. Omogoča vpogled v stopnjo variabilnosti, ki bi jo lahko pričakovali v vzorčni sredini, če bi se študija ponovila z uporabo novih vzorcev, odvzetih iz iste populacije.

Standardna napaka povprečja je najpogosteje navedena oblika standardne napake, podobne mere pa obstajajo tudi za druge statistične parametre, kot so mediane ali razmerja. Standardna napaka deluje kot razširjeno merilo napake vzorčenja, ki prikazuje neskladje med populacijskim parametrom in vzorčno statistiko.

Za zmanjšanje standardne napake je priporočljivo povečati velikost vzorca. Uporaba velikega naključnega vzorca je najučinkovitejša strategija za zmanjšanje napake pri vzorčenju in povečanje zanesljivosti ugotovitev.

Standardna napaka in standardni odklon sta merili variabilnosti:

- **Standardni odklon** opisuje variabilnost **znotraj posameznega vzorca**.
- **Standardna napaka** ocenjuje variabilnost **v več vzorcih** populacije.

Standardni odklon je opisna statistika, ki izhaja neposredno iz vzorčnih podatkov, medtem ko je standardna napaka inferenčna statistika, ki je običajno ocenjena, če ni znan natančen parameter populacije.

2.11 Formula za standardno napako

Standardna napaka povprečja se določi tako, da se poleg velikosti vzorca uporabi tudi standardni odklon. S pomočjo formule je razvidno, da sta velikost vzorca in standardna napaka v obratnem sorazmerju. Poenostavljeno povedano, z večanjem velikosti vzorca se standardna napaka zmanjšuje. Do tega pojava pride zato, ker je večji vzorec praviloma bližji vzorčni statistiki populacijskega parametra.

Glede na to, ali je standardni odklon populacije znan, se uporabljajo različne formule. Te formule se uporabljajo za vzorce, ki vsebujejo več kot 20 elementov (n > 20).

Ko so parametri populacije znani

Ko je standardni odklon populacije znan, ga lahko uporabite v spodnji formuli za natančen izračun standardne napake.

Formula

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

Razlaga:

- *SE* je standardna napaka
- σ je standardni odklon populacije
- *n* je število elementov v vzorcu

Ko so parametri populacije neznani

Kadar standardni odklon populacije ni znan, lahko s spodnjo formulo samo ocenite standardno napako. Ta formula upošteva standardni odklon vzorca kot točkovno oceno standardnega odklona populacije.

Formula

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

Razlaga:

- *SE* je standardna napaka
- *s* je standardni odklon vzorca

Formula **Razlaga:**

- *n* je število elementov v vzorcu

Primer: Uporaba formule za standardno napako za oceno standardne napake za matematične rezultate SAT. Izvedite naslednja dva koraka.

Najprej poiščite kvadratni koren velikosti vzorca (*n)*.

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

Nato standardni odklon vzorca delite s številom, ki ste ga ugotovili v prvem koraku.

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

Standardna napaka matematičnih rezultatov SAT je 12,8.

Standardno napako lahko predstavite poleg povprečja ali pa jo vključite v interval zaupanja, da izrazite negotovost, ki spremlja povprečje.

Na primer: Povprečni rezultat na naključnem vzorcu udeležencev preizkusa je 550 ± 12,8 (SE).

Poročanje o standardni napaki znotraj intervala zaupanja je bolj priporočljivo, saj tako bralcem ni treba opravljati dodatnih izračunov, da bi dobili smiseln razpon.

Interval zaupanja označuje razpon vrednosti, za katere se pričakuje, da bodo najpogosteje veljale, če bi študijo ponovili z novimi naključnimi vzorci.

Pri 95-odstotni stopnji zaupanja se pričakuje, da bo 95 % vseh vzorčnih povprečij znotraj intervala zaupanja, ki obsega ± 1,96 standardne napake vzorčnega povprečja. Ta interval služi kot ocena, znotraj katere naj bi s 95- odstotnim zaupanjem bil pravi populacijski parameter.

Na primer: Zgradite 95-odstotni interval zaupanja Zgradite 95-odstotni interval zaupanja (CI), da ocenite povprečno oceno populacije za matematični izpit SAT. Pri normalno porazdeljeni lastnosti, kot je ocena SAT, približno 95 % vseh vzorčnih povprečij je znotraj približno 4 standardnih napak vzorčnega povprečja.

Formula za interval zaupanja

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

$\bar{x}$ =vzorčno povprečje = 550

SE = standardna napaka = 12,8

Spodnja meja	**Zgornja meja**
$\bar{x} - (1{,}96 \times SE)$	$\bar{x} + (1{,}96 \times SE)$
550 - (1.96 × 12.8) = **525**	550 + (1.96 × 12.8) = **575**

Pri naključnem vzorčenju vam 95-odstotni indeks zaupanja [525 575] pove, da obstaja 0,95-odstotna verjetnost, da je povprečni rezultat populacije na testu SAT iz matematike med 525 in 575.

Literatura poglavje 2

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

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

Dodatne povezave do literature in videoposnetkov na YouTubu

Poglavje 2

- 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 podatkov

Pri elektronski izmenjavi podatkov B2B (EDI) se med partnerji v oskrbovalni verigi izmenjujejo sporočila, ki vsebujejo oznake izdelkov ali storitev, identifikacije prevoznih enot in pravne dokumente. Običajno so v obliki alfanumeričnih nizov.

3.1 Informacije-podatki-znanje

V računalniško podprtih informacijskih sistemih se predstavitev informacij razlikuje glede na njihov namen in uporabo. Lahko je alfanumerična ali binarna, glede na to, ali predstavlja besedilo, sliko, zvok ali izvedljiv program ... Da bi lahko shranjevali, priklicali, obdelovali in prenašali podatke različnih vrst (npr. številke, znake, datume, valute itd.), jih je treba ustrezno kodirati. Alfanumerični formati za predstavitev podatkov izvirajo iz abecede ASCII (ask-key), ki se je razvila v smislu nacionalnih naborov znakov (npr. standardov 8859-1, Latin 1 in 8859-2, Latin 2) in nazadnje napredovala v mednarodne formate UTF-8 in UTF-16. Tako omogočajo skupno razlago podatkov s strani poslovnih partnerjev, ki pripadajo različnim etničnim skupinam in geografskim okoljem. Medtem ko so alfanumerični nizi odvisni od njihovega kodiranja, se numerični podatki razlikujejo predvsem po svoji velikosti in/ali natančnosti.

Postopek pretvorbe podatkov iz prvotne analogne v digitalno obliko se popularno imenuje digitalizacija. Ustrezno zasnovani uporabni in aplikativni programi, ki sprejemajo podatke iz različnih virov (npr. optični čitalniki, električni senzorji, EDI itd.), omogočajo organizacijam, da avtomatizirajo pridobivanje, shranjevanje, obdelavo in prenos podatkov znotraj svojih računalniško podprtih informacijskih sistemov in med njimi.

Ko so zbrani, se lahko podatki različnih vrst združijo in organizirajo v podatkovne tabele, podatkovne baze, podatkovna skladišča (kronološki vrstni red) ali baze znanja (konceptualni vrstni red). Višje ravni organizacije podatkov omogočajo avtomatizirano klasifikacijo, sklepanje in predstavitev tako zbranega znanja za analitične namene.

3.2 Logistični podatki

V logistiki se EDI uporablja za prenos podatkov o transakcijah med poslovnimi partnerji. Ker ti lahko uporabljajo različne jezike in aplikacije, je treba podatke pretvoriti v skupno obliko (npr. XML, JSON), da jih lahko interpretirajo informacijski sistemi različnih partnerjev (W3schools, 2023). Za hitro identifikacijo in manipulacijo so bile razvite črtne kode in oznake RFID.

Da bi omogočili mednarodno sodelovanje, je bilo treba za logistične namene opredeliti svetovno sprejete podatkovne formate. Logistični podatkovni formati ustrezajo oznakam storitev in izdelkov, identifikacijam prevoznih enot in kodam transakcij, ki so običajno v obliki alfanumeričnih nizov s časovnim žigom. Zaradi enostavnosti manipulacije in hitrosti obdelave so bile te kode standardizirane in kodirane kot optično berljive črtne kode ali elektromagnetno berljive kode za radiofrekvenčno identifikacijo (RFID).

Črtna koda (EAN/UCC) je več-sektorska in mednarodna oblika številčenja predmetov (POS EAN-8 in EAN-13, spremenljiva EAN-128, podatkovna vrstica, embalaža ITF-14, QR, podatkovna matrika itd.). Uporabljajo se za identifikacijo izdelkov, serij izdelkov ali pošiljk (kode 1D) in storitev (kode 2D). Med črtnimi kodami 2D je najbolj uveljavljena koda QR, ki jo je mogoče prebrati tudi s pametnimi telefoni, kar povečuje njihovo uporabnost na različnih področjih uporabe.

Kode RFID se uporabljajo predvsem na enak način kot črtne kode. Z njimi je mogoče enolično identificirati predmete ali storitve. Običajno je na nalepkah RFID poleg izbirne črtne kode še več informacij na čipu, velikem kot glavica pincete. Poleg identifikacije oznake RFID omogočajo beleženje informacij o sledenju, kar se pogosto zahteva v logističnih aplikacijah. V nasprotju s črtnimi kodami omogoča RFID njihovo skeniranje brez neposredne vidljivosti in tudi več etiket hkrati.

S standardom GS1 EPC Gen2 (ISO/IEC 18000-6:2013) je bil vzpostavljen tehnološki standard, ki določa komunikacijo med oznakami RFID in bralniki. Podobno kot pri črtnih kodah standardi EPCglobal povezujejo tehnologijo RFID z označevanjem izdelkov, logističnih transportnih enot, lokacij, zalog, vračljivih predmetov, dokumentov itd. z oznako EPC za neposredno, avtomatizirano identifikacijo in sledenje logističnih enot v oskrbovalnih verigah.

Standardi EPCglobal so tudi osnova za GDSN (Global Data Synchronization Network). Omogoča avtomatizirano pridobivanje in izmenjavo podrobnih podatkov o izdelkih in njihovi embalaži, s čimer podjetjem omogoča centralno upravljanje teh podatkov, ki jih lahko izmenično uporabljajo sama in njihovi partnerji.

V preglednici 3.1 so povzete različne tehnologije identifikacije in njihove uporabe. Razkriva različne eno- in dvodimenzionalne črtne kode ter različne razrede kod RFID z njihovimi zmogljivostmi.

Preglednica 3.1 Tehnologije označevanja.

Tehnologija	Aplikacija
Bar 1D	Maloprodajni izdelki in sestavni deli izdelkov
Bar 2D	storitve (npr. UPS, letalske vozovnice), blago na debelo, ki zahteva sledenje
RFID razreda 1 (pasivni, R-oznake)	Predmeti, ki zahtevajo množično identifikacijo, nadzor dostopa
RFID razreda 2 (pasivne, RW-oznake)	Predmeti, ki jih je treba spremljati
RFID razreda 3 (pol-aktivne, RW-oznake)	Nadzor dostopa z dodanimi informacijami o sledenju
RFID razred 4 (aktivni, RW-oznake)	Sledenje in sledenje v zaprtem prostoru
RFID razred 5 (aktivne oznake/bralci)	Sledenje in sledenje prostemu prostoru, storitve bližine z omogočenimi napravami, storitve, ki temeljijo na lokaciji.

Prihodnji trendi na področju označevanja, sledenja in izsleditve sledijo dvema glavnima smerema: miniaturizaciji in raznolikosti. Podatkovne črtne (1D) kode omogočajo tudi označevanje miniaturnih predmetov (npr. medicinskih kapsul). Nove matrične kode (2D) ne bodo omogočale le odpravljanja napak med skeniranjem, temveč tudi šifriranje podatkov.

RFID se še naprej širi na druga področja uporabe, kot so identifikacija s strani ponudnikov storitev (npr. železniška kartica, registracija na delovnem mestu itd.), brezstična plačila (npr. brezžični prenosi denarja, plačila na prodajnih avtomatih) in pametne rešitve (npr. upravljanje pametnega doma, daljinsko upravljanje pametnih naprav itd.) ter e-valute.

3.3 Organizacija podatkov

Poleg določene oblike so lahko podatki organizirani na različne načine, da se olajša njihovo upravljanje, obdelava in predstavitev. Čeprav so podatki na vhodu večinoma nestrukturirani, se z njihovim shranjevanjem, prenosom in obdelavo povečuje njihova organiziranost. V nadaljevanju so predstavljene običajne oblike organizacije podatkov z naraščajočo kompleksnostjo od pol-strukturiranih (npr. CSV) do strukturiranih (npr. preglednice, podatkovne zbirke itd.) oblik.

Preglednice

Prva oblika organizacije podatkov je z dvodimenzionalnimi polji, imenovanimi tudi tabele ali preglednice. Običajno prva vrstica preglednice označuje pomen vrednosti, shranjenih v osnovnih stolpcih, ki jim sledijo vrstice s podatki.

Polje ali celica tabele je najmanjša enota podatkov. Ima določeno vrsto (niz, številka, datum, valuta itd.). Njegovo vsebino je mogoče nasloviti z oznakami vrstic in stolpcev (npr. A1, ki predstavlja prvo vrstico stolpca A).

Vsaka vrstica tabele je skupina povezanih polj, ki predstavljajo zapis (npr.: transakcija, zapis študenta, podatki o izdelku itd.). Ker imajo vse vrstice tabele enako strukturo, lahko tip zapisa opredelimo kot seznam atributov (npr. podatki o študentu (ime, priimek, datum rojstva, kraj rojstva, ID ...)) ustreznih podatkovnih tipov.

Podatkovne zbirke

Datoteka ali tabela podatkovne zbirke je zbirka zapisov istega tipa. Podatkovna baza (DB) je sestavljena iz več medsebojno povezanih tabel. Tako je zbirka podatkov opredeljena v standardu ANSI:

- Podatki v DB so medsebojno povezani in razvrščeni.
- Podatke iz DB lahko hkrati uporablja več uporabnikov.
- Podatki v DB se ne ponavljajo
- DB je shranjena v računalniku.

Iz zgornje opredelitve lahko potegnemo nekaj zaključkov o arhitekturi odjemalec-strežnik, kjer ima strežnik DB, do katerega dostopajo odjemalci. Seveda je treba za dostop do DB vzpostaviti komunikacijsko omrežje med strežnikom in njegovimi odjemalci. Strežnik DB se običajno imenuje "zaledni del", medtem ko odjemalci predstavljajo njegov "sprednji del". Sistem za upravljanje

podatkovnih zbirk (DBMS) na strežniku omogoča svojim odjemalcem dostop do podatkov, shranjenih v DB, prek vmesnika za aplikacijske programe (API) in funkcij DBM. Funkcije DBM so mehanizmi, ki omogočajo vnos, pridobivanje, obdelavo in predstavitev podatkov v DB. Za priklic teh funkcij so bili opredeljeni standardni poizvedovalni jeziki (SQL).

Relativni model podatkovne baze

Obstajajo različne oblike organizacije DB, med katerimi je najpogostejši relacijski model (RDB). Osnovna ideja tega modela je dejstvo, da uporabnik ne more vnaprej poznati vseh možnosti uporabe podatkov, shranjenih v DB. Ker običajno ni fiksnih poti iskanja po datotekah zbirke podatkov, so bili za iskanje in manipulacijo s podatki razviti različni poizvedovalni jeziki. Model RDB temelji na konceptu entitet in relacij:

- Entiteta je oseba, stvar ali koncept, ki ga je mogoče enolično identificirati in ima atribute.
- Relacija predstavlja način povezovanja dveh ali več entitet.

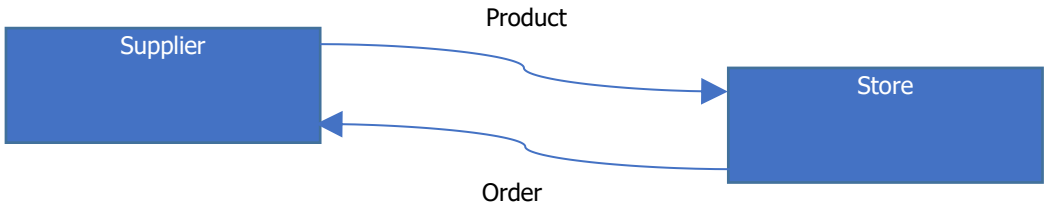
Tabele RDB, ki predstavljajo entitete ali relacije, so med seboj povezane s ključi. Nabor atributov, ki enoznačno identificirajo entiteto, se imenuje njen primarni ključ. Kadar se primarni ključ pojavi kot polje v drugi preglednici, da bi izpolnil razmerje s svojo prvotno preglednico, se imenuje sekundarni ali tuji ključ. Medtem ko lahko tabela vsebuje samo en primarni ključ za enolično identifikacijo svojih zapisov, lahko vsebuje več sekundarnih ključev.

Na splošno obstajata dva pristopa k izgradnji RDB: analitični in sintetični.

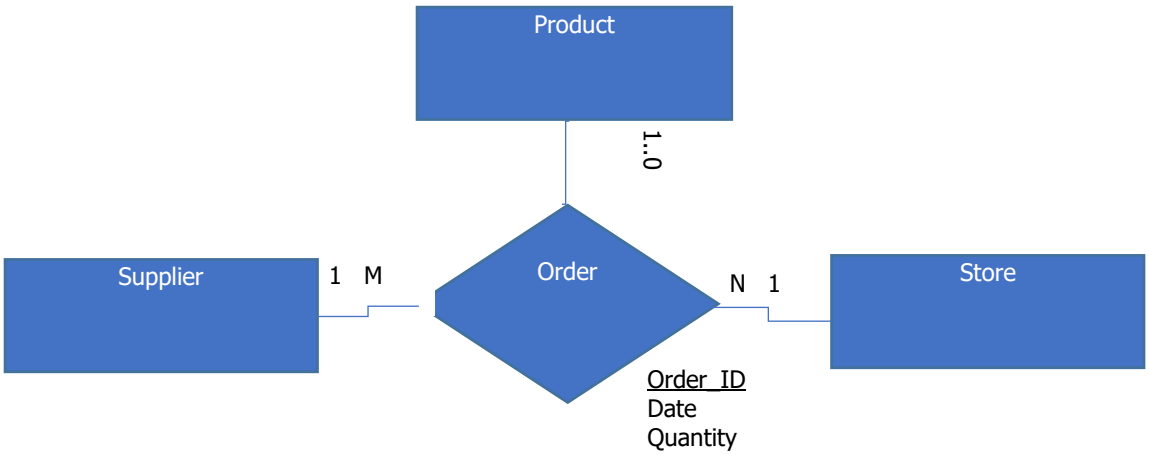
Analitični pristop k izdelavi RDB obsega naslednje štiri korake:

1. Analiza realnega sveta - globalni model
2. Določite entitete in odnose - konceptualni model (npr. E-R diagram)
3. Določitev logičnega modela - relacijska shema
4. Izdelava podatkovne zbirke (DBMS) - fizični model

Za ponazoritev pristopa si oglejmo primer verige trgovin s hitro prehrano in njihovih dobaviteljev (slika 3.1). Vsaka trgovina ima več dobaviteljev. Vsak dobavitelj lahko oskrbuje različne trgovine. Cikel dopolnjevanja zalog se začne z naročilom iz trgovine. V zameno dobavitelj dostavi blago v trgovino. Naročila so transakcije, v katerih so združeni podatki o trgovini, dobavitelju in dobavljenem blagu (slika 3.2).



Slika 3.1 Globalni model.



Slika 3.2 Konceptualni model.

SUPPLIER (Supplier_ID#, Supplier_name, Supplier_contact)

STORE (Store_ID#, Store_name, Store_address)

ORDER (Supplier_ID#, Store_ID#, Order_ID#, Date, Quantity, EPC)

PRODUCT (EPC#, Product_name, Product_price)

Slika 3.3 Logični model.

Logični model predstavlja tabele RDB z vrstami njihovih zapisov. V logičnem modelu (slika 3.3) nekateri atributi vsebujejo simbol hashtag (#). To pomeni, da je polje, ki ga predstavljajo, primarni (sestavljeni) ključ ali mu pripada. Nekatera polja so podčrtana. Imenujemo jih sekundarni ali tuji ključi, saj se sklicujejo na primarne ključe povezanih tabel.

Sintetični pristop k RDB obsega naslednje tri korake:

1. Analiza podatkov - seznam vseh ustreznih atributov
2. Določitev logičnega modela z normalizacijo - relacijska shema

3. Fizični model (DBMS)

Normalizacija ali kanonična sinteza (Kent, 1983) zagotavlja, da se s povratnim inženiringom iz ustreznih atributov oblikuje DB, ki izpolnjuje pogoje RDB (prim. sliko 4). Medtem ko začetna normalna oblika (0NF) atributov predstavlja tabelo neurejenih atributov, poznejše normalne oblike, kot jih opredeljuje (Codd, 1970), predstavljajo višje ravni organizacije podatkov. Lahko trdimo, da je z dosegom tretje normalne oblike dosežena shema, ki izpolnjuje zahteve za logični model RDB.

Tabela je v 1NF, če predstavlja relacijo. S tem je zagotovljeno, da se vse ponavljajoče se skupine podatkov hranijo ločeno in se torej ne ponavljajo.

Tabela v 2NF je v 1NF. Poleg tega noben atribut ključa ne sme biti delno funkcionalno odvisen od primarnega ključa. Pri tem so vsi ključi, ki enolično identificirajo določene atribute, ločeni. S tem se predvsem zagotovi, da so v eni tabeli samo atributi, ki so odvisni od vseh (delov) primarnega ključa.

Tabela v 3NF je v 2NF. Poleg tega noben atribut, ki ni ključ, ni prehodno odvisen od primarnega ključa. To pomeni, da so vsi neključni atributi, ki lahko predstavljajo ključ za nekatere druge neključne atribute, shranjeni v ločeni tabeli, pri čemer se v prvotni tabeli kot tuji ključ ohranja le njihov ključ.

Z normalizacijo od 1NF do 3NF dobimo logični model, ki ustreza predpisom relacijske podatkovne baze (RDB). Višje oblike normalizacije so namenjene predvsem optimizaciji modela RDB.

Tabela v Boyce-Codd NF (BCNF) je v 3NF; poleg tega je vsaka determinanta ključ. S tem se iz prvotne tabele odstranijo vse relacije, ki niso zajete z obstoječim vrstnim redom ključev, s čimer se za vsak kandidatni ključ oblikujejo dodatne tabele. Tabela v 4NF je v 3NF in BCNF. Poleg tega je vsak večvrednostni atribut, ki je delno odvisen od ključa, v svoji tabeli. Namen 4NF je odstraniti vse možne preostale večvrednostne atribute iz prvotne tabele. Tabela 5NF je v 4NF; poleg tega je vsaka operacija JOIN predvidena s ključi. Tabela v 6NF je v 5NF; poleg tega so upoštevane vse netrivialne odvisnosti JOIN.

Opazimo lahko, da se pri višjih oblikah organizacije število tabel povečuje z vsakim korakom. Zato je smiselno opazovati razdrobljenost podatkov, da se prepreči ustvarjanje nepotrebnih tabel, do katerih se dostopa le redko.

Preglednica 3.2 Primer normalizacije RBD.

0NF	1NF
NAROČILO • Supplier_ID*	NAROČILO • Supplier_ID#

• Ime dobavitelja • Dobavitelj_kontakt • Store_ID* • Ime_trgovine • Store_address • Order_ID* • Datum • EPC • Količina • Ime_izdelka • Cena_izdelka	• Ime dobavitelja • Dobavitelj_kontakt • Store_ID# • Ime_trgovine • Store_address • Order_ID# • Datum • EPC • Količina • Ime_izdelka • Cena_izdelka
* Kandidatski ključi ponavljajoče se skupine atributov, ki enolično identificirajo naročilo	
2NF DOBAVITELJ • Supplier_ID# • Ime dobavitelja • Dobavitelj_kontakt TRGOVINA • Store_ID# • Ime_trgovine • Store_address	NAROČILO • Order_ID# • Supplier_ID# • Store_ID# • EPC • Ime_izdelka • Cena_izdelka • Datum • Količina
3NF NAROČILO • Supplier_ID# • Store_ID# • Order_ID# • Datum • Količina • ID_izdelka	IZDELEK • EPC# • Ime_izdelka • Cena_izdelka

Jeziki poizvedb

Večinoma so dveh vrst: Strukturirani poizvedovalni jezik (SQL) in poizvedovanje po zgledu (QBE). Medtem ko SQL velja za programski jezik in API DBMS, se QBE večinoma uporablja neposredno z DBMS za upravljanje DB in skladiščenje podatkov.

Standard SQL (ISO/IEC 9075, 1986-2016) je programski jezik četrte generacije za manipulacijo s podatkovnimi bazami. Omogoča iskanje, dodajanje, spreminjanje in brisanje podatkovnih zapisov.

Kljub njegovi standardizaciji obstajajo manjše razlike v njegovi implementaciji pri različnih sistemih za upravljanje podatkovnih zbirk (DBMS).

V nadaljevanju je jezik na kratko predstavljen z najpogostejšimi možnostmi. Po dogovoru so ključne besede SQL zapisane z velikimi črkami, vsak stavek pa se konča s podpičjem. Stavki so predstavljeni s povezavami do povezanih referenčnih gradiv, ki ponujajo dodatne informacije.

Vsaka manipulacija s podatkovno zbirko se začne z njenim ustvarjanjem. Stavek

CREATE DATABASE *Ime_podatkovne_baze*;

ustvari novo prazno zbirko podatkov z navedenim imenom.

Kot je opisano zgoraj, so podatki v zbirkah podatkov organizirani v tabelah podatkovnih zapisov določenega tipa, kjer imajo vse vrstice skupno strukturo. Za ustvarjanje tabele se uporablja naslednji stavek:

CREATE TABLE *ime_tabele* (
column1 type1
, column2 type2,
column3 type3,
....);

Vsak poimenovani stolpec predstavlja atribut z določenega podatkovnega tipa. Na primer, v:

CREATE TABLE *TRGOVINA* (*Store_ID* int NOT NULL PRIMARNI KLJUČ,...);

CREATE TABLE *NAROČILO* (*Order_ID* int NOT NULL PRIMARNI KLJUČ, ..., *Product_Id* int FOREIGN KEY REFERENCES *Product* (*EPC*));

ustvarjeni sta dve tabeli. Prva vsebuje podatke o strankah, druga pa podatke o njihovih naročilih in se sklicuje na prvo tabelo prek številke stranke kot tujega ključa.

Medtem ko so podatki v tabeli že razvrščeni po primarnem ključu, jih je mogoče dodatno razvrstiti po drugih atributih, če so indeksirani. Indeksiramo ga lahko tako, da ustvarimo indeks na danem atributu (atributih) z naslednjim stavkom:

CREATE INDEX *index_name* ON *table_name* (*column_name*);

Vsaka manipulacija s podatki v indeksirani tabeli traja nekoliko dlje, saj je treba za doslednost preveriti ne le podatkov iz ključev in jih ustrezno urediti, temveč tudi druge atribute iz določenega indeksa.

Najpogostejša operacija v zbirki podatkov je poizvedba podatkov, ki jo omogoča stavek SELECT:

SELECT *stolpec1, stolpec2*, ...
FROM *ime tabele*;

Ta podatkovna poizvedba vrne podatke iz tabele v stolpcu1, stolpcu2 itd. Stavki poizvedbe so običajno oblikovani z zagotavljanjem dodatnih možnosti, filtriranjem podatkov, izpolnjevanjem določenih pogojev:

WHERE določa pogoj, ki določa merila za izbor zapisov.

GROUP BY združi zapise, ki imajo skupno lastnost, ki omogoča združevalne funkcije.

Funkcija HAVING določa zbirne funkcije za skupine, opredeljene z izjavami GROUP BY.

ORDER BY določa atribute, na podlagi katerih so urejeni zapisi za vračilo.

Na primer:

SELECT "Store". "Ime_trgovine", "Izdelek". "Ime_izdelka", "Naročilo". "Količina" FROM "Naročilo", "Izdelek", "Dobavitelj", "Trgovina" WHERE "Naročilo". "ID_izdelka" = "Izdelek". "EPC" IN "Naročilo". "ID_dobavitelja" = "Dobavitelj". "ID_dobavitelja" IN "Naročilo". "ID_trgovine" = "Trgovina". "ID_trgovine" ORDER BY "Store". "Ime_trgovine" ASC

vrne seznam trgovin z naročenimi izdelki in količinami, urejen po imenu trgovine.

Najpomembnejša operacija v postopku izbire je operacija JOIN. Pogosto se namesto pogoja WHERE uporablja operacija JOIN ON, ki ji sledi pogoj. Primerja vrednosti stolpcev in na podlagi primerjave določi, ali jih je treba vključiti v rezultat ali ne. Pri operaciji LEFT JOIN se vrne zapis, če so merila izpolnjena v levi tabeli, in obratno pri operaciji RIGHT JOIN. Kot je opisano zgoraj, mora biti pogoj izpolnjen v obeh tabelah, da je skladen z operacijo INNER JOIN ali FULL JOIN. Ker se slednja najpogosteje uporablja, lahko kot sopomenko uporabimo JOIN. Glede na pogoje normalnih oblik 5th in 6th gre za isto operacijo JOIN, ki mora biti izpolnjena, da je skladna s pogoji ustrezne NF.

Za vnos novih podatkov v tabelo se uporablja operacija INSERT INTO:

INSERT INTO *ime tabele (stolpec1*, [*stolpec2*, ...])
VALUES (*value1*, [*value2*, ...]);

Da bi bila operacija uspešna, morajo vrednosti v operaciji izpolnjevati vse pogoje atributov, ki jih označujejo imena stolpcev. Če so navedene vse vrednosti, imen stolpcev ni treba navesti. Če so v tabeli predvidene nekatere vrednosti DEFAULT, jih ni treba navesti, razen če so drugačne.

Ko so podatki vneseni, jih lahko spremenite z ukazom UPDATE:

UPDATE ime_tabele
SET *column1=vrednost1, column2=vrednost2*,...
WHERE *stolpec=neka_vrednost*;

V izjavi so podane nove vrednosti za polja v navedenih stolpcih. Kriteriji za izbiro vrstic so označeni s specifikatorjem WHERE, ki določa vse vrednosti stolpcev, na katere se nanaša izjava UPDATE. Da bi preprečili neželene spremembe, je treba biti pri oblikovanju izbirnih meril še posebej previden.

Z operacijo DELETE lahko iz tabele izbrišete zapis ali več zapisov:

DELETE FROM *ime* tabele
WHERE *some_column=some_value*;

Tako kot pri stavku UPDATE se za določitev vseh vrstic, ki jih je treba izbrisati, uporabi določilo WHERE.

Upravljanje podatkovne zbirke se tu seveda ne konča. Vsak element zbirke podatkov lahko tudi odstranite, spremenite in/ali nadomestite z novim. Če je treba odstraniti indeks, tabelo ali zbirko podatkov, lahko uporabite naslednje stavke:

DROP INDEX *index_name* ON *table_name*;

DROP TABLE *ime_tabele*;

DROP DATABASE *podatkovne baze ime_podatkovne baze*;

Če želite iz tabele odstraniti le podatke, lahko uporabite stavek TRUNCATE:

TRUNCATE TABELE *ime_tabele*;

Če želimo dodati ali odstraniti atribut (stolpec) v/iz tabele, lahko to storimo z ukazom ALTER:

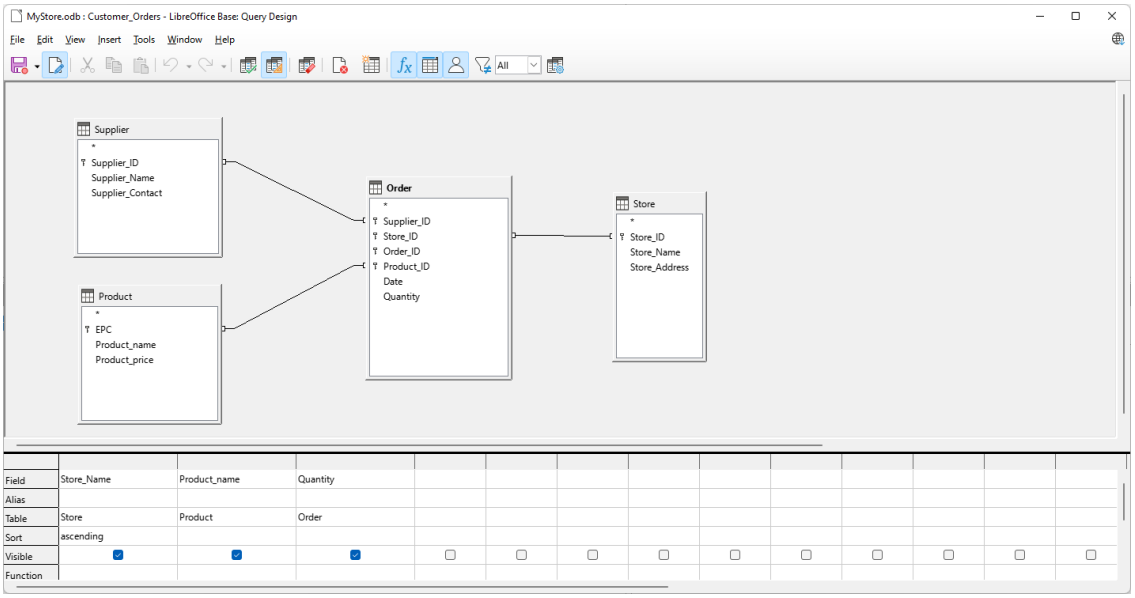
ALTER TABLE *table_name* ADD *column_name datatype*;

ALTER TABLE *table_name* DROP COLUMN *column_name*;

S tem zaključujemo kratek pregled jezika SQL in najpogostejših scenarijev njegove uporabe. Jezik SQL se pogosto uporablja v arhitekturah odjemalec-strežnik, kjer je DBMS nameščen v strežniku. Za dostop do nje se izdajajo stavki SQL, bodisi s programom odjemalske aplikacije bodisi s spletnim vmesnikom strežnika DBMS.

Po drugi strani pa se QBE pogosto uporablja tudi z relacijskimi DBMS z grafičnim uporabniškim vmesnikom (GUI), kot sta MS Access ali LibreOffice Base. S QBE se zbirka podatkov in njene tabele ustvarjajo veliko bolj interaktivno, njihova struktura pa se lažje vzdržuje. Kot pove že njeno ime,

ponuja tudi preprostejšo obliko vnosa in odkrivanja podatkov. Za izvedbo iskanja je treba sestaviti vse tabele, ki se uporabljajo pri iskanju, nato pa določiti pogoje kot vzorce v poljih stolpcev za filtriranje ustreznih podatkov (slika 3.4). Formulacija poizvedbe je dopolnjena z obstoječimi razmerji med tabelami. Kot običajno je rezultat take poizvedbe druga tabela z dobljenimi podatki, ki jih je mogoče pozneje dodatno obdelati. Na ta način se lahko oblikujejo tudi kaskadne ali večfazne poizvedbe.



Slika 3.4 Poizvedba z zgledom, enakovredna zgornji poizvedbi SQL.

Filtri in maske podatkov

Da bi preprečili napačne ali nepopolne podatke, ki bi lahko vplivali na njihovo obdelavo in razlago, je treba uporabiti dodatne previdnostne ukrepe:

1. Filtriranje praznih vrstic in stolpcev
2. Uporaba strogega tipiziranja podatkov za preprečevanje računskih napak.
3. Opredelitev vhodnih mask za preprečevanje vnosa napačnih podatkov
4. Prilagajanje podatkov za preprečevanje napačnih rezultatov

Prazne vrstice in stolpci so pogost vir napak, ki je predvsem posledica slabih vmesnikov v aplikacijah za zbiranje podatkov. Preklicane ali nepopolne transakcije običajno povzročijo prazne vrstice ali manjkajoče podatke v dnevnikih transakcij. Le delno jih je mogoče obravnavati z aplikacijami preglednic, kjer je mogoče prazne vrstice in manjkajoče podatke odkriti s preverjanjem skupnega števila v primerjavi s številom neničelnih podatkov v vrsticah/stolpcih. Lahko jih filtriramo tako, da

odstranimo prazne vrstice/stolpce, vendar to ni vedno želeno dejanje, saj lahko izgubimo tudi nekaj dragocenih podatkov. To lahko najbolje preprečimo tako, da uporabimo sistem DBMS, ki bi na eni strani dovoljeval vnos le popolnih podatkov o transakcijah, na drugi strani pa bi preprečeval tudi vnos praznih vrstic/stolpcev, saj jih v podatkovnih zbirkah ni.

Slabo tipiziranje podatkov je še en pogost vir napak. Če namesto številčnih podatkov vnesete alfanumerične podatke, na primer datum ali znesek v valuti, lahko pride do napak pri obdelavi teh podatkov. Tako v preglednicah kot v podatkovnih zbirkah lahko posameznim podatkovnim celicam, ki predstavljajo vrednosti atributov v podatkovnih zapisih, pripišemo podatkovne tipe, kar bi nas ob vnosu podatkov v napačni obliki alarmiralo. Tako lahko s tem ukrepom preprečimo, da bi napačni podatki ovirali njihovo obdelavo.

Pri oblikovanju podatkovnih zbirk lahko za podatkovna polja, ki predstavljajo atribute entitete ali relacije, uporabimo omejitve. Poleg tega, da se jim dodeli ustrezen tip, se lahko določijo maske za vnos, ki omogočajo vnos podatkov, kot so datumi, valute, kode EAN itd., samo v določeni obliki. S tem se običajno odpravijo številne napake, do katerih bi sicer lahko prišlo pri obdelavi podatkov.

Drug pogost vir napak so nepristranski podatki, ki predstavljajo podatke, ki so za red velikosti večji ali manjši od pričakovanih. Tudi ti lahko motijo našo obdelavo in povzročijo napačne rezultate. Težje jih je odkriti in jih je mogoče izločiti le s pregledom podatkov. Pri uporabi preglednic je dobra običajna praksa, da določimo najmanjše, največje in srednje vrednosti podatkov v ustreznih stolpcih, da bi odkrili morebitna odstopanja. Če odkrijemo le malo takih primerov, jih lahko poudarimo in obravnavamo ročno. Če jih je veliko, pa jih lahko filtriramo in spremenimo s poizvedbo v tabeli podatkovne zbirke. V vsakem primeru jih je treba skrbno oceniti, da ne bi še poslabšali stanja, saj bi bilo v tem primeru boljše te podatke odstraniti.

Podatkovna skladišča in baze znanja

Podatki v podatkovnih skladiščih so zbrani iz RDB in katalogizirani v kronološkem vrstnem redu. Običajno se na podatkih opravijo nekatere poslovne analize, rezultate pa analitični oddelek shrani tudi za poznejše sklicevanje. Ko so ti podatki shranjeni v skladišču, se običajno ne spreminjajo, da se ohrani njihova doslednost.

Poleg kronološkega vrstnega reda se pri gradnji baz znanja upoštevajo tudi kontekstualni redi. Pri tem so entitete predstavljene kot izpeljanke entitete najvišje ravni ali njene podrejene entitete. Njihova razmerja se vzpostavljajo bolj svobodno, saj so namenjena posodabljanju in nadgrajevanju ob njihovi uporabi. Vzpostavljeni so v obliki pravil, ki temeljijo na lastnostih entitet. Zato je oblika, v kateri so shranjeni, nekoliko drugačna. Pogosto so shranjeni v obliki ontologij, ki vsebujejo

poglobljeno znanje o zbranih podatkih. Podobno kot pri shranjevanju rezultatov poizvedb v podatkovnih skladiščih so tudi poizvedbe v bazah znanja shranjene za poznejšo uporabo za prikaz trenutnih rezultatov, saj se entitete, razmerja in primerki podatkov spreminjajo.

V nasprotju s podatkovnimi bazami in skladišči, ki so vezani na posamezno aplikacijo, so lahko baze znanja neodvisne od aplikacije in se pogosto uporabljajo v različnih aplikacijah. Primer je predstavljen v (Gumzej et. al., 2023).

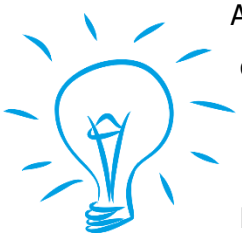
3.4 Zaključek

V tem poglavju so bili obravnavani različni vidiki upravljanja podatkov v logistiki. Poleg predstavitev podatkov in standardov za shranjevanje podatkov so bili predstavljeni tudi mehanizmi za organizacijo in pridobivanje podatkov. Na koncu so bile obravnavane nekatere pogoste napake pri avtomatizirani obdelavi podatkov, da bi bil zaveden bralec pozoren. Poleg naštetih primerov lahko v pripadajočem učnem gradivu odkrijete še več.

Literatura poglavje 3

- 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]

4. Simulacijsko modeliranje in analiza



S simulacijskim modeliranjem in analizo (SMA) si prizadevamo izpolniti zahteve Conant-Ashbyjevega teorema (Conant in Ashby, 1970), ki opredeljuje model sistema kot dobrega regulatorja, ki ima toliko ročic, delov in stanj kot njegov izvirni fizični ustreznik; tako je mogoče zgraditi njegov digitalni model in vzpostaviti digitalni laboratorij, ki bo omogočal njegovo raziskovanje, prilagajanje in optimizacijo. Nastali simulacijski modeli so abstraktni, dinamični in v večini primerov stohastični, saj so njihove sistemske spremenljivke modelirane z verjetnostnimi porazdelitvami.

4.1 Simulacija v logistiki

V logistiki lahko SMA zagotovi dragocene podatke za optimizacijo oskrbovalne verige (SC) in prometnega omrežja (TN). Simulacijsko modeliranje se lahko uporablja za grafično vizualizacijo časovnih tokov skozi kompleksne procese in vire v SC in TN, kar omogoča napovedovanje in količinsko opredelitev možnih rezultatov različnih scenarijev. To subjektom SC in TN pomaga pridobiti dragocene vpoglede in razumeti učinke njihovih morebitnih odločitev na delovanje SC in TN, vključno s časi izvedbe SC, potovalnimi časi TN in stroški. Zato lahko SMA pri modeliranju SC in TN prispeva k analizi SC in TN ter izboljšanju njihovih zasnov za doseganje večje učinkovitosti in trajnosti.

SC ima več vidikov, ki predstavljajo različne perspektive upravljanja SC. Pogled vodje proizvodnje na SC se razlikuje od pogleda vodje trženja, ki se spet razlikuje od pogleda vodje oskrbe itd. Zato so uporabljeni modeli različni celo za isto podjetje, kaj šele za celoten SC.

Pri reševanju problemov SMA v logistiki morajo menedžerji sprejemati odločitve na strateški, taktični in operativni ravni glede na njihov učinek na KS ali TN kot celoto. Zaradi njihove medsebojne odvisnosti menedžerji pogosto ne morejo reševati problemov na nobeni posamezni ravni. Hkrati je tudi težko opazovati vse tri ravni z vidika posamezne enote. Z vidika SMA je mogoče opazovati KS ali TN na dveh ravneh:

1. Makro raven

- samoorganizacija,
- so-razvoj entitet,

- odvisnost od povezav/prevoznih poti.

2. Mikro raven

- številne in heterogene entitete,
- lokalne interakcije med entitetami,
- strukturirane subjekte,
- prilagodljive entitete.

Čeprav se izvajajo v realnem času, je časovni vidik operacij SC nekoliko dvoumen. Odvisno od ravni in vidika se lahko trajanje operacij meri v dnevih, tednih ali celo mesecih, kadar gre za medorganizacijske dejavnosti, medtem ko se po drugi strani notranje organizacijske operacije merijo v urah ali celo sekundah. Glede na naravo modeliranega problema trajanje najkrajše operacije ali največja pogostost vhodnih/izhodnih zahtevkov ne določa le predstavitve časa v modelu SMA, temveč tudi njegovo granulacijo. Čim krajše je najkrajše trajanje najkrajše operacije ali čim večja je najvišja frekvenca zahtevkov, tem večja je granularnost časa ali z drugimi besedami natančnost vodenja časa v modelu. To je pomembno za izdelovalca modela, saj reakcijski čas modela ne more biti krajši od vnaprej določene granulacije časa. Zato je treba vnaprej oceniti trajanje vseh operacij in čas med prihodi vhodnih/izhodnih signalov, da lahko pravilno določimo časovne enote modela sistema.

V simulacijskem modelu lahko čas napreduje po kritičnih dogodkih od transakcije do transakcije ali neprekinjeno. V slednjem primeru je potek časa v modelu neodvisen od pogostosti operacij. Pri časovnem toku, ki ga sprožijo kritični dogodki, se operacije sprožijo glede na čas njihovega pojavljanja, tj. kritičnih dogodkov. Prednost SMA je v tem, da lahko med simulacijo pospešimo potek časa v modelu, tako da procesi potekajo hitreje kot v realnem času. Tako je mogoče zgodaj napovedati naslednje dogodke.

Časi med vhodnimi simulacijskimi enotami in časi njihove obdelave/tranzita so lahko rezultat opazovanj in meritev. Če se ne spreminjajo, so deterministični. Vendar so običajno po svoji naravi stohastični. Zato je treba uvesti konstrukcije, ki modelirajo njihove verjetnostne porazdelitvene funkcije (npr. trikotno, enakomerno, eksponentno itd.).

4.2 Simulacija diskretnih dogodkov

Analiza DES ponuja vodji proizvodnje najbolj podroben vpogled v logistični (proizvodni) proces z doslednim in skladnim modelom. Tako je DES zelo cenjeno orodje za določanje obnašanja v realnem času in uporabe virov v procesni industriji, vključno z logistiko.

Konstrukti:

- Enote pretoka predstavljajo simulacijske enote (npr. naročila, materiali itd.), ki vstopajo v sistem na vhodu(-ih) in napredujejo skozi model sistema.
- Procesorji predstavljajo mobilne (npr. ljudje, viličarji itd.) in fiksne (npr. stroji, proizvodne linije itd.) vire, ki obdelujejo simulacijske enote.
- V čakalnih vrstah so enote pretoka shranjene do njihovega prehoda na naslednji razpoložljivi procesor.
- Konektorji določajo napredovanje enot po modelu sistema.

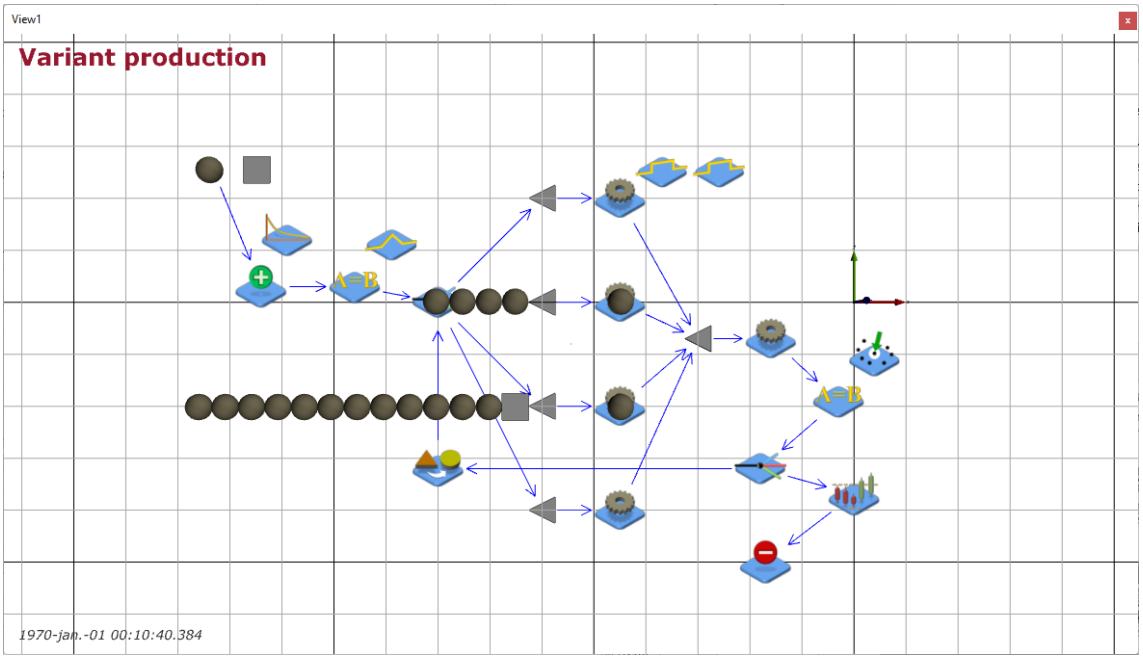
Lastnosti:

- Usmerjenost v proces.
- Osredotoča se na podrobno modeliranje procesov.
- Heterogeni subjekti.
- Mikrosubjekti so pasivni predmeti.
- Dogodki v sistem vnašajo dinamiko.
- Diskretno časovno napredovanje; od enega (časovnega) dogodka do naslednjega.
- Prilagodljivost je dosežena s spreminjanjem strukture modela; struktura sistema med simulacijo je fiksna.

Primer

Primer DES (slika 4.1, povzeta iz simulacijskega okolja JaamSim (JaamSim Development Team, 2023)) vključuje model proizvodnje variant, kjer se proizvajajo štirje različni izdelki (Gumzej in Rakovska, 2020). V skladu s proizvodnim načrtom se proizvede približno 10, 30, 40 in 20 % vrst izdelkov 1, 2, 3 oziroma 4. Izbira vrste izdelka je posledica trikotniške porazdelitve med 1 in 4 z modulom pri 3. Vsaka vrsta izdelka ima posebno proizvodno linijo. Proizvodna naročila se izpolnijo

v skladu z eksponentno porazdelitvijo okoli 30-s povprečne časovne vrednosti. Proizvodnja vsakega posameznega izdelka traja 100-120 s v skladu z enakomerno porazdelitvijo. Ko so izdelki dokončani, se njihova kakovost preveri na posebnem testnem mestu. Preverjanje kakovosti traja 10 s. Po izkušnjah podjetja v povprečju vsak 1 od 10 izdelkov ne ustreza zahtevam pregleda. Izdelki nezadostne kakovosti se prepeljejo nazaj na prvotno proizvodno linijo. Njihova ponovna obdelava traja 120-130 s glede na enakomerno porazdelitev. Trajanje proizvodnje in pregleda kakovosti ter ponovne obdelave ni odvisno od vrste izdelka. Po uspešno opravljenem pregledu kakovosti se končni izdelki prepeljejo s proizvodnega mesta v skladišče končnih izdelkov. Ponovna proizvodnja izdelkov z napako, ko so ti še v proizvodnji, je učinkovit način za zmanjšanje vplivov na okolje in proizvodnih stroškov.



Slika 4.1 Proizvodnja variant z nadzorom kakovosti.

Povzetek

Z DES je mogoče analizirati in optimizirati naslednje parametre postopka:

- Čas proizvodnega cikla in učinkovitost.
- Uporaba proizvodnih celic in prostorov.
- Zmogljivost prostorov za shranjevanje in čas bivanja enote za shranjevanje.
- Uporaba mobilnih virov (npr. upravljavcev, transporterjev, viličarjev).

4.3 Dinamika sistema

Analiza SD predstavlja pogled vodje SC na proizvodni proces z doslednim in skladnim modelom. SD velja za orodje, ki je najprimernejše za določanje strukture in optimalnih količin (kdaj in koliko) vhodov, zalog in izhodov posameznega mesta. Zato omogoča učinkovito uporabo proizvodnih in skladiščnih zmogljivosti.

Konstrukti:

- Zaloge predstavljajo rezervoarje, v katerih se lahko shranjujejo predmeti dobave v oskrbovalni verigi.
- Tokovi predstavljajo dobavne poti.
- Povratne zanke predstavljajo parametre za natančno uravnavanje dopolnjevanja zalog.

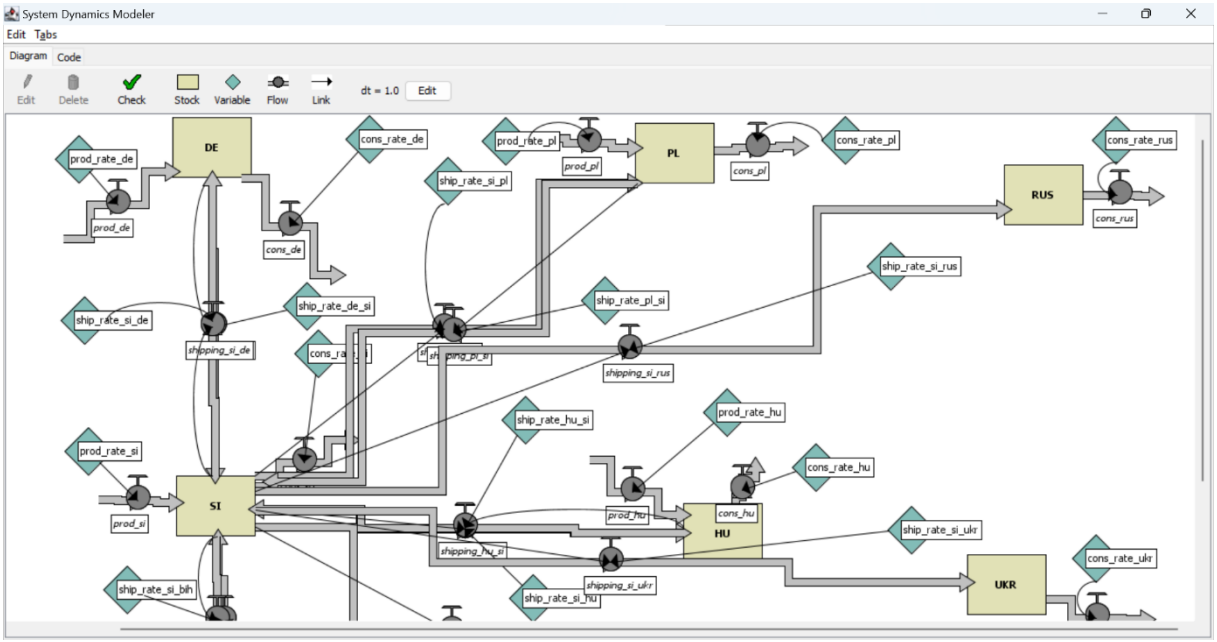
Lastnosti:

- Osredotočenost na sistem.
- Ključni kazalniki uspešnosti, usmerjeni v modeliranje sistemskih spremenljivk.
- Homogene enote.
- Subjekti na mikroravni se ne upoštevajo.
- Dinamika se uvaja s povratno zanko.
- Neprekinjeno časovno napredovanje; čas napreduje sinhrono za vse komponente modela sistema.
- Prilagodljivost se doseže s spreminjanjem strukture modela.
- Struktura sistema med simulacijo je fiksna.

Primer

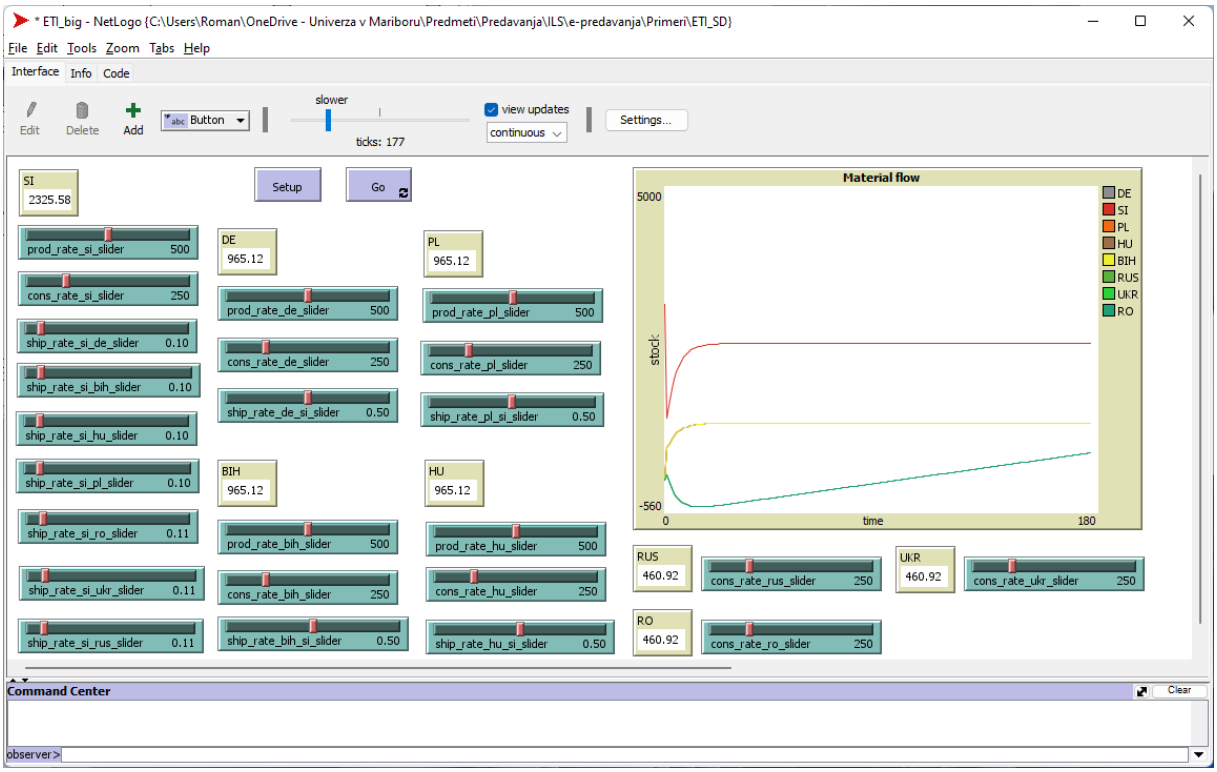
Primer SD (slika 4.2, povzeta iz simulacijskega okolja NetLogo (Wilensky, 1999)) zajema SC podjetja za gospodinjske aparate in opisuje materialne tokove med njegovimi podružnicami (Gumzej in Rakovska, 2020). Podjetje ima več proizvodnih obratov: glavni obrat v Sloveniji (SI) ter hčerinska podjetja v Nemčiji (DE), na Poljskem (PL), Madžarskem (H) ter v Bosni in Hercegovini (BIH). Poleg proizvodnih obratov so obrati za bruto prodajo v Rusiji (RUS), Ukrajini (UKR) in

Romuniji (RU). Proizvodni obrati oskrbujejo svoje trge s končnimi izdelki, drug drugemu pa dobavljajo sestavne dele izdelkov.



Slika 4.2 Razporeditev SC.

Povezana nadzorna plošča NetLogo (slika 4.3) služi kot orodje za podporo odločanju (DST), ki omogoča povezovanje količin proizvodnje in zalog s predispozicijami in njihovo fizično porazdelitvijo. Časovni tok je neprekinjen skozi vsakodnevne transakcije, tj. vsak dan se določeno število sestavnih delov pošlje med proizvodnimi lokacijami in določeno število končnih izdelkov se porabi na lokaciji ali pošlje na distribucijske lokacije. Na podlagi začetne zaloge 300 enot na lokaciji SI in 0 zalog na drugih lokacijah ter distribucijskega modela količine zalog na posameznih lokacijah predstavljajo povprečno zalogo glede na dane stopnje proizvodnje (v kosih), porabe (v %) in pošiljanja (v %).



Slika 4.3 Nadzorna plošča SC.

Sinopsis

Simulacija sistemske dinamike omogoča:

- Načrtovanje postavitve SC.
- Optimizacija proizvodnih in distribucijskih zmogljivosti.
- Ocena obremenitve distribucijskih kanalov in s tem povezanih stroškov.

4.4 Simulacija na podlagi agentov

Analiza ABS ponuja pogled strateškega menedžerja ali tržnega regulatorja na trg. Zato ABS velja za orodje, ki je najprimernejše za določanje optimalne strukture in razporeditve/asortimenta posameznega trga in/ali SC ob upoštevanju njihovih globalnih značilnosti (npr. demografija, podnebje, BDP, kakovost, ozaveščenost itd.).

Konstrukcije:

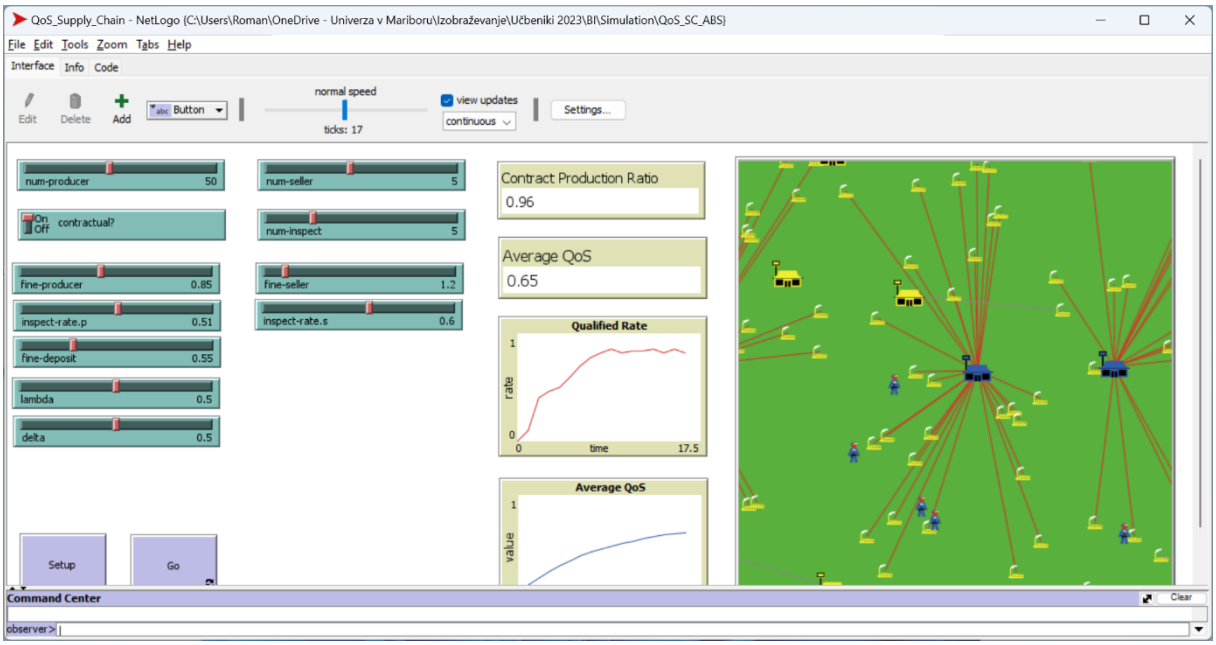
agenti, ki predstavljajo vozlišča oskrbovalne verige (npr. dobavitelji, trgovci na drobno in inšpektorji) s svojimi lastnostmi, odnosi in obnašanjem.

Lastnosti:

- Osredotočenost na entiteto.
- Problemsko usmerjeno modeliranje entitet in njihovih interakcij.
- Heterogenost subjektov.
- Mikro-objekti so aktivni predmeti, ki delujejo v svojem okolju, komunicirajo med seboj in samostojno sprejemajo odločitve.
- Odločitve in interakcije med agenti vnašajo dinamiko v sisteme.
- Agenti in njihova okolja so formalni modeli.
- Časovni tok je diskreten in univerzalen na ravni modela; časovni potek modela je skladen s pogostostjo transakcij SC in življenjskimi cikli vozlišč SC.
- Prilagodljivost modela je dosežena s spreminjanjem strukture sistema in obnašanja agentov.
- Struktura sistema med simulacijo je spremenljiva.

Primer

Primer ABS (slika 4.4, povzet iz simulacijskega okolja NetLogo (Wilensky, 1999)) je bil uporabljen za analizo obnašanja ešalonov SC na odprtem trgu (Gumzej in Rakovska, 2020) glede na njihovo kakovost storitev (QoS). V tem primeru so bile različne politike v zvezi z upravljanjem celovite kakovosti podjetja raziskane z modelom, ki je vključeval njegove dobavitelje, odjemalce in tržne regulatorje.



Slika 4.4 Ureditev trga.

Sinopsis

Simulacija, ki temelji na agentih, omogoča:

- Načrtovanje postavitve SC.
- Modeliranje dinamične rasti SC.
- Modeliranje obnašanja partnerjev v SC.
- Optimizacija globalnih kazalnikov.

4.5 Simulacija omrežja

Analiza NS ponuja pogled regulatorja omrežja na omrežje. Zato NS velja za najprimernejše orodje za določanje optimalne strukture, postavitve in izbora omrežja ob upoštevanju njegovih globalnih značilnosti (npr. prepustnosti, emisij, kazalnikov QoS itd.).

Konstrukcije:

Agenti, ki predstavljajo predmete toka z njihovimi lastnostmi, odnosi in obnašanjem.

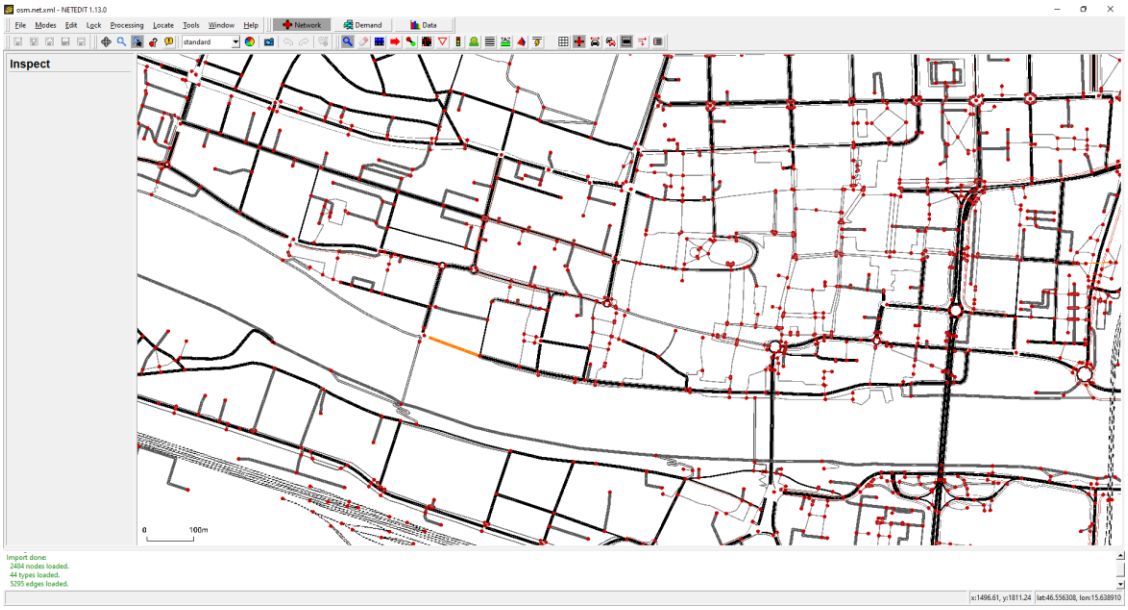
Omrežje, ki predstavlja prekrivno omrežje (npr. prometno omrežje), v katerem se predmeti pretoka pretakajo.

Lastnosti:

- Osredotočenost na sistem.
- Problemsko usmerjeno modeliranje entitet in njihovih interakcij.
- Heterogenost subjektov.
- Mikroobjekti so aktivni predmeti, ki delujejo v svojem okolju, komunicirajo med seboj in samostojno sprejemajo odločitve.
- Odločitve in interakcije med agenti vnašajo dinamiko v sisteme.
- Agenti in njihova okolja so formalni modeli.
- Časovni tok je diskreten in univerzalen na ravni modela; časovni potek je skladen z relativnimi hitrostmi predmetov toka.
- Prilagodljivost modela je dosežena s spreminjanjem strukture omrežja, ki je med simulacijo nespremenljiva, in obnašanja agentov, ki se spreminja glede na stanje (prometnega) omrežja in njihove cilje.

Primer

Predstavljeni primer (slika 4.5, vzeta iz simulacijskega okolja SUMO (Pablo et.al., 2018)) je bil uporabljen za določitev prometnih tokov in prepustnosti ulic v mestnem središču, ki jih je prizadela načrtovana zapora ceste (Šinko in Gumzej, 2021). Poleg tega so bili izmerjeni s prometom povezani kazalniki, kot so potovalni časi, poraba goriva in emisije.



Slika 4.5 Prometne razmere in omrežje.

Sinopsis

Simulacija omrežja omogoča:

- Načrtovanje postavitve omrežja.
- Modeliranje dinamičnega obnašanja omrežja za določitev ozkih grl in šibkih členov.
- Modeliranje pretoka omrežnih elementov.
- Optimizacija kazalnikov globalnega omrežja.

4.6 Logistični simulacijski projekti

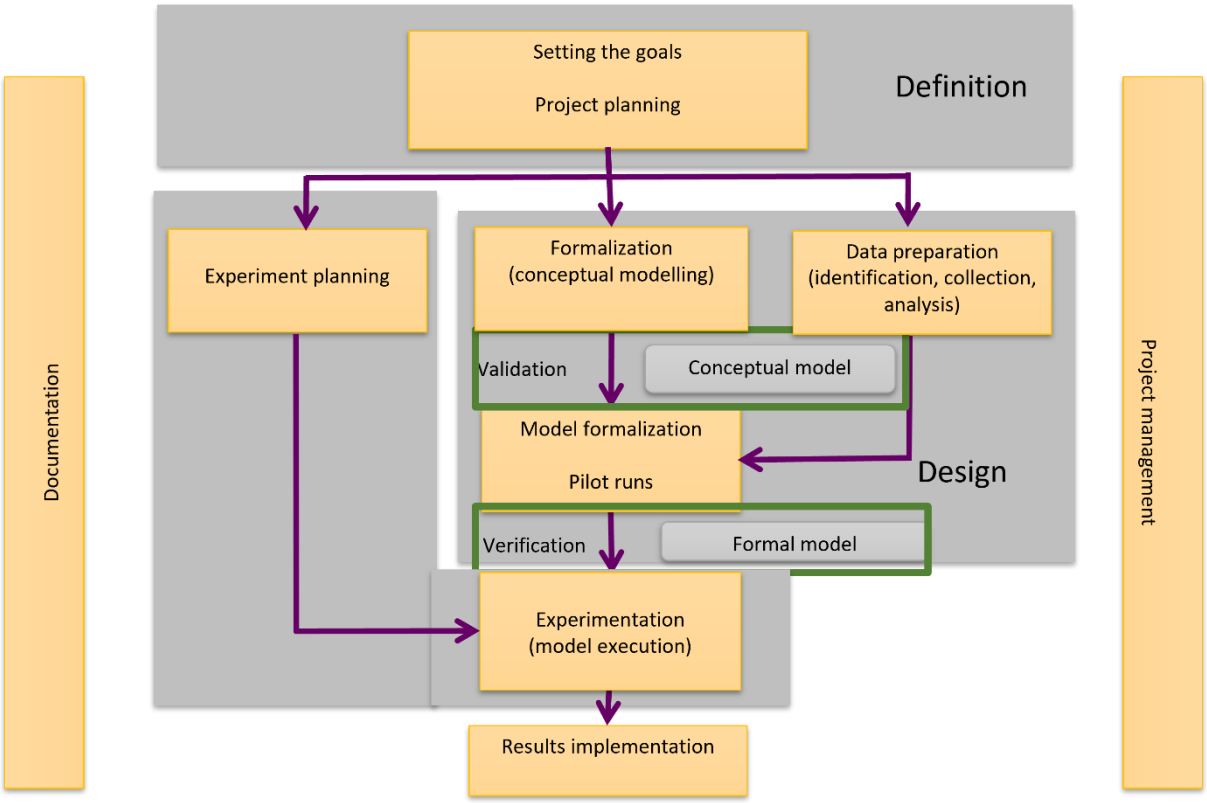
Logistični simulacijski projekti so zasnovani v skladu s paradigmo Design for Six Sigma (DFSS) in temeljijo na Demingovem ciklu izboljšav:

- Načrtovanje: opredelitev sistema in ciljev.
- Izvedba: oblikovanje simulacijskega modela.
- Analiza: eksperimentiranje s simulacijskim modelom in ocenjevanje alternativ.
- Ukrep: uporaba rezultatov simulacije za izvajanje izboljšav.

Vsak projekt logistične simulacije (slika 4.6) je sestavljen iz sedmih faz:

1. Strateški načrt: analiza obstoječih in predlaganih virov in procesov.

2. Konceptualni model: opredelitev abstraktnega modela sistema in predispozicij, zbiranje podatkov.
3. Logični model: diagram pretoka objektov, zalog in tokov ali mrežni diagram modela sistema.
4. Simulacijski model: razvoj ustreznega simulacijskega modela.
5. Preverjanje in potrjevanje simulacijskega modela: preverjanje doslednosti in skladnosti modela.
6. Analiza na podlagi simulacijskega modela: načrtovanje in izvedba poskusov.
7. Uporaba rezultatov simulacije za pripravo akcijskega načrta: projekcija izboljšav sistema.



Slika 4.6 Postopek simulacijskega modeliranja in analize (SMA).

4.7 Zaključek

V logistiki je SMA pomembna sestavina operacijskih raziskav (OR), ki omogoča optimizacijo procesov.

Sistemska dinamika (SD) je metodologija za analizo kompleksnih, dinamičnih in nelinearnih interakcij v sistemih, ki omogoča oblikovanje novih struktur in politik za izboljšanje obnašanja

sistema. Pri tem se obravnavajo fizični in informacijski tokovi, da bi se zmanjšala njihova zamuda, in nazadnje zaloge SC.

Druga priljubljena procesno usmerjena metodologija je simulacija diskretnih dogodkov (DES). To je eno od najbolj razširjenih in prilagodljivih analitičnih orodij v SMA proizvodnih sistemov. Uspešno obravnava negotovost in zagotavlja možnosti za primerjavo alternativnih načinov za skrajšanje dobavnega časa ter optimizacijo izkoriščenosti strojev in virov.

Uporabna metodologija za razumevanje vedenja organizacij in njihovih interakcij (npr. SC in njihovih entitet) je simulacija na podlagi agentov (ABS). Simulacija omrežja (NS), ki je posebna vrsta ABS, omogoča modeliranje in optimizacijo (prometnih) prekrivnih omrežij.

To vodi do zaključka, da celostni pristop k uporabi SMA v logistiki pomembno prispeva k odločitvam o oblikovanju kompleksnih sistemov, kjer je veliko spremenljivk, ki medsebojno vplivajo druga na drugo. Uporaben celostni pristop, ki vključuje metodologije SD, DES in ABS ter lahko količinsko opredeli delovne tokove na različnih ravneh oskrbovalne verige, je bil predstavljen v (Gumzej in Rakovska, 2020). Pri analizi in optimizaciji prometnih tokov metodologija NS zagotavlja potreben okvir glede spremljanja in natančnega prilagajanja ključnih kazalnikov uspešnosti (Šinko in Gumzej, 2021).

Literatura poglavje 4

- 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 z enim in več regresorji

Upravljavske odločitve pogosto temeljijo na razmerju med dvema ali več spremenljivkami. Vodja trženja lahko na primer poskuša napovedati prodajo pri določeni ravni izdatkov za oglaševanje, potem ko preuči razmerje med temi izdatki in prodajo.

V drugem primeru lahko javno podjetje za napovedovanje porabe električne energije uporabi razmerje med najvišjo dnevno temperaturo in povpraševanjem po električni energiji. Včasih se upravljavec zanaša na intuicijo. Intuitivno oceni, kako sta spremenljivki povezani. Če pa je mogoče pridobiti podatke, je smiselno uporabiti statistični postopek, imenovan regresijska analiza, ki pokaže, kako sta spremenljivki povezani med seboj.

V regresijski terminologiji se napovedana spremenljivka imenuje odvisna spremenljivka.

Spremenljivka ali spremenljivke, ki se uporabljajo za napovedovanje vrednosti odvisne spremenljivke, se imenujejo neodvisne spremenljivke.

Pri analizi učinka izdatkov za oglaševanje na prodajo bi bila torej prodaja odvisna spremenljivka. Neodvisna spremenljivka so izdatki za oglaševanje. V statističnem zapisu y označuje odvisno spremenljivko, in x označuje neodvisno spremenljivko.

V tem razdelku si bomo ogledali najpreprostejšo vrsto regresijske analize, ki vključuje eno neodvisno spremenljivko in eno odvisno spremenljivko. Razmerje med spremenljivkama se približa s premico. To se imenuje preprosta linearna regresija. Regresijska analiza, ki vključuje dve ali več neodvisnih spremenljivk, se imenuje multipla regresijska analiza.

5.1 Enostavni linearni regresijski model

Best Burger je veriga restavracij s hitro prehrano na območju več držav. Lokacije Best Burger so v bližini univerzitetnih kampusov. Poslovodje menijo, da četrtletna prodaja teh restavracij (označena z y) je pozitivno povezana z velikostjo študentske populacije (označeno z x). Restavracije v bližini študentskih kampusov z velikim številom študentov običajno ustvarijo večjo prodajo kot restavracije v bližini kampusov z

majhnim številom študentov. Z regresijsko analizo lahko razvijemo enačbo, ki pokaže, kako odvisna spremenljivka y povezana z neodvisno spremenljivko x.

5.2 Regresijski model in regresijska enačba

V primeru restavracije Best Burger so prebivalstvo vse restavracije Best Burger. Za vsako restavracijo v populaciji obstaja vrednost x (študentska populacija) in ustrezna vrednost y (četrtletna prodaja). Enačba, ki opisuje, kako y je povezana z x se imenuje regresijski model.

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

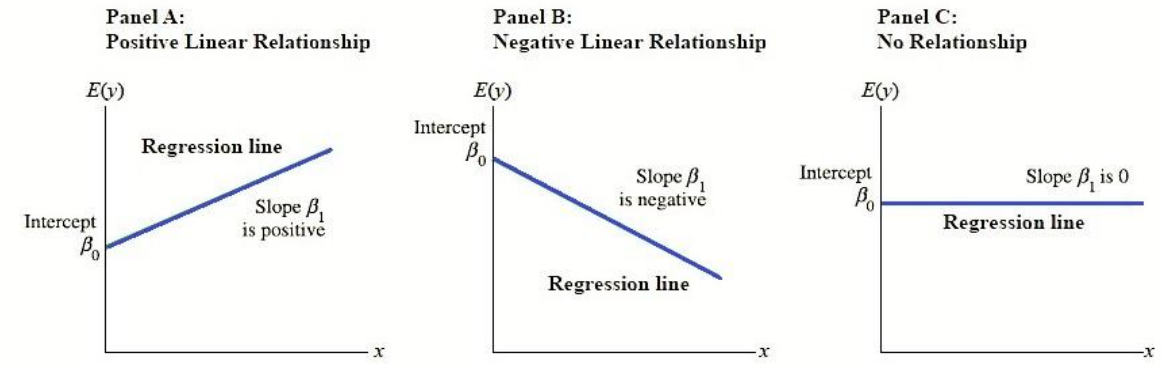
β_0 in $\beta_1$imenujemo parametri modela, ϵ (grška črka epsilon) je naključna spremenljivka, imenovana napaka modela. Napaka predstavlja variabilnost y ki je ni mogoče pojasniti z linearno povezavo med $\mathrm{x\ in\ y}$.

Populacijo vseh restavracij Best Burger lahko obravnavamo tudi kot zbirko pod-populacij, po eno za vsako vrednost posebej. x. Ena podpopulacija je na primer sestavljena iz vseh restavracij Best Burger v bližini univerzitetnih kampusov z 8000 študenti. Drugo podpopulacijo sestavljajo vse restavracije Best Burger v bližini univerzitetnih kampusov z 9000 študenti in tako naprej. Vsaka podpopulacija ima ustrezno porazdelitev vrednosti y. Vsaka porazdelitev vrednosti y ima svojo srednjo ali pričakovano vrednost. Enačba, ki opisuje pričakovano vrednost, je y, označena z $E(y)$, ki je povezana z x se imenuje regresijska enačba. Regresijska enačba za preprosto linearno regresijo je naslednja

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

Graf preproste linearne regresijske enačbe je ravna črta. β_0 predstavlja začetno vrednost regresijske premice, β_1 je smerni koeficient premice in $E(y)$ srednja vrednost ali pričakovana vrednost y za dano vrednost x.

Primeri možnih regresijskih linij so prikazani na spodnji sliki 5.1. Regresijska črta v primeru A kaže, da je vrednost y je pozitivno povezana z x. Ko se vrednosti povečujejo x, se vrednosti povečujejo tudi $E(y)$. Regresijska premica v primeru B kaže, da je vrednost y, ki je negativno povezana z x. Kadar so manjše vrednosti $E(y)$ so povezane z višjimi vrednostmi x. Regresijska črta v panelu C prikazuje primer, ko je vrednost y ni povezana z x. To pomeni, da je vrednost y je enaka za vsako vrednost x.



Slika 5.1 Primeri grafov linearnega razmerja.

5.3 Ocenjena regresijska enačba

Če bi bile vrednosti parametrov populacije znane β_0 in β_1, bi lahko z zgornjo enačbo izračunali vrednosti y za dano vrednost x. V praksi so ti parametri težko dostopni, zato jih preprosto ocenimo z uporabo vzorčnih podatkov. Vzorčna statistika (označena z b_0 in b_1) se izračunajo kot ocene populacijskih parametrov β_0 in β_1. Zamenjava vrednosti vzorčnih statistik b_0 in b_1 namesto $\beta_0$in β_1v regresijski enačbi dobimo novo, ocenjeno regresijsko enačbo. Ocenjena regresijska enačba za enostavno linearno regresijo je naslednja

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

Graf ocenjene enostavne linearne regresije se imenuje ocenjena regresijska črta. b_0 predstavlja začetno vrednost regresijske premice, b_1 je smerni koeficient premice.

V nadaljevanju prikazujemo, kako z metodo najmanjših kvadratov izračunati vrednosti b_0 in b_1 v ocenjeni regresijski enačbi.

Na splošno $\hat{y}$ (ocena za $E(y)$) povprečna vrednost y za dano vrednost x. Če bi zdaj želeli oceniti pričakovano vrednost četrtletne prodaje za vse restavracije Best Burger, ki se nahajajo v bližini kampusov z 10000 študenti, bi bila vrednost x v zadnji enačbi nadomestili z vrednostjo 10000. V nekaterih primerih pa nas morda bolj zanima napoved prodaje samo za eno določeno restavracijo. Recimo, da želite napovedati četrtletno prodajo za restavracijo, ki jo nameravate zgraditi v bližini kolidža z 10000 študenti. Kot se izkaže, je tudi v tem primeru najboljši napovedovalec vrednosti y za dano vrednost x vrednost $\hat{y}$.

5.4 Metoda najmanjših kvadratov

Metoda najmanjših kvadratov je postopek, pri katerem na podlagi vzorčnih podatkov poiščemo enačbo ocenjene regresijske premice. Za ponazoritev metode najmanjših kvadratov predpostavimo, da so bili podatki zbrani na vzorcu 10 restavracij Best Burger v bližini univerzitetnih kampusov. S x_i označimo velikost študentske populacije (v tisočih) in z y_i velikost četrtletne prodaje (v tisočih EUR). $x_i \, in \, y_i$ za 10 vzorčnih restavracij so povzete v spodnji preglednici. Vidimo, da je restavracija 1, z x_1 = 2 in y_1 = 58, blizu kampusa z 2000 študenti in ima četrtletno prodajo v višini 58 000 EUR. Restavracija 2, z x_2 = 6 in y_2 = 105, je blizu kampusa s 6000 študenti in ima četrtletno prodajo 105.000 €. Restavracija z najvišjo vrednostjo prodaje je restavracija 10, ki je blizu kampusa s 26.000 študenti in ima četrtletno prodajo 202.000 €.

V nadaljevanju je prikazana razpršena slika podatkov na sliki 5.2 spodaj. Na vodoravni osi je prikazana populacija študentov, na navpični osi pa četrtletna prodaja. Diagrama razpršitve za regresijsko analizo sta sestavljena z neodvisno spremenljivko x na vodoravni osi in odvisno spremenljivko y na navpični osi. Diagram razpršitve nam tako omogoča, da naredimo predhodne sklepe o možnem razmerju med spremenljivkama.

Restaurant i	Student Population (1000s) x_i	Quarterly Sales (€1000s) 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 5.2 Diagram razpršitve podatkov.

Kakšne predhodne ugotovitve lahko potegnemo iz spodnje slike 5.3? V kampusih z večjim številom študentov je četrtletna prodaja višja. Poleg tega obstaja konstantna povezava med številom študentov in četrtletno prodajo, ki jo je mogoče opisati z ravno črto. Med $x \, in \, y$ pozitivnim linearnim razmerjem je res mogoče sklepati na pozitivno linearno razmerje. Zato smo izbrali preprost linearni regresijski model za predstavitev povezave med četrtletno prodajo in številom študentov. Glede na to izbiro je naša naslednja naloga, da s pomočjo preglednice z vzorčnimi podatki določimo vrednosti b_0 in b_1, ki sta

pomembna parametra pri ocenjevanju enačbe preproste linearne regresije. Za i-to restavracijo je ocenjena regresijska enačba naslednja

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

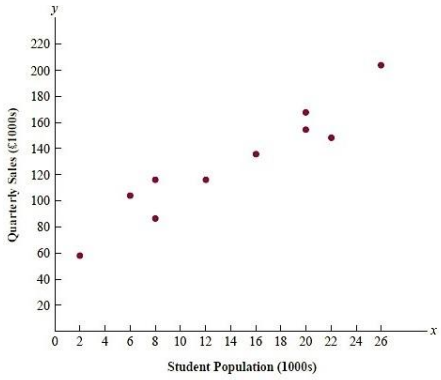
Kje:

$\hat{y}_i$ _ ocenjena vrednost četrtletne prodaje (1000 EUR) za i-to restavracijo

b_0 _ začetna vrednost ocenjene regresijske premice

b_1 _ smerni koeficient ocenjene regresijske premice

x_i _velikost populacije učencev (1000) za i-to restavracijo



Slika 5.3 Graf razpršitve.

y_i označuje opazovano (dejansko) prodajo v restavraciji i in $\hat{y}_i$, ki predstavlja ocenjeno vrednost prodaje za restavracijo $i$vsaka restavracija v vzorcu bo imela opazovano vrednost prodaje y_i in predvideno vrednost prodaje $\hat{y}_i$. Da bi ocenjena regresijska premica zagotovila dobro prileganje podatkom, želimo, da so razlike med opazovanimi vrednostmi prodaje in napovedanimi vrednostmi prodaje čim manjše.

Metoda najmanjših kvadratov uporablja vzorčne podatke, ki zagotavljajo vrednosti b_0 in b_1.

Minimizirati vsoto kvadratov odstopanj med opazovanimi vrednostmi odvisne spremenljivke y_i in napovedano vrednostjo odvisne spremenljivke $\hat{y}_i$. Izhodišče za izračun najmanjše vsote po metodi najmanjših kvadratov je podano z izrazom

Merilo najmanjše vsote: $\min \sum(y_i - \hat{y}_i)^2$

Kje:

y_i = opazovana vrednost odvisne spremenljivke za i-to opazovanje

$\hat{y}_i$ =napovedana vrednost odvisne spremenljivke za i-to opazovanje

smerni koeficient regresijske premice in začetna 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 neodvisne spremenljivke za i-to opazovanje

y_i _ vrednost odvisne spremenljivke za i-to opazovanje

$\bar{x}$ _ povprečna vrednost za neodvisno spremenljivko

$\bar{y}$ _ povprečna vrednost odvisne spremenljivke

n _skupno število opazovanj

V nadaljevanju so prikazani nekateri izračuni, potrebni za izračun ocenjene regresijske premice najmanjših kvadratov. Pri vzorcu 10 restavracij imamo n=10 opazovanj. Zgornje enačbe najprej zahtevajo izračun povprečne vrednosti x in povprečno vrednost 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 enačba za izračun b_1:

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

Z uporabo zadnjih enačb in informacij na sliki 5.4 lahko izračunamo smerni koeficient regresijske premice za primer restavracije Best Burger. Izračun naklona (b_1) je naslednje.

Slika 5.5 prikazuje graf te enačbe na diagramu razpršitve.

Naklon ocenjene regresijske enačbe ali smerni koeficient enačbe ($b_1 = 5$) je pozitiven.

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

Temu sledi izračun začetne vrednosti (b_0).

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

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

Slika 5.4 Graf enačbe na diagramu razpršitve.

Tako se oceni regresijska enačba:

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

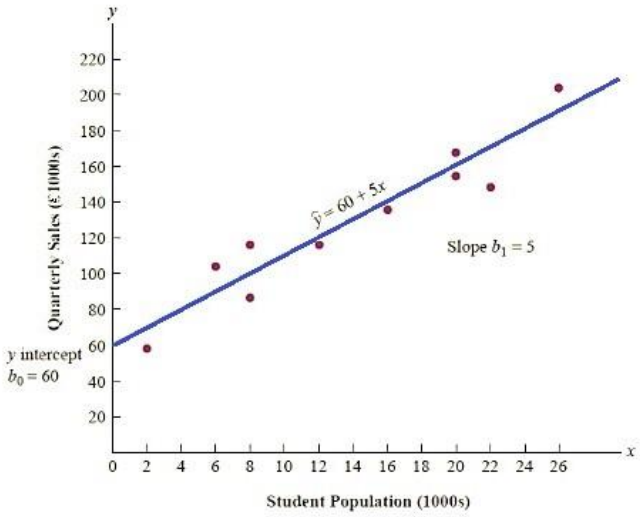
Na sliki je prikazan graf te enačbe na diagramu razpršitve.

Naklon ocenjene regresijske enačbe (b_1 = 5) je pozitivna, kar pomeni, da se s povečevanjem števila študentov povečuje tudi prodaja. Dejansko lahko sklepamo (na podlagi izmerjene prodaje v tisočakih in števila študentov v tisočakih), da je povečanje števila študentov za 1000 povezano s povečanjem pričakovane prodaje za 5000; tj. četrtletna prodaja naj bi se povečala za 5 EUR na študenta.

Če menimo, da regresijska enačba, ocenjena z metodo najmanjših kvadratov, ustrezno opisuje razmerje med x in y, se zdi smiselno, da z ocenjeno regresijsko enačbo napovemo vrednost y za dano vrednost x. Na primer, če bi želeli napovedati četrtletno prodajo za restavracijo, ki je v bližini kampusa s 16.000 študenti, bi izračunali z

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

Zato predvidevamo, da bo četrtletna prodaja te restavracije znašala 140.000 evrov. V naslednjih razdelkih obravnavamo metode za ocenjevanje ustreznosti uporabe ocenjene regresijske enačbe za ocenjevanje in napovedovanje.



Slika 5.5 Graf razpršitve števila študentov in četrtletne prodaje.

5.5 Koeficient determinacije

Za primer restavracij Best Burger smo razvili ocenjeno regresijsko enačbo $y = 60 + 5x$ za približno linearno razmerje med velikostjo študentske populacije x in četrtletno prodajo y. Vprašanje je, kako dobro se ocenjena regresijska enačba ujema s podatki. V tem razdelku bomo pokazali, da koeficient determinacije zagotavlja merilo dobrega ujemanja ocenjene regresijske enačbe. Za i-to opazovanje je razlika med opazovano vrednostjo odvisne spremenljivke $y_i$in napovedano vrednostjo odvisne spremenljivke imenujemo i-ti ostanek.

Vsota kvadratov teh ostankov ali napak je količina, ki se minimizira z metodo najmanjših kvadratov. Ta količina, znana tudi kot vsota kvadratov napak, je označena s SSE.

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

Vrednost SSE je merilo napake pri uporabi ocenjene regresijske enačbe za napovedovanje vrednosti odvisne spremenljivke v vzorcu. Slika 5.6 prikazuje izračune, potrebne za izračun vsote kvadratov zaradi napake za primer Best Burger.

Restaurant i	x_i = Student Population (1000s)	y_i = Quarterly Sales (€1000s)	Predicted Sales $\hat{y}_i = 60 + 5x_i$	Error $y_i - \hat{y}_i$	Squared Error $(y_i - \hat{y}_i)^2$
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 5.6 Kvadrati napak v primeru Best Burger.

Recimo, da moramo pripraviti oceno četrtletne prodaje, ne da bi poznali velikost študentske populacije. Brez poznavanja povezanih spremenljivk bi kot oceno četrtletne prodaje v kateri koli restavraciji uporabili povprečje vzorca. Tabela na sliki 5.6 je pokazala, da za podatke o prodaji y_i =1300. Zato je povprečna vrednost četrtletne prodaje za vzorec 10 restavracij Best Burger y_i/n = 1300/10 = 130. V preglednici 14.4 prikazujemo vsoto kvadratov odstopanj, ki jih dobimo z uporabo vzorčnega povprečja 130 za napovedovanje vrednosti četrtletne prodaje za vsako restavracijo v vzorcu. Za i-to restavracijo v vzorcu je razlika y_i predstavlja merilo napake, ki je vključena v aplikacijo za napovedovanje prodaje. Ustrezna vsota kvadratov, imenovana skupna vsota kvadratov, je označena s SST.

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

Restaurant i	x_i = Student Population (1000s)	y_i = Quarterly Sales (€1000s)	Deviation $y_i - \bar{y}$	Squared Deviation $(y_i - \bar{y})^2$
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

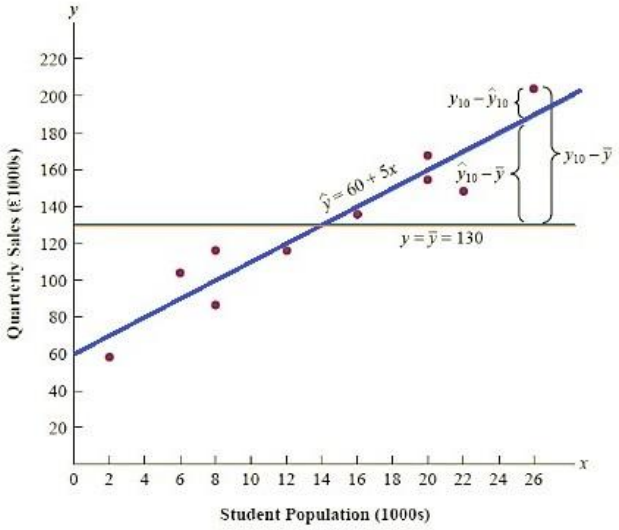
Slika 5.7 Vsota kvadratov.

Vsota na dnu zadnjega stolpca na sliki 5.7 je skupna vsota kvadratov za restavracije BestBurger SST = 15 730. Na sliki 5.8 je prikazana ocenjena regresijska črta $y = 60 + 5x$ in črto, ki ustreza y = 130. Upoštevajte, da so točke bolj zbrane okoli ocenjene regresijske premice kot okoli premice y = 130. Na primer za deseto restavracijo v vzorcu vidimo, da je napaka veliko večja, če za

napovedovanje y = 10 uporabimo 130, kot če uporabimo 130 $y = 60 + 5x$ in znaša 190. Na SST lahko gledamo kot na merilo, kako dobro so opazovanja zbrane okoli premice, na SSE pa kot na merilo, kako dobro so opazovanja zbrana okoli premice.

Da bi izmerili, koliko vrednosti na ocenjeni regresijski premici odstopajo od naslednjih, se izračuna še ena vsota kvadratov. Ta vsota kvadratov, imenovana vsota kvadratov zaradi regresije, se označi kot SSR.

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



Slika 5.8 Regresijska črta za primer Best Burger.

Na podlagi prejšnje razprave lahko pričakujemo, da so SST, SSR in SSE povezani. Dejansko je povezava med temi tremi vsotami kvadratov eden najpomembnejših rezultatov v statistiki.

5.6 Razmerje med SST, SSR in SSE:

$$SST = SSR + SSE$$

Kje je:

SST = skupna vsota kvadratov

SSR = vsota kvadratov zaradi regresije

SSE = vsota kvadratov zaradi napake

Enačba ($SST = SSR + SSE$) kaže, da lahko skupno vsoto kvadratov razdelimo na dve komponenti, vsoto kvadratov zaradi regresije in vsoto kvadratov zaradi napake. Če sta torej znani vrednosti

katerih koli dveh od teh vsot kvadratov, lahko tretjo vsoto kvadratov zlahka ugotovimo z izračunom. V primeru restavracij Best Burger na primer že vemo, da je SSE = 1530 in SST = 15,730; zato z reševanjem za SSR v zgornji enačbi ugotovimo, da je vsota kvadratov zaradi regresije

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

Poglejmo, kako lahko s pomočjo treh vsot kvadratov, SST, SSR in SSE, zagotovimo merilo ustreznosti ocenjene regresijske enačbe. Ocenjena regresijska enačba bi se popolnoma ujemala, če bi vsaka vrednost odvisne spremenljivke $yi\_$ ležala naključno na ocenjeni regresijski premici. V tem primeru bi bila za vsako opazovanje enaka nič, kar bi pomenilo SSE =0. Ker je SST = SSR + SSE, vidimo, da mora biti za popolno ujemanje SSR enak SST, razmerje (SSR/SST) pa mora biti enako ena. Slabše prileganje bo imelo za posledico večje vrednosti SSE. Če v enačbi (14.11) rešimo SSE, vidimo, da je SSE = SST - SSR. Zato je največja vrednost SSE (in s tem najslabše prileganje), ko je SSR = 0 in SSE = SST.

Za ocenjevanje se uporablja razmerje SSR/SST, ki ima vrednosti med nič in ena, ki ustrezajo ocenjeni regresijski enačbi.

To razmerje se imenuje koeficient determinacije in se označuje z r^2.

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

Za primer restavracij Best Burger je vrednost koeficienta determinacije

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

Ko je koeficient determinacije izražen v odstotkih, lahko r^2 lahko interpretiramo kot odstotek skupne vsote kvadratov, ki ga lahko pojasnimo z ocenjeno regresijsko enačbo. Za restavracije Best Burger Restaurants lahko sklepamo, da je 90,27 % skupne vsote kvadratov mogoče pojasniti z ocenjeno regresijsko enačbo. $y = 60 + 5x$ za napoved četrtletne prodaje. Z drugimi besedami, 90,27 % variabilnosti prodaje je mogoče pojasniti z linearno povezavo med velikostjo študentske populacije in prodajo. Zadovoljni moramo biti, da se tako dobro ujema z ocenjeno regresijsko enačbo.

5.7 Korelacijski koeficient

Korelacijski koeficient je opisno merilo moči linearne povezave med dvema spremenljivkama, x in y, z vrednostmi med -1 in +1. Vrednost +1 pomeni popolno pozitivno linearno povezavo, kjer vse podatkovne točke ležijo na premici a s pozitivnim naklonom, medtem ko vrednost -1 nakazuje popolno negativno linearno povezavo, pri kateri vse točke ležijo na premici a z negativnim naklonom. Vrednosti blizu nič kažejo, da x in y nista linearno povezani.

Če je regresijska analiza že izvedena in je koeficient determinacije r^2 znan, lahko koeficient korelacije vzorca izračunamo na naslednji način

$$r_{xy} = (sign\ of\ b_1)\sqrt{coefficient\ of\ determination}$$

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

PEARSONOV KORELACIJSKI KOEFICIENT: VZORČNI PODATKI

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

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

Kje so:

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

Znak za vzorec korelacijskega koeficienta je pozitiven, če ima ocenjena regresijska enačba pozitiven naklon ($b_1 > 0$) in negativen, če ima ocenjena regresijska enačba

negativni naklon ($b_1 < 0$).

Za primer Best Burger je vrednost koeficienta determinacije, ki ustreza ocenjeni regresijski enačbi $y = 60 + 5x$ znaša 0,9027. Ker je naklon ocenjene regresijske enačbe pozitiven, enačba (14.13) kaže, da je koeficient korelacije vzorca Z koeficientom korelacije vzorca

Rxy=0,9501, bi sklepali, da obstaja močna pozitivna linearna povezava med $x\ in\ y$.

V primeru linearnega razmerja med dvema spremenljivkama tako koeficienti determinacije kot koeficient vzorčne korelacije zagotavljajo merilo za moč razmerja.

Koeficient determinacije predstavlja mero med nič in ena, medtem ko koeficient vzorčne korelacije predstavlja mero med -1 in +1. Čeprav je koeficient vzorčne korelacije omejen na linearno razmerje med dvema spremenljivkama, se koeficient determinacije lahko uporablja za nelinearna razmerja in razmerja, ki imajo dve ali več neodvisnih spremenljivk. Tako koeficient determinacije omogoča širši obseg uporabe.

5.8 Večkratni regresijski model

V naslednjih razdelkih nadaljujemo s preučevanjem regresijske analize in obravnavamo situacije, ki vključujejo dve ali več neodvisnih spremenljivk. To tematsko področje, imenovano multipla regresijska analiza, nam omogoča, da upoštevamo več dejavnikov in tako dobimo boljše napovedi, kot so mogoče z enostavno linearno regresijo.

Večkratna regresijska analiza je študija povezanosti odvisne spremenljivke y z dvema ali več neodvisnimi spremenljivkami. V splošnem primeru bomo s p označili število neodvisnih spremenljivk.

5.9 Regresijski model in regresijska enačba

Pojma regresijski model in regresijska enačba, predstavljena v prejšnjem razdelku, veljata tudi v primeru večkratne regresije. Enačba, ki opisuje, kako je odvisna spremenljivka y povezana z neodvisnimi spremenljivkami $x_1, x_2, \ldots, x_p$ in izrazom napake, se imenuje model večkratne regresije. Začnemo s predpostavko, da ima model večkratne regresije naslednjo obliko.

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

V modelu multiple regresije so $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ parametri in izraz napake (ϵ) je naključna spremenljivka. Natančen pregled tega modela pokaže, da je y linearna funkcija spremenljivk $x_1, x_2, \ldots, x_p$ in izraza napake ϵ epsilon. Člen napake upošteva variabilnost y, ki je ni mogoče pojasniti z linearnim učinkom p neodvisnih spremenljivk.

V razdelku 5.10 obravnavamo predpostavke za model multiple regresije in epsilon. Ena od predpostavk je, da je povprečna ali pričakovana vrednost (ϵ) enaka nič. Posledica te predpostavke je, da je povprečna ali pričakovana vrednost y, označena z $E(y)$, je enaka $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots +$

$\beta_p x_p$. Enačba, ki opisuje, kako je srednja vrednost y povezana z $x_1, x_2, \ldots, x_p$ se imenuje enačba multiple regresije.

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

5.10 Ocenjena enačba multiple regresije

Če so vrednosti $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ znane, lahko enačbo (5.9) uporabimo za izračun povprečne vrednosti y pri danih vrednostih $x_1, x_2, \ldots, x_p$. Žal te vrednosti parametrov običajno niso znane in jih je treba oceniti iz vzorčnih podatkov. Za izračun vzorčne statistike se uporabi enostavni naključni vzorec $b_0, b_1, b_2, \ldots, b_p$ki se uporabijo kot točkovne ocene parametrov $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$. Te vzorčne statistike zagotavljajo naslednje ocenjevanje enačb multiple regresije:

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

Kjer so

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

$\hat{y} =$ napovedana vrednost odvisne spremenljivke

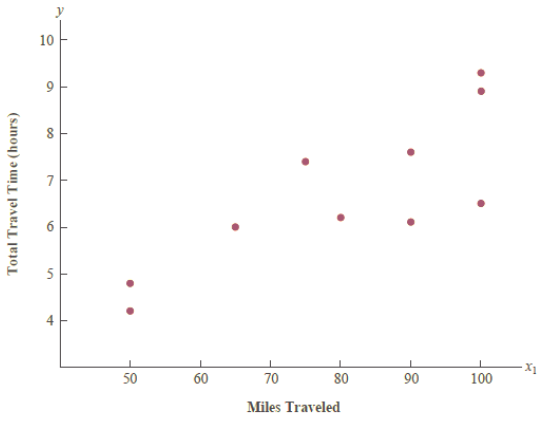
Primer: Prevozno podjetje Frigo

Za ponazoritev večkratne regresijske analize bomo obravnavali problem, s katerim se je soočalo podjetje Frigo Trucking Company, neodvisno prevozniško podjetje v južni Italiji. Največji del poslovanja podjetja Frigo vključuje dostavo na lokalnem območju. Za boljši razvoj delovnih razporedov želijo vodje načrtovati skupni dnevni potovalni čas za svoje voznike.

Sprva so vodje menili, da bo skupni dnevni čas potovanja tesno povezan s številom prevoženih kilometrov pri dnevnih dostavah. S preprostim naključnim vzorcem 10 vozniških nalog so bili pridobljeni podatki, prikazani na sliki 5.9, in razpršeni diagram. Po pregledu tega razpršenega diagrama so vodje domnevali, da je mogoče za opis povezave uporabiti preprost linearni regresijski model $y = \beta_0 + \beta_1 x_1 + \epsilon$ med skupnim časom vožnje (y) in številom prevoženih kilometrov (x_1).

Driving Assignment	x_1 = KM Traveled	y = Travel Time (hours)
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 5.9 Podatki za primer prevoznega podjetja Frigo.



Slika 5.10 Diagram razpršitve za primer podjetja Frigo Transport

Za oceno parametrov β_0 in . β_1 je bila za pripravo ocenjene regresije uporabljena metoda enačbe najmanjših kvadratov.

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

Na zgornji sliki je prikazan rezultat programa Minitab z uporabo preproste linearne regresije za podatke v zgornji tabeli. Ocenjena regresijska enačba je

$$\hat{y} = 1.27 + 0.0678 x_1$$

Pri stopnji pomembnosti 0,05 F-vrednost 15,81 in ustrezna p-vrednost 0,004 kažeta, da je razmerje pomembno. To pomeni, da lahko zavrnemo $H_0: \beta_1 = 0$ ker je p-vrednost manjša od $\alpha = 0{,}05$. Upoštevajte, da enak sklep izhaja iz vrednosti $t = 3{,}98$ in pripadajoče p-vrednosti 0,004. Tako lahko sklepamo, da je povezanost med skupnim časom potovanja in številom prevoženih kilometrov pomembna. Daljši čas potovanja je povezan z več prevoženimi kilometri. S koeficientom determinacije (izraženim v odstotkih) $R - Sq = 66{,}4\,\%$, vidimo, da lahko 66,4 % variabilnosti potovalnega časa pojasnimo z linearnim učinkom števila prevoženih kilometrov.

Ta ugotovitev je dokaj dobra, vendar bodo vodje morda želeli razmisliti o dodajanju druge neodvisne spremenljivke, da bi pojasnili nekatere preostale spremenljivke v odvisni spremenljivki.

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 5.11 Rezultati z eno neodvisno spremenljivko.

Pri iskanju druge neodvisne spremenljivke so menedžerji menili, da lahko k skupnemu času potovanja prispeva tudi število dostav. Podatki družbe Frigo Trucking z dodanim številom dostav so prikazani na spodnji sliki. (x_1) in število dostav (x_2) kot neodvisni spremenljivki sta prikazani na sliki 5.12. Ocenjena regresijska enačba je

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

DATA FOR FRIGO TRUCKING WITH KM TRAVELED (x_1) AND NUMBER OF DELIVERIES (x_2) AS THE INDEPENDENT VARIABLES

Driving Assignment	x_1 = kM Traveled	x_2 = Number of Deliveries	y = Travel Time (hours)
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 5.12 Podatki in neodvisne spremenljivke družbe Frigo Trucing.

Podrobneje si oglejmo vrednosti b_1 = 0,0611 in b_2 = 0,923 v zadnji enačbi.

Opomba o razlagi koeficientov

Na tej točki lahko komentiramo razmerje med ocenjeno regresijsko enačbo z neodvisno spremenljivko samo prevoženi kilometri in enačbo, ki vključuje število dostav kot drugo neodvisno spremenljivko. Vrednost b_1 v obeh primerih ni enaka. Pri preprosti linearni regresiji razlagamo b_1 kot oceno spremembe y za spremembo neodvisne spremenljivke za eno enoto. Pri večkratni regresijski analizi je treba to razlago nekoliko spremeniti. To pomeni, da se pri večkratni regresijski analizi vsak regresijski koeficient razlaga na naslednji način: predstavljal bi oceno spremembe v y ki ustreza spremembi v x_i za eno enoto, če so vse druge neodvisne spremenljivke nespremenjene.

V primeru družbe Frigo Trucking to vključuje dve neodvisni spremenljivki, b_1=0,0611 in 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

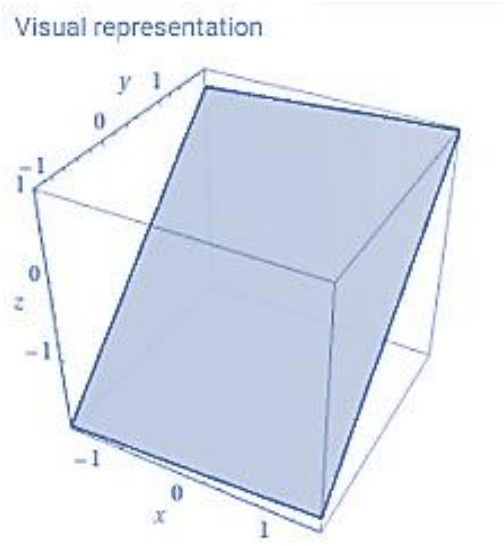
S = 0.573142   R-Sq = 90.4%   R-Sq(adj) = 87.6%

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 5.13 Rezultati za Frigo Trucking z dvema neodvisnima spremenljivkama.

Tako je 0,0611 ure ocena pričakovanega podaljšanja potovalnega časa, ki ustreza povečanju za eno miljo na prevoženo razdaljo, če je število dostav konstantno. Podobno, ker je b_2 =0,923, je ocena pričakovanega povečanja potovalnega časa, ki ustreza povečanju za eno dostavo, ko je število prevoženih milj konstantno, 0,923 ure.



Slika 5.14 Vizualni prikaz rezultatov za primer Frigo Trucking.

Literatura poglavje 5

- 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

6. Uvod v raziskovanje operacij

Operativno raziskovanje (britanska angleščina: operational research, oznaka specialnosti ameriških zračnih sil: Operacijska analiza), pogosto skrajšana na inicialko OR, je disciplina, ki se ukvarja z razvojem in uporabo analitičnih metod za izboljšanje odločanja. Kot sopomenka se občasno uporablja izraz management science.

Metode OR se uporabljajo za analizo zmogljivosti virov, ozkih grl, časov izvedbe in ciklov, vzorcev povpraševanja, zalog, porazdelitve virov, vzdrževanja, razporejanja operaterjev, mešanja izdelkov, proizvodnje izdelkov, zanesljivosti, uporabe virov, pravil in politik, učinkovitosti razporeda in razporejanja, prepustnosti procesa itd. (Ueda, 2010).

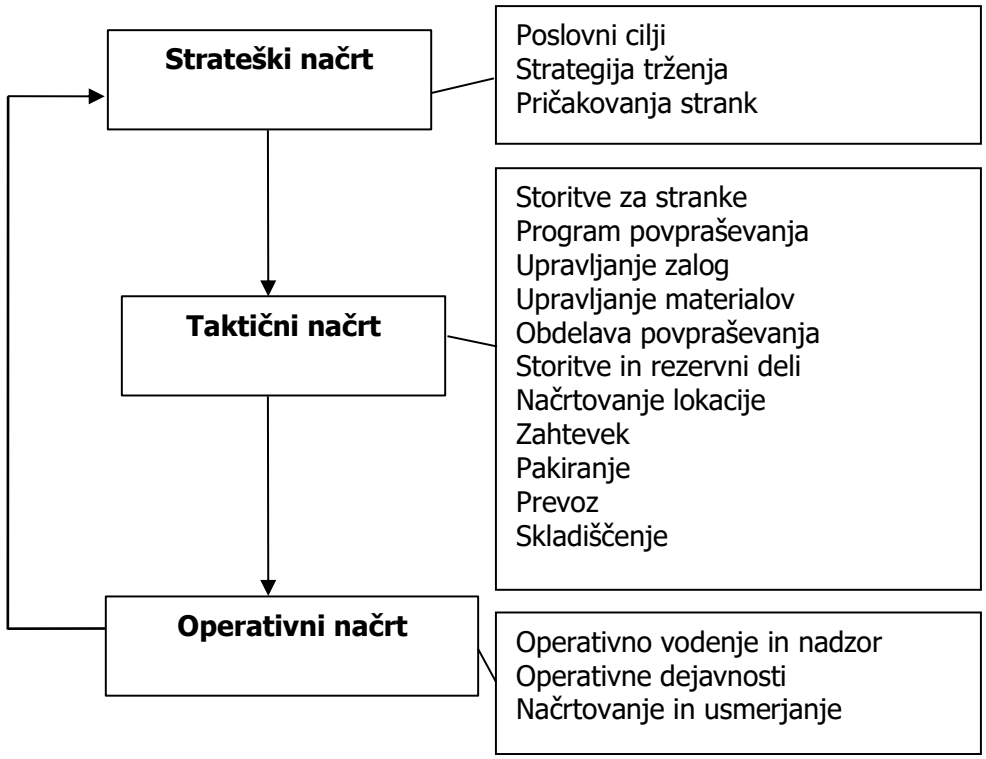
Z uporabo tehnik iz drugih matematičnih ved, kot so modeliranje, statistika in optimizacija, operacijske raziskave iščejo optimalne ali skoraj optimalne rešitve problemov odločanja. Zaradi poudarka na praktični uporabi se operacijske raziskave prekrivajo s številnimi drugimi disciplinami, zlasti z industrijskim inženiringom in logistiko, zaradi česar so postale sestavni del njihovih sistemov za upravljanje znanja (KMS).

6.1 Strateško logistično načrtovanje

Po mnenju (Robinson, 2004) se raziskovanje operacij v glavnem ukvarja z interakcijo med načrtovanimi tehničnimi in človeškimi sistemi. Zanjo so značilne spremenljivost, soodvisnost med sestavnimi deli ter strukturna in vedenjska kompleksnost. Za obvladovanje teh značilnosti je bila razvita široka paleta metod, ki jih ustrezno obravnavajo kot celoto. Uporabljajo se pri njihovem strateškem načrtovanju za:

- omogočiti zapleteno analizo "kaj, če";
- upravljanje kompleksnosti: soodvisnost + spremenljivost + dinamika;
- vključuje manj stroškov in posegov v proces kot eksperimentiranje z resničnim sistemom;
- osredotočiti se na podrobnosti;
- izboljšanje razumevanja sistema;
- izboljšanje komunikacije med vodstvom in strokovnjaki.

Strateško logistično načrtovanje (slika 6.1) zajema vse dejavnosti, ki jih je treba izvajati na strateški, taktični in operativni ravni, da bi zagotovili celovito upravljanje kakovosti (TQM) (Ciampa, 1992) in pravočasno proizvodnjo (JIT) (Britannica, 2023).



Slika 6.1 Strateško logistično načrtovanje.

6.2 Six-Sigma

Pri strateškem logističnem načrtovanju referenčni model SCOR (AIMS, 2021) podjetjem pomaga oceniti in izpopolniti upravljanje oskrbovalne verige za zanesljivost, doslednost in učinkovitost. Prepoznava šest glavnih poslovnih procesov - načrtovanje, pridobivanje, izdelava, dobava, vračanje in omogočanje.

Postopek načrtovanja SCOR zajema vse dejavnosti, povezane z razvojem načrtov za upravljanje in izboljšanje oskrbovalne verige. Stalna prizadevanja za doseganje stabilnih in predvidljivih rezultatov procesov z zmanjševanjem variabilnosti procesov (6-sigma) so ključnega pomena za poslovni uspeh.

Razlaga DMAIC in DMADV

Proizvodni procesi (pridobivanje, izdelava, dobava, omogočanje in ravnanje z vračili) imajo značilnosti, ki jih je mogoče opredeliti, izmeriti, analizirati, izboljšati in nadzorovati. Zato te faze sestavljajo metodologijo upravljanja proizvodnih procesov, skrajšano imenovano DMAIC.

Nekateri praktiki so ideje 6-sigme združili z vitko proizvodnjo in ustvarili metodologijo, imenovano Lean Six Sigma (Wheat & Mills & Carnell, 2003). Metodologija Lean Six Sigma obravnava vitko proizvodnjo ("just-in-time" proizvodnja), ki obravnava učinkovitost procesov, in 6-sigma, ki se osredotoča na zmanjševanje variacij in odpadkov, kot komplementarni disciplini, ki spodbujata poslovno in operativno odličnost.

Metodologija DMADV (definiraj, izmeri, analiziraj, oblikuj in preveri), znana tudi kot DFSS ("Design for Six Sigma"), je skladna s KBE ("Knowledge Based Engineering"). Metodologija DFSS (Chowdhury, 2002) ima naslednje faze:

1. Opredelite cilje oblikovanja, ki so skladni z zahtevami strank in strategijo podjetja.
2. Izmerite in prepoznajte značilnosti, ki so ključne za kakovost (CTQ), izmerite zmogljivosti izdelka, zmogljivost proizvodnega procesa in izmerite tveganja.
3. Analizirajte, da razvijete in oblikujete alternativne možnosti.
4. Oblikujte izboljšano alternativo, ki je najprimernejša glede na analizo v prejšnjem koraku.
5. Preverite zasnovo, pripravite poskusno izvedbo, izvedite proizvodni proces in ga predajte lastniku(-kom) procesa.

Projekti za izboljšanje poslovanja Six Sigma (Tennant, 2001), ki jih je navdihnil cikel W. Edwardsa Deminga "Načrtuj - Izvedi - Preučuj - Deluj" (Tague, 2005), glede na svojo naravo sledijo eni od zgoraj omenjenih metodologij, od katerih ima vsaka pet faz:

1. DMAIC se uporablja za projekte, katerih cilj je izboljšati obstoječi poslovni proces.
2. DMADV se uporablja pri projektih, namenjenih ustvarjanju novih izdelkov ali načrtovanju procesov.

6.3 Poslovno obveščanje

Poslovna inteligenca (BI) zajema vse strategije in tehnologije, ki jih podjetja uporabljajo za podatkovno analizo preteklih in trenutnih poslovnih informacij (Tableau, 2023). Podpirajo jo sistemi za upravljanje znanja (KMS), ki predstavljajo del logističnih informacijskih sistemov (LIS), ki strokovnjakom z različnih področij omogočajo svetovanje in zagotavljanje podpore različnim ravnem upravljanja.

Poslovna analitika

Poslovna analitika (BA) je proces, ki temelji na BI in omogoča nove vpoglede v poslovne procese ter boljše strateško odločanje za prihodnost. Izvira iz podatkovnega rudarjenja (DM), ki je proces iskanja anomalij, vzorcev in korelacij v večjih podatkovnih nizih za napovedovanje rezultatov.

Postopek BA je sestavljen iz:

1. Združevanje podatkov: pred analizo je treba podatke najprej zbrati, urediti in filtrirati, in sicer s prostovoljnimi podatki ali transakcijskimi zapisi.
2. Podatkovno rudarjenje: podatkovno rudarjenje z uporabo podatkovnih zbirk, statistike in strojnega učenja razvršča velike podatkovne nize, da bi prepoznalo trende in ugotovilo povezave.
3. Identifikacija povezav in zaporedij: identifikacija predvidljivih dejanj, ki se izvajajo v povezavi z drugimi dejanji ali zaporedno.
4. Rudarjenje besedila: raziskuje in organizira velike nabore nestrukturiranih besedilnih podatkov za namene kvalitativne in kvantitativne analize.
5. Napovedovanje: analizira zgodovinske podatke iz določenega obdobja, da bi pripravila utemeljene ocene, ki so napovedne pri določanju prihodnjih dogodkov ali obnašanja.
6. Napovedna analitika: napovedna poslovna analitika uporablja različne statistične tehnike za ustvarjanje napovednih modelov, ki pridobivajo informacije iz zbirk podatkov, prepoznavajo vzorce in zagotavljajo napovedne rezultate za vrsto organizacijskih rezultatov.
7. Optimizacija: ko so trendi opredeljeni in napovedi izdelane, lahko podjetja uporabijo simulacijske tehnike za testiranje najboljših možnih scenarijev.

8. Vizualizacija podatkov: omogoča vizualne predstavitve, kot so grafikoni in diagrami, za enostavno in hitro analizo podatkov.

Načrtovanje prodaje in poslovanja

Načrtovanje prodaje in poslovanja (SOP) je prilagodljivo orodje za napovedovanje in načrtovanje proizvodnih dejavnosti. Koraki SOP:

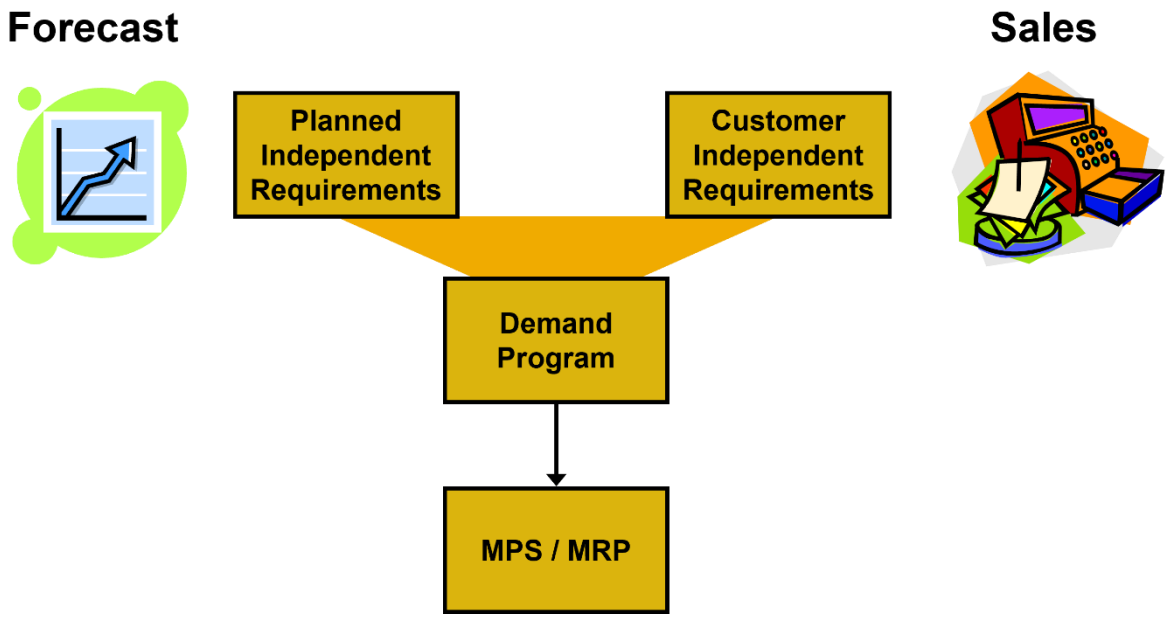
1. načrt prodaje,
2. načrt proizvodnje in
3. načrtovanje zmogljivosti.

SOP deluje na podlagi podatkov iz različnih virov informacij v podjetju: V podjetju se uporabljajo različni viri informacij iz različnih služb: prodaje, trženja, proizvodnje, računovodstva, kadrovske službe in naročilnic. Običajno jih zagotovijo ustrezni oddelki prek sistema za načrtovanje virov podjetja (ERP).

Medtem ko SOP deluje na strateški ravni, ERP deluje na taktični ravni logističnega informacijskega sistema podjetja. Združuje ju program za upravljanje povpraševanja, ki povezuje strateško načrtovanje prodaje in poslovanja (SOP) ter podrobno načrtovanje proizvodnje (glavno načrtovanje proizvodnje / načrtovanje materialnih potreb) na operativni ravni. Pri tem sta v ospredju prej omenjeno načrtovanje in simulacija, ki omogočata izdelavo izvedljivega in optimalnega proizvodnega načrta.

Program upravljanja povpraševanja (slika 6.2) je sestavljen iz dveh vrst napovedi:

1. načrtovane neodvisne potrebe (PIR) iz predvidenega obsega prodaje na podlagi trženja in
2. neodvisne zahteve strank (CIR) iz podatkov na podlagi obstoječih in načrtovanih prodajnih naročil.



Slika 6.2 Upravljanje povpraševanja.

Primer SOP

Podjetje v svoji prodajni mreži trguje z več vrstami izdelkov. Prodajne transakcije se shranjujejo centralno, da je mogoče spremljati stanje zalog in izvajati analizo prodaje za upravljanje povpraševanja. Beležijo se v obliki CSV (Comma Separated Values - vrednosti, ločene z vejico), ki jo je enostavno obdelati v sistemu ERP podjetja in analitičnem oddelku, ki uporablja preglednice.

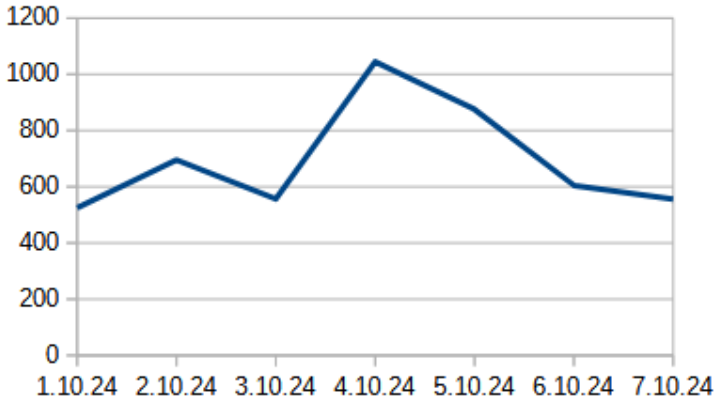
Pri analizi prodaje se prodajne transakcije najprej filtrirajo, da se ugotovi, ali so popolne in pravilno oblikovane. Šele nato so pripravljene za statistično oceno, saj bi morebitni manjkajoči ali slabo oblikovani podatki lahko povzročili napačno razlago rezultatov.

Date	Seller ID	Customer ID	Transaction ID	Product ID	Product Price
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 6.3 Podatki o tedenski prodaji.

Običajno se z analitiko omogočijo različni smiselni vpogledi v zbrane "surove podatke" (slika 6.3). To lahko dosežemo z vrtilnimi tabelami, ki omogočajo združevanje podatkov po izbranih atributih in izvajanje statističnih analiz. Načeloma lahko za vrtilno točko štejemo vsak atribut (stolpec) vhodnih podatkov. Zato pogosto govorimo o "podatkovni kocki" z več dimenzijami. Ker ne moremo grafično prikazati več kot dveh ali treh dimenzij, so najpreprostejše, vendar običajno najbolj uporabne predstavitve vhodnih podatkov sestavljene iz dveh ali treh atributov pivotov. Na podlagi danega vzorca podatkov je v nadaljevanju podanih nekaj primerov vrtilnih tabel.

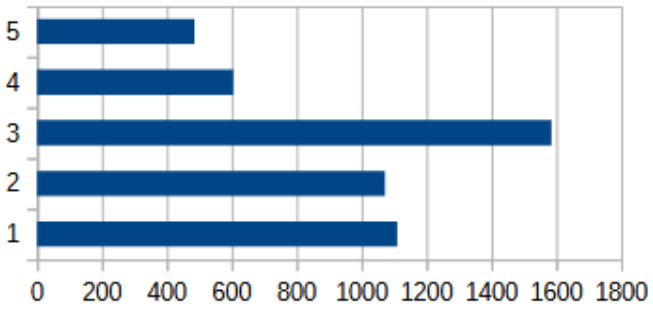
Date	Vsota - Product Price
1.10.24	525,00 €
2.10.24	695,00 €
3.10.24	556,00 €
4.10.24	1.045,00 €
5.10.24	875,00 €
6.10.24	605,00 €
7.10.24	556,00 €
Skupaj Rezul	**4.857,00 €**



Slika 6.4 Statistični podatki o prodaji po dnevih v tednu.

Statistični podatki o prodaji po delovnih dneh ali mesecih (slika 6.4) omogočajo vpogled v sezonske trende.

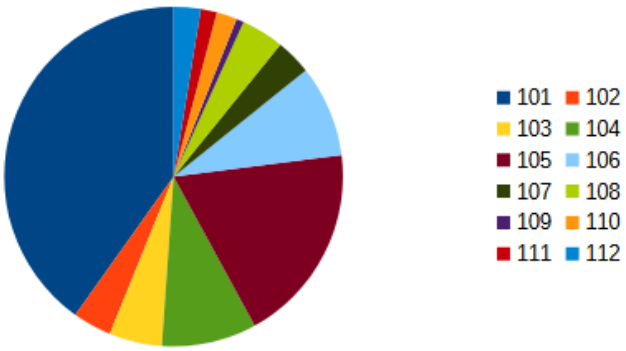
Seller ID	Vsota - Product Price
1	1.110,00 €
2	1.072,00 €
3	1.585,00 €
4	605,00 €
5	485,00 €
Skupaj Rezul	**4.857,00 €**



Slika 6.5 Statistični podatki o prodaji po prodajnih uradih.

Statistični podatki o prodaji po prodajnih uradih (slika 6.5) določajo, kateri uradi so najbolj obremenjeni in/ali ustvarijo največ prihodkov.

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 6.6 Statistični podatki o prodaji po izdelkih.

Statistični podatki o prodaji po izdelkih (slika 6.6) določajo izdelke, ki so najbolj iskani ali predstavljajo pomemben delež v portfelju.

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

Slika 6.7 Pregled transakcij.

Pregled transakcij (slika 6.7) po dnevih, prodajnih uradih, transakcijah in kupcih omogoča strukturiran vpogled v podatke, kar je koristno pri reševanju poizvedb o določenem izdelku ali prodajni transakciji.

6.4 Sistemi za podporo odločanju

Sistem za podporo odločanju (DSS) je interaktivni sistem, ki odločevalcem pomaga pri uporabi podatkov in modelov za reševanje nestrukturiranih ali delno strukturiranih problemov. Ekspertni sistem (ES) je aplikativni program ali okolje, ki učinkovito podpira reševanje problemov na specializiranem problemskem področju, ki zahteva strokovno znanje in veščine.

Poleg statističnih in simulacijskih metod, ki se uporabljajo v BA, DSS pogosto vključuje modele, ki omogočajo sprejemanje odločitev na podlagi niza posebnih meril. Ti modeli so lahko preprosti (npr. odločitvene tabele, drevesa itd.), ki vodijo do ene same pravilne rešitve. Po drugi strani pa je pri obravnavi več (po možnosti nasprotujočih si) meril in več rešitev potreben bolj zapleten model.

Ustrezen model, ki se pogosto uporablja v poklicnem in zasebnem življenju, je model odločanja po več merilih.

Sprejemanje odločitev na podlagi več meril

Večkriterijsko odločanje (MCDM) ali večkriterijska analiza odločanja (MCDA) je pod-disciplina OR, ki pri odločanju izrecno ocenjuje več (lahko tudi nasprotujočih si) meril nad nizom kandidatnih rešitev ali variant.

Odločitveni model MCDM je sestavljen iz:

- Kriterijev - parametri vhodnih variant, ki so ključni za našo zasnovo.
- Uteži - relativna pomembnost izbranih meril.
- Funkcija uporabnosti - funkcija, ki združuje ponderirane parametre različic v vrednost primernosti.
- Podatki - podatki, ki predstavljajo variante; vhodni podatki za naš model MCDM.

Podatkovne vrste v MCDM so lahko:

- Kvantitativni - predstavljajo vrednosti, ki jih je mogoče primerjati kot take.
- Kvalitativni - predstavljajo relativne primerjalne vrednosti (npr. visoka, udobna, nizka temperatura itd.), ki jih je treba količinsko opredeliti, da bi dobili edinstvene vrednosti.
- Binarni - predstavljajo binarna merila; lastnost je izpolnjena (1) ali ne (0).

Postopek večkriterijskega odločanja:

1. Predstavitev variant (V) z njihovimi značilnimi parametri (P): $\{V_i\ (P_{i,1}\ ;\ P_{i,2}\ ;\ \dots\ P_{i,n}\ );\ i=1\dots m\}$.
2. Normalizacija parametrov z izračunom relativne lokalne ocene $p_{i,j}$ za vsak $P_{i,j}$ (j=1...n) glede na največjo vrednost parametra j^{th} $P_{i,j}$ iz vseh i vzorcev:
   a) $p_{i,j} = P_{i,j} / \max\{P_{i,j}\}$, če je večja vrednost $P_{i,j}$ ugodnejša.
   b) $p_{i,j} = 1 - P_{i,j} / \max\{P_{i,j}\}$, če je manjša vrednost Pi,j ugodnejša.
3. Ocene so ponderirane glede na preference: $x_{i,j} = p_{i,j} * U_j$ za vsak j=1...n, s ponderji U_j , katerih vsota mora biti 1, tj. 100 %.
4. Ponderirane ocene vseh različic se seštejejo: $X_i = \sum x_{i,j}$ za vsak i=1...m, da dobimo sestavljene ocene v skladu z našo funkcijo koristnosti.

5. Izbrana je najboljša varianta $Y=\max\{X\}_i$

Primer MCDM

Pri izbiri nove opreme v podjetju se moramo pogosto odločati po več kriterijih. Oglejmo si primer izbire stroškovno najugodnejše (pod 300 EUR) mobilne platforme z operacijskim sistemom Android za naše podjetje. V preglednici 6.1 so navedeni primerki, ki so bili izbrani na podlagi poizvedbe med zaposlenimi, da bi zožili našo ponudbo. Za vsakega od njih so navedeni parametri, ki so bili izbrani kot najpomembnejši. V nadaljevanju so podatki o izbranih parametrih normalizirani, da dobimo primerljive vrednosti, ponderirani, da poudarimo vrednosti, ki so pomembnejše, in sešteti, da dobimo ocene za izbrane primerke.

Preglednica 6.1 MCDM za stroškovno učinkovito mobilno platformo Android.

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 %

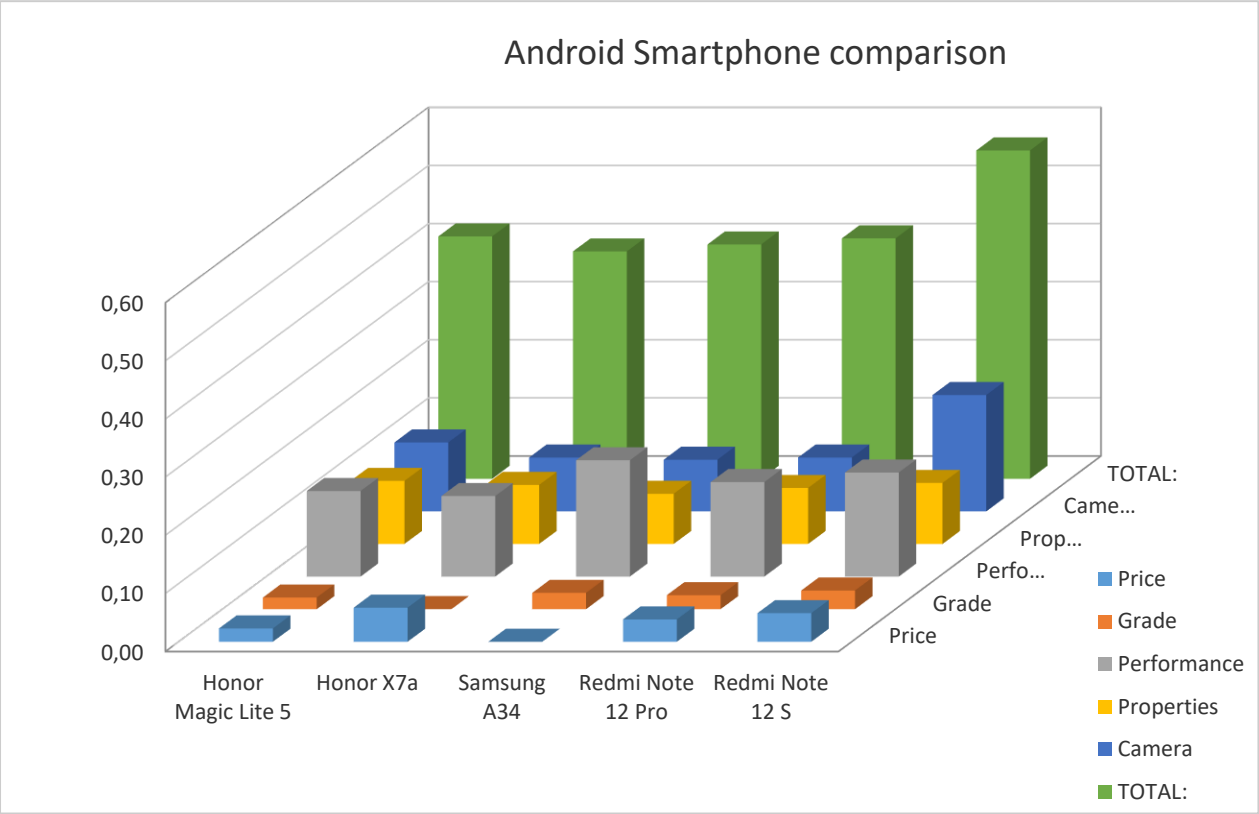
NORMALIZED PARAMETERS									
	Price (€)	Grade*	Performance			Properties			Camera
Model		(1-10)	proc.speed (GHz)	RAM (GB)	int.mem. (GB)	weight (g)	size (mm3)	bat.cap. (mAh)	(MP)
Honor Magic Lite 5	0,11	0,20	0,85	0,75	0,50	0,12	0,11	0,85	0,59
Honor X7a	0,29	0,00	0,88	0,50	0,50	0,02	0,00	1,00	0,46
Samsung 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

Končni rezultat naše analize je zbirna preglednica (preglednica 6.2) in po možnosti graf (slika 6.8), ki povzema postopek odločanja in predstavlja najboljšo izbiro ter prednosti in slabosti posameznih variant. Pogosto se zgodi, da vzorec z najboljšo oceno ni tisti, ki se je najbolje odrezal v vseh kategorijah, temveč tisti, ki v povprečju najbolj ustreza našim merilom za izbor ter tehtanju parametrov. To je tudi glavna prednost metode MCDM, saj nas lahko izbira po posameznem parametru zavede.

Preglednica 6.2 Povzetek MCDM za naš primer izbire mobilne platforme.

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



Slika 6.8 Izbira najboljše mobilne platforme.

Najboljša izbira: 0,56 (Redmi Note 12 S).

6.5 Inženiring, ki temelji na znanju

Na znanju temelječi inženiring (KBE) je inženirska metodologija za sistematično vključevanje inženirskega znanja v sistem načrtovanja (Andersson, Larsson in Ola, 2011).

Življenjski cikel upravljanja izkušenj

Potreba po zajemanju, upravljanju in uporabi znanja o oblikovanju ter avtomatizaciji procesov, ki so edinstveni za izkušnje proizvajalca pri razvoju izdelkov, je privedla do razvoja tehnologije na znanju temelječega inženiringa (KBE) (Prasad, 2005). Namen KBE je obogatiti institucionalno znanje z upravljanjem izkušenj. Faze upravljanja izkušenj so po (Andersson & Larsson & Ola, 2011) naslednje:

1. Prepoznavanje: izbrana je neskladnost z želenim stanjem, ki se pojavi v proizvodnem procesu zaradi slabo opredeljenega izdelka ali procesa.
2. Zajem: izkušnja s svojimi lastnostmi je zajeta.
3. Analiza: opravi se analiza temeljnih vzrokov zajetih izkušenj, da se določi ustrezna strategija za odpravo nepravilnosti in njena ponovna uporaba za preprečevanje ponavljajočih se nepravilnosti.
4. Shranjevanje: vpogledi iz analize so arhivirani skupaj z izkušnjo.
5. Iskanje in priklic: izkušnja se poišče in prikliče.
6. Uporaba: uporabljen je element izkušnje.
7. Ponovna uporaba: zaključi cikel upravljanja znanja in začne novega.

Na znanju temelječi inženiring na splošno dopolnjujejo še druge discipline, katerih podrobnejša obravnava presega obseg tega poglavja:

- Računalniško podprto upravljanje projektov (PS).
- Računalniško podprto načrtovanje (CAD), proizvodnja (CAM) in robotika (CIM).
- Računalniško simulacijsko modeliranje in analiza (SMA).
- Računalniško podprto podrobno načrtovanje proizvodnje (MPS/MRP).

6.6 Zaključek

Kot je predstavljeno v tem poglavju, so glavne aplikacije BI v upravljanju podjetij povezane s poslovno analitiko (BA) in sistemi za podporo odločanju (DSS). Običajno jih imenujemo operacijske raziskave (OR). Poleg osnovnih tehnik BI, opisanih v tem poglavju, je v poglavjih 3 in 4 o upravljanju podatkov ter simulacijskem modeliranju in analizi (SMA) podanih nekaj dodatnih premislekov o zbiranju, obdelavi in predstavitvi podatkov, ki prav tako podpirajo odločanje. Če povzamemo, aplikacije BI najdemo v sistemih za upravljanje znanja (KMS), ki vključujejo:

- Sistemi za podporo odločanju (DSS),
- poslovna analitika (BA) kot nadgradnja podatkovnega rudarjenja (DM) in
- na znanju temelječe inženirstvo (KBE) kot nadgradnja računalniško podprtega inženirstva (CAE).

Na podlagi rezultatov BA, DSS in SMA se oblikujejo izkušnje, ki krepijo institucionalno znanje in sestavljajo njihove na znanju temelječe strokovne sisteme (KBS). Kot je prikazano v (Gumzej et. al., 2023), jih lahko KBE uporabi za uvedbo načela "pridobljenih izkušenj" v upravljanje izboljšav v podjetju s strateškim logističnim načrtovanjem.

Literatura poglavje 6

- 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. Statistična obdelava podatkov SPSS

Do zdaj ste že pridobili temeljno razumevanje statistike, ravnanja s podatki, vzpostavitve simulacije, modeliranja in analize v logističnih oskrbovalnih verigah ter preprostih metod linearne regresije. Čeprav statistika ponuja raznoliko paleto modelov in tehnik za izboljšanje vaših optimizacijskih

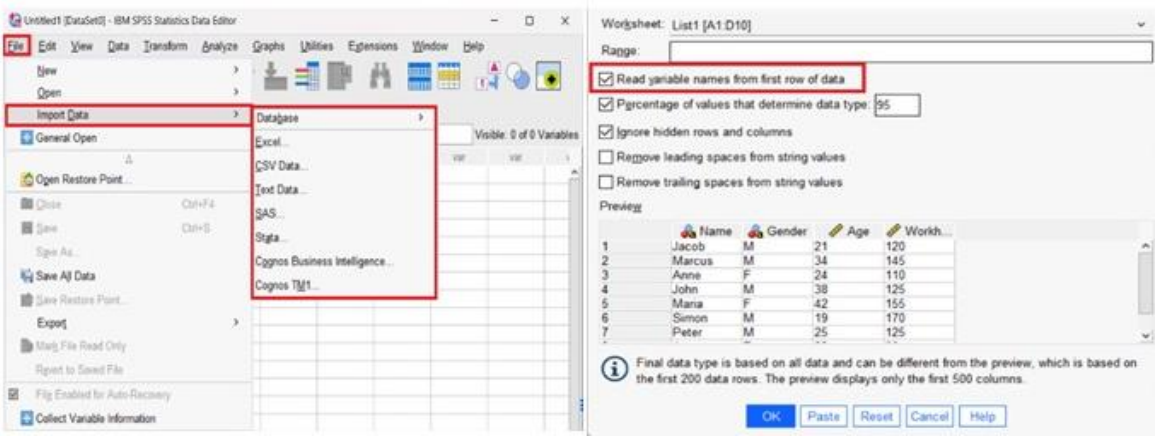
prizadevanj, izvajanje analiz in ugotavljanje morebitnih izboljšav, ste morda opazili, da lahko z naraščanjem kompleksnosti analiziranih podatkov in izračunov tradicionalni pristopi postanejo vse bolj zapleteni in zahtevni za izračunavanje. S stopnjevanjem zapletenosti podatkov in izračunov lahko običajne metode postanejo zastarele in v nekaterih primerih ogrozijo zanesljivost vaših rezultatov. Za reševanje tega problema statistika uporablja različne računalniške programe, ki avtomatizirajo analizo in razlago zbranih podatkov, hkrati pa zagotavljajo številne modele in funkcije za zagotavljanje zanesljivih rezultatov. Eden od takih programov je IBM-ov SPSS, ki bo ključno orodje v tem poglavju. V tem poglavju bomo na kratko predstavili osnovno uporabo programske opreme SPSS ter raziskali njene funkcionalnosti in praktično uporabo. Začetnemu uvodu bo sledila praktična uporaba programa s štirimi temeljnimi testi za izračun rezultatov: T-test, korelacije, Chi-kvadrat in ANOVA. Za lažje učenje bomo predstavili preproste probleme in njihove rešitve, ki vam bodo pomagali pri spoznavanju teh testov.

7.1 Osnove IBM-ovega programa SPSS

Morda ste že imeli nekaj izkušenj s programsko opremo SPSS. Če pa jo še vedno potrebujete, naj vam ponudimo kratko predstavitev. SPSS, podobno kot njegov bolj znan ekvivalent Excel, omogoča manipulacijo, analizo in vizualizacijo podatkov. Kljub temu pa za razliko od programa Excel, ki je lahko včasih naporen in zapleten pri programiranju funkcij, SPSS ponuja uporabniku prijazen vmesnik za statistično analizo (IBM, 2021). SPSS ponuja različne funkcije in metodologije za učinkovito obdelavo zagotovljenih podatkov. Čeprav se programska oprema SPSS odlikuje z zagotavljanjem obsežnih zmožnosti statistične analize, je navigacija pri manipulaciji s podatki in konfiguracija začetnih nastavitev za analizo včasih lahko zahtevna (IBM, 2021). Zato se bomo poglobili v osnove uvoza podatkov in priprave podatkov za poznejše statistične teste.

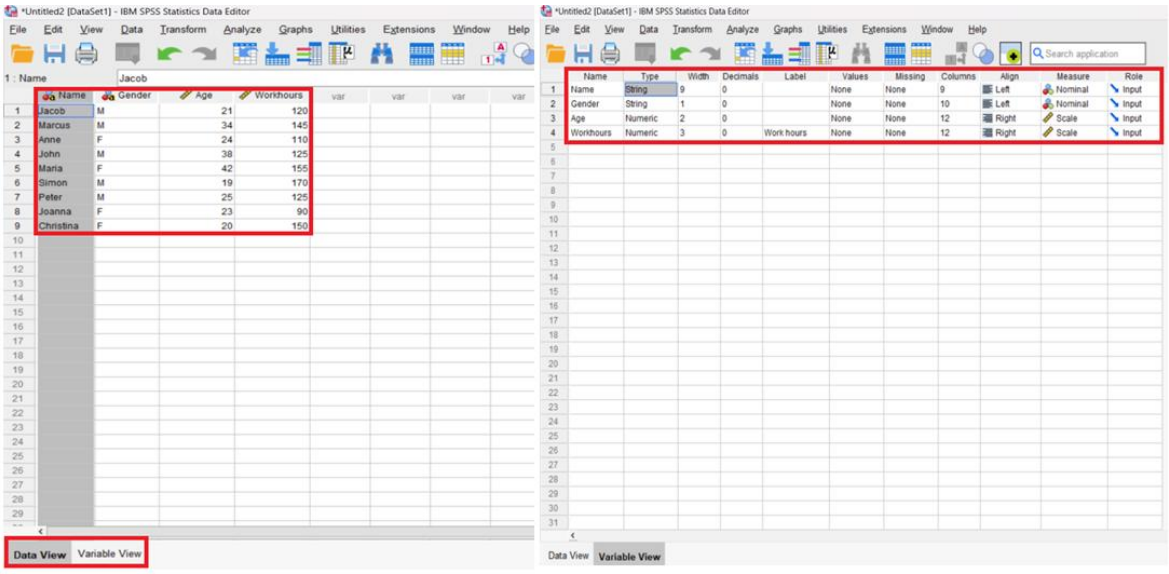
Glede na razširjeno uporabo programa Excel za obdelavo številčnih podatkov so lahko vaši podatki pridobljeni ali pripravljeni v Excelovi preglednici. Na srečo lahko SPSS v svoje preglednice uvozi podatke iz različnih oblik datotek. Ko imate pripravljene končne preglednice Excel, odprite

programsko opremo SPSS. Na začetnem zaslonu pojdite na zavihek "File" (Datoteka) in izberite "Import Data" (Uvozi podatke). V naslednjem oknu lahko izberete obliko podatkov, ki jo nameravate uvoziti (glejte sliko 7.1). Po tem koraku poiščite svojo pripravljeno datoteko, jo izberite in nadaljujte z naslednjim oknom. V tem oknu boste morali konfigurirati dodatne nastavitve. Če ste imena stolpcev že vključili v prvo vrstico podatkov, izberite možnost "Preberi imena spremenljivk iz prve vrstice podatkov" (glejte sliko 7.1) in nato kliknite "Dokončaj", da se podatki prikažejo v preglednici (IBM, 2021).



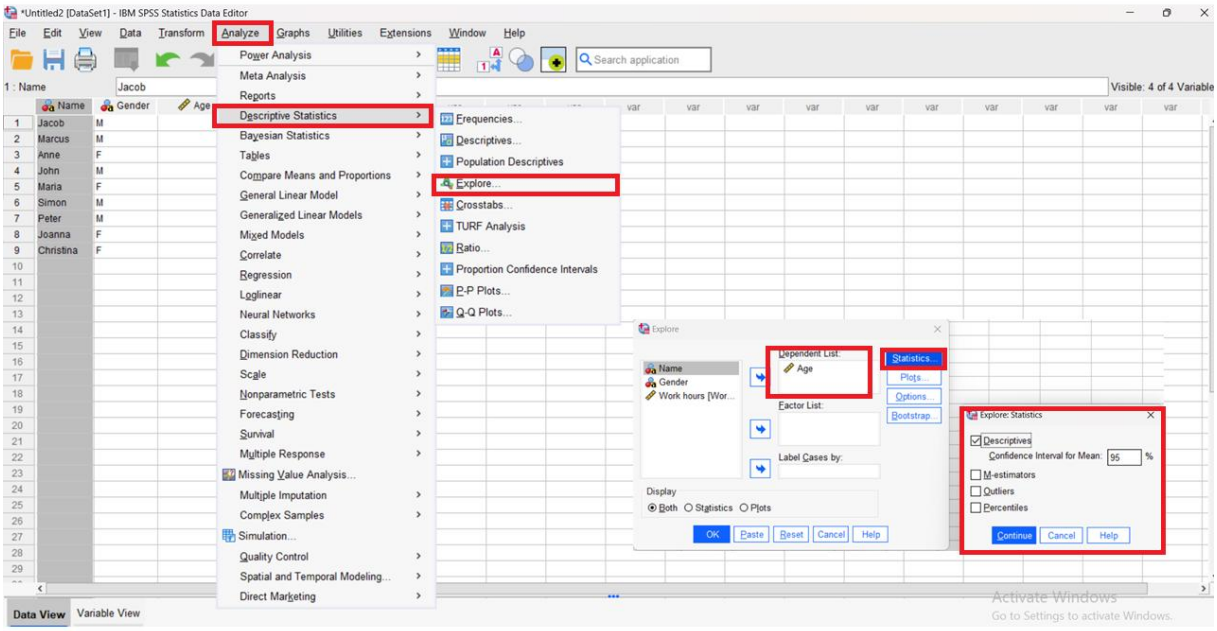
Slika 7.1 Nastavitve uvoza SPSS.

Zdaj, ko imamo podatke v preglednici, boste opazili jasno razliko v predstavitvi v primerjavi z Excelom. SPSS razvršča podatke v dve osnovni vrsti, vsaka pa ima še dve dodatni podvrsti. Kot je prikazano na sliki 7.2, lahko podatke razvrstimo kot numerične ali kategorične. Numerični podatki so sestavljeni iz števil in jih lahko razvrstimo kot diskretne (s končnimi možnostmi) ali zvezne (ponujajo neskončno možnosti). Po drugi strani so kategorični podatki sestavljeni iz besed in se lahko nadalje ločijo na ordinalne (s hierarhijo) ali nominalne (brez hierarhije). Glede na naravo vaših podatkov boste morda morali nastaviti spremenljivke tako, da bodo ustrezale želeni analizi. V večini primerov SPSS samodejno ustrezno razvrsti spremenljivke. Predpostavimo, da želite dodatno manipulirati s podatkovnimi vrstami. V tem primeru lahko dostopate do možnosti "Pogled" in v razdelku "Pogled spremenljivke" prilagodite informacije o spremenljivkah, kot so ime, vrsta, širina, mera in drugo (IBM, 2021).



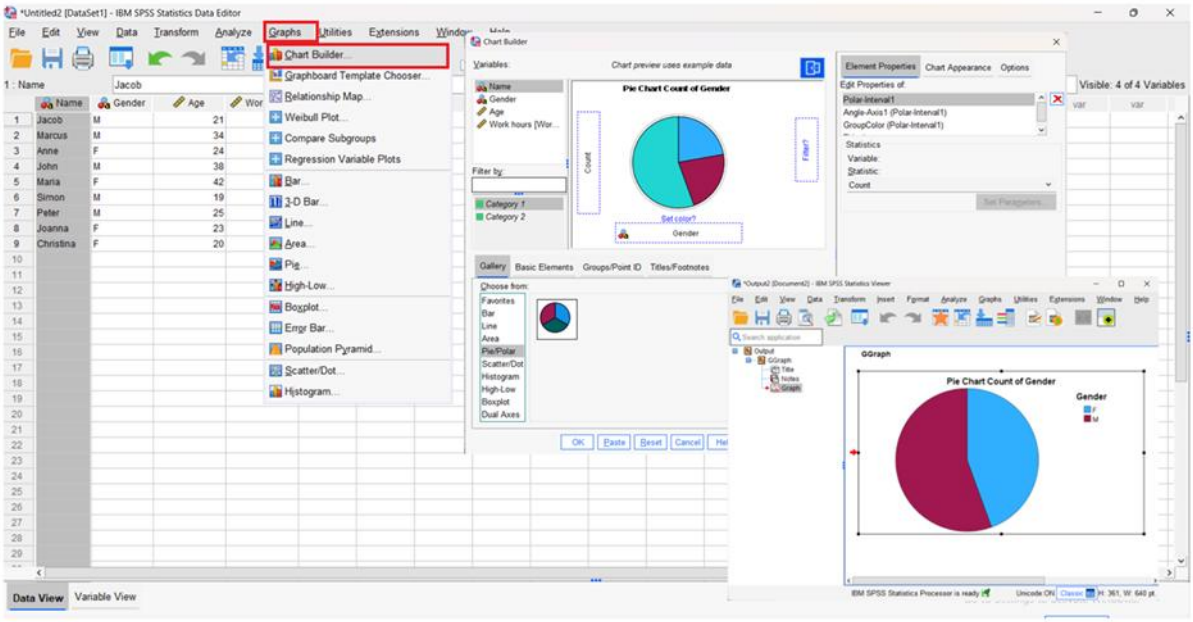
Slika 7.2 Okna za prikaz podatkov in spremenljivk.

Ko pravilno nastavite podatke, jih lahko raziskujete v programu SPSS. SPSS uporabnikom omogoča izvajanje temeljnih statističnih analiz, ne da bi se pri tem zanašali na vnaprej določene funkcije. Na začetnem zaslonu (glejte sliko 7.3) pojdite na "Analyze" (Analiza), nato na "Descriptive Statistics" (Opisna statistika) in izberite "Explore" (Raziskovanje). V razdelku "Explore" (Raziskovanje) boste našli različne možnosti, ki so odvisne od značilnosti predloženih podatkov. V tem načinu vam bo SPSS zagotovil informacije "opisne statistike" o vaših podatkih. To je sicer dragoceno za začetno analizo podatkov, vendar ponuja le temeljni vpogled in se ne spušča v podrobnejšo statistično analizo, ki bo obravnavana v naslednjih poglavjih. Preden nadaljujemo, bomo raziskali še eno funkcijo v SPSS - vizualizacijo grafov (IBM, 2021).



Slika 7.3 Nastavitve opisne statistike.

SPSS ponuja številne možnosti vizualizacije podatkov, vključno s histogrami, škatlastimi diagrami, stolpčnimi diagrami, diagrami razpršitve, linijskimi diagrami, krožnimi diagrami in drugimi. Do te točke bi morali imeti osnovno znanje o tem, kaj predstavljajo posamezne vrste grafov in kako razlagati rezultate, ki jih zagotavljajo. Zato se bomo osredotočili na to, kako ustvariti te grafe v programu SPSS. Za ustvarjanje grafov na začetnem zaslonu izberite zavihek "Graphs" (Grafi) in nato "Chart Builder" (Graditelj grafov). V novem oknu lahko izberete vrsto grafa, ki ga želite ustvariti, in izberete spremenljivke, ki jih želite vključiti. Po izbiri možnosti "Finish" (Končaj) se bo prikazalo novo okno z rezultati, vizualiziranimi v izbrani obliki grafa. V tem novem oknu lahko aktivno sodelujete z grafom, kar vam omogoča spreminjanje barv in pisav spremenljivk, raziskovanje porazdelitve spremenljivk po grafu in drugo (IBM, 2021).



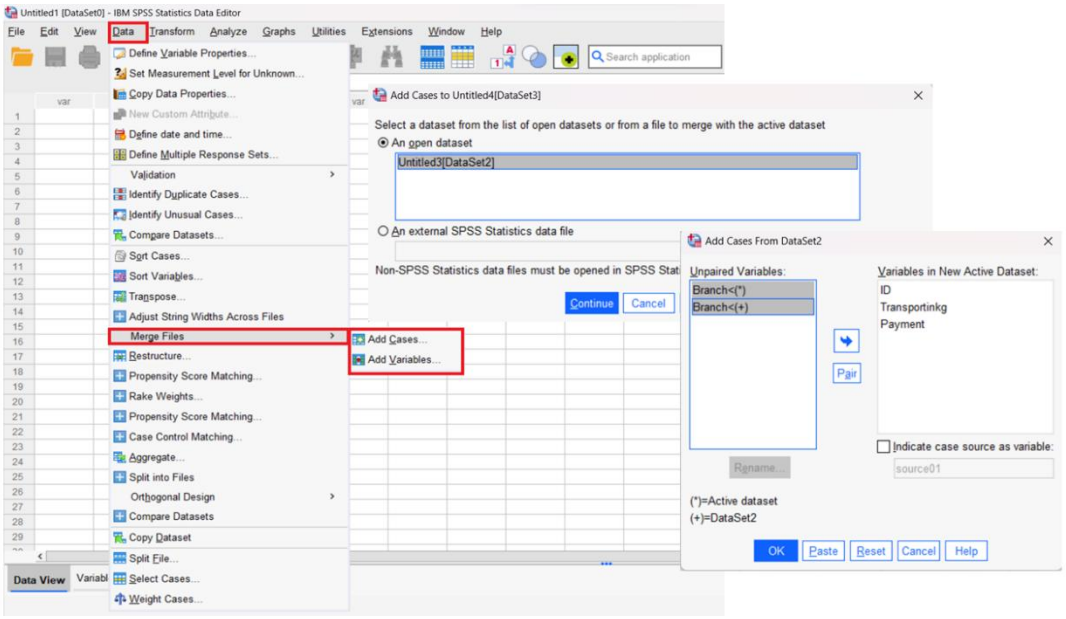
Slika 7.4 Nastavitve programa Chart Builder v SPSS.

Do te točke smo obravnavali tri od štirih pravil za "raziskovanje podatkov", ki vključujejo pregledovanje podatkov (raziskovanje surovih podatkov), prepoznavanje podatkov (določanje vrst podatkov) ter do neke mere grafično prikazovanje in opisovanje podatkov z opisno statistiko in izdelavo grafov. Zadnje pravilo je "oblikovanje vprašanj", pri katerem se vprašamo, kaj želimo doseči z analizo podatkov, ter ustrezno nastavimo grafe in opisno statistiko, da dobimo odgovore na naša specifična vprašanja. V našem trenutnem primeru bi se lahko na primer vprašali: "Ali v naši analizirani populaciji prevladujejo ženske?" Z uporabo grafov in opisne statistike lahko ugotovimo, da našo populacijo sestavljajo predvsem moški posamezniki. Pri oblikovanju vprašanj vedno upoštevajte razpoložljive podatke in opredeljene spremenljivke (Garth, 2008). S tem smo zaključili prvi del analize podatkov v programu SPSS, zdaj pa bomo nadaljevali pripravo testa.

7.2 Upravljanje podatkov

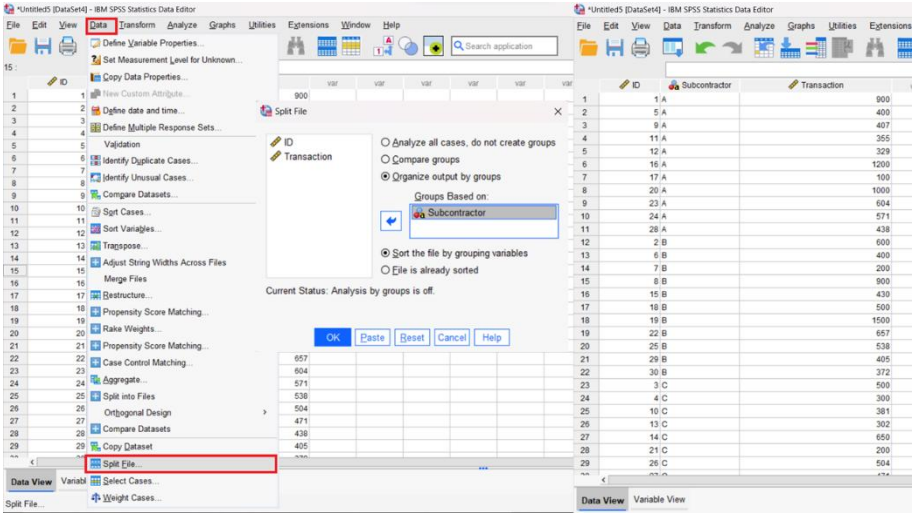
Pri delu s ključnimi podatki v programski opremi SPSS je ključno razumeti tehnike za manipulacijo informacij v posameznih aktivnih podatkovnih nizih. Program SPSS zagotavlja funkcije, ki omogočajo manipulacijo obstoječih podatkov, vsebovanih v aktivnih naborih podatkov. Občasno se lahko zgodi, da sta dve podatkovni zbirki ločeno uvoženi v podatkovne nize, vendar ju je treba zaradi boljše analize združiti. Razmislite o logističnem podjetju z dvema podružnicama, od katerih vsaka prispeva podatke o stroških in prevozu tovora v kilogramih. Vodstveni cilj je analizirati splošno

učinkovitost podjetja. V programu SPSS to dosežemo tako, da se pomaknemo na "Podatki", izberemo "Združiti datoteke" in imamo na voljo dve različni možnosti. Ena vključuje izbiro "Cases" (Primeri) in določitev spremenljivke za združevanje ter odstranitev te spremenljivke ob združevanju drugih. Druga možnost je izbira "Variable" (Spremenljivka), pri kateri se spremenljivka ohrani v novem podatkovnem nizu. Praktična uporaba je očitna v našem logističnem scenariju, kjer združevanje podatkovnih nizov poenostavi celovito analizo uspešnosti podjetja.



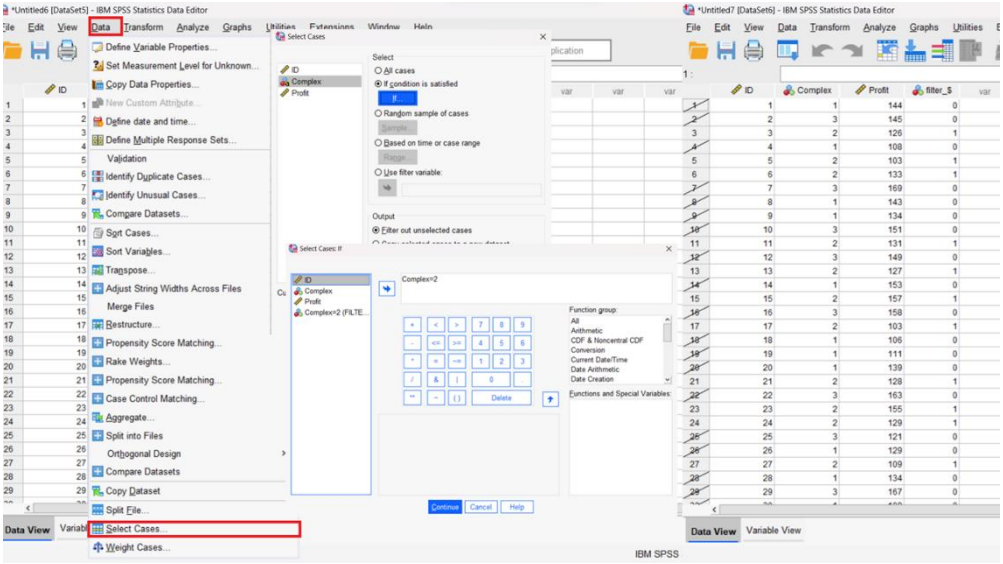
Slika 7.5 Okno za združevanje datotek.

Funkciji združevanja in delitve omogočata posebno ravnanje s podatki, vendar ima možnost "Izberi primere" posebne prednosti. Predstavljajte si, da imate podatke za trgovine B, C in D v eni podatkovni zbirki in se osredotočate izključno na primerjavo trgovine A in trgovine C. Z izbiro "Podatki" in "Izberi primere" lahko določite spremenljivke, ki vas zanimajo, in tako učinkovito filtrirate neželene podatke. Če na primer določite trgovino C kot 2, se programska oprema osredotoči izključno na trgovino C, pri čemer se ustvarijo rezultati, ki so na voljo za nadaljnje analize, kot je opisna statistika, ki se osredotoča izključno na izbrane primere. Takšen pristop omogoča tudi primerjalno analizo samo med vrednostmi v trgovini A in trgovini C.



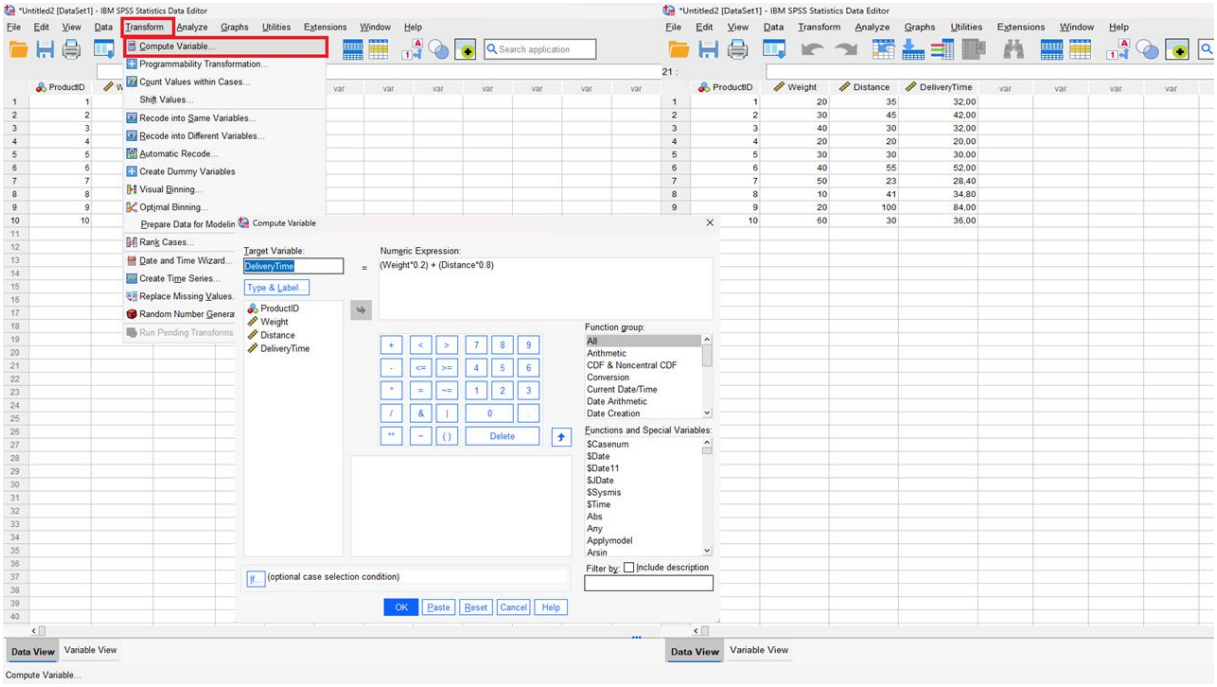
Slika 7.6 Okno za delitev datoteke.

Funkciji združevanja in delitve omogočata določeno ravnanje s podatki, na voljo pa je tudi možnost "Izberite primere". Predstavljajte si, da zagotovo vemo, da ima trgovina A v povprečju 120 EUR dobička, in to želimo primerjati s trgovino C. Žal imamo v naši zbirki podatkov podatke za trgovine B, C in D v eni zbirki podatkov in analiza bi vključevala podatke vseh treh trgovin. S klikom na "Podatki" in "Izberi primere" lahko izberemo, na katero spremenljivko se želimo osredotočiti. V naših primerih smo določili, da mora biti trgovina C nastavljena kot 2, in nato ustvarili funkcijo za programsko opremo, da se osredotoči samo na trgovino C. Rezultat lahko nato z izbiro tega novega stolpca (npr. opisna statistika) uporabimo za poznejšo analizo.



Slika 7.7 Izbira postopka za primer.

Včasih lahko podatkovni nizi že vsebujejo spremenljivke, vendar je treba uvesti nove spremenljivke na podlagi obstoječih. Vzemimo na primer vodjo logističnega podjetja, ki ima podatke o teži in prevoženi razdalji različnih izdelkov, vendar za optimizacijo poti potrebuje čas dostave. V programu SPSS to dosežemo tako, da kliknemo "Transform" in nato "Compute Variables". V novem oknu se z nastavitvijo številskih izrazov ustvari nova spremenljivka, DeliveryTime. V tem primeru se z dodelitvijo lestvice 0,8 razdalji in 0,2 teži ustvari nova spremenljivka, ki predstavlja čas dostave, kar je ključni dodatek k podatkovni zbirki. Obstaja možnost izračunavanja dodatnih spremenljivk, ustvarjenih za potrebe statističnih testov.



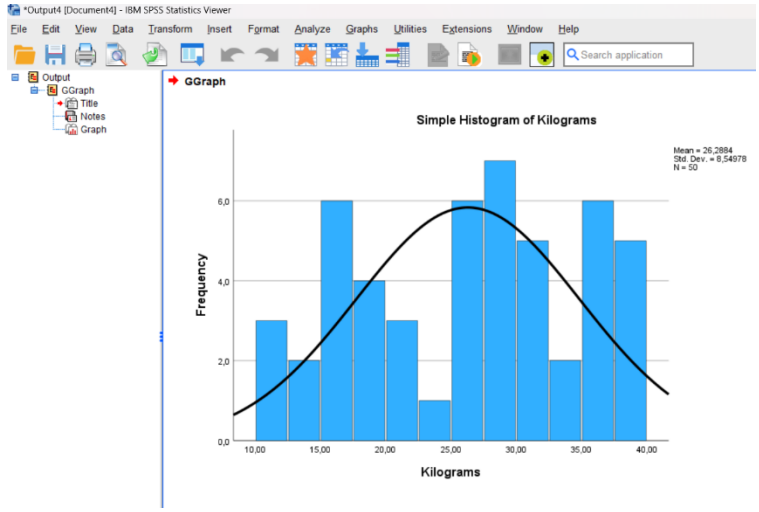
Slika 7.8 Postopek izračuna spremenljivk.

S tem zaključujemo kratek pregled funkcij za upravljanje podatkov, ki jih pokriva SPSS in bi lahko bili koristni pri nadaljnjih testih modelov, ki jih obravnavamo v tem poglavju. Nadaljevali bomo s fazami, ki so potrebne, preden lahko izvedemo statistični test v programu SPSS.

7.3 Priprava na testiranje

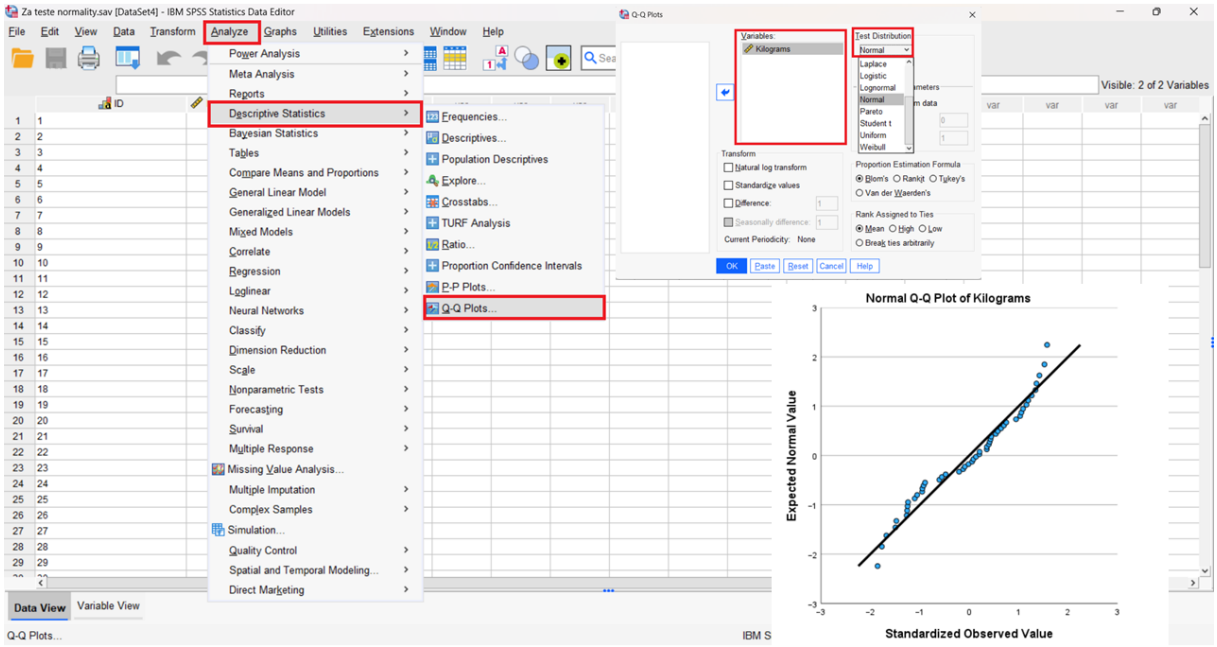
Preden se lotimo statističnih testov, se je treba držati standardnega postopka analize podatkov, ki vključuje raziskovanje podatkov (kot je opisano v poglavjih 7.1 in 7.2), analizo podatkov in interpretacijo rezultatov (Garth, 2008; George in Mallery, 2022). V tem poglavju se osredotočamo na analizo podatkov z uporabo programske opreme SPSS. Ker so bile hipoteze obravnavane že v prejšnjih poglavjih, se bomo osredotočili predvsem na izvajanje testov normalnosti v programu

SPSS. Obstajajo tri metode za ocenjevanje normalnosti: histogram, QQ-plot in test normalnosti. Priporočljivo je uporabiti vsaj dve, če ne vseh treh možnosti, saj vsaka od njih zagotavlja različne informacije (Ghasemi in Zadesiasl, 2012). Če želite ustvariti histogram, pojdite na "Grafi" in nato na "Grafični gradnik". V novem oknu izberite "Histogram". Če imate več spremenljivk, morate ta postopek ponoviti za vsako spremenljivko posebej, da dobite rezultate. Histogram potrdi preskus normalne porazdelitve, če so stolpci, ki predstavljajo vrednosti spremenljivk, podobni zvonovi krivulji. Če so palice bolj nagnjene na levo ali desno stran, to lahko pomeni eksponentno porazdelitev. Na primer, ustvarili smo zbirko podatkov s 100 identifikacijskimi podatki, vsak s spremenljivko, ki predstavlja težo v kilogramih. Po navodilih smo ustvarili histogram, kot je prikazano na sliki 7.9. Kot je razvidno iz slike, so stolpci razporejeni po grafu, in čeprav morda ne odražajo popolnoma krivulje, vseeno kažejo na normalno porazdelitev in pozitiven rezultat testa (George & Mallery, 2022; Goeman & Solari, 2021).



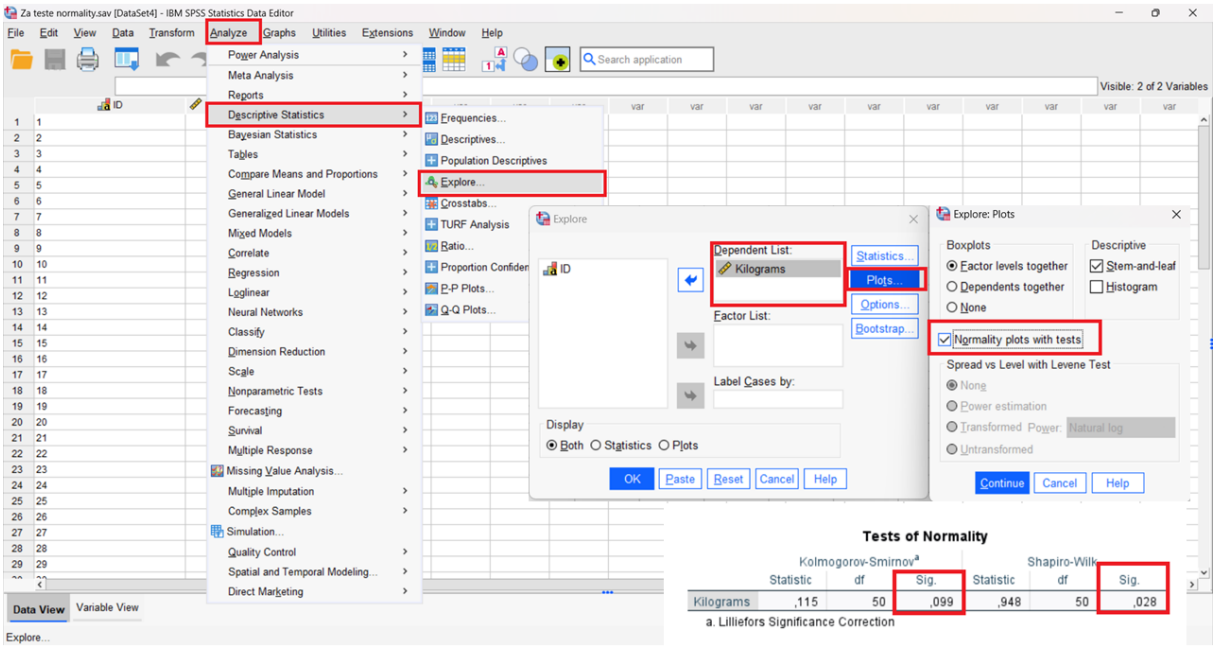
Slika 7.9 Histogram rezultatov testa normalnosti.

Druga možnost za izvajanje testov normalnosti je diagram QQ, ki ga lahko sprožite tako, da kliknete "Analyze", nato "Descriptive Statistics" in izberete "Q-Q Plots". Prednost tega pristopa je, da omogoča hkratno ocenjevanje več spremenljivk (Williamson, b.d.). Test se šteje za uspešnega, če se točke na grafu tesno združijo okoli ravne črte, ki predstavlja normalno porazdelitev. Če točke tvorijo "repe", to pomeni, da test normalnosti ni uspel (Andersen in Dennison, 2018). Z uporabo iste podatkovne zbirke iz testa histogramskega grafa smo izvedli test Q-Q Plot. Na spodnji sliki 7.10 lahko opazite, da se večina točk grozda za našo spremenljivko ujema z ravno črto, kar kaže na normalno porazdelitev naših podatkov. Čeprav bi že na tej stopnji lahko sklepali, da je test normalnosti pozitiven, smo se odločili, da poiščemo potrditev z vsemi tremi testi.



Slika 7.10 Nastavitve in rezultati preskusa normalnosti Q-Q ploskve.

Zadnja možnost za izvedbo testa normalnosti je tako imenovani test normalnosti, ki velja za statistični test. Običajno se uporablja test Kolmogorov-Smirnov, pri majhnih vzorcih pa se lahko uporabi test Shapiro-Wilk (Goeaman in Solari, 2021). V programu SPSS lahko ta test izvedete tako, da kliknete "Analyze" (Analiza), nato "Descriptive Statistics" (Opisna statistika) in nato "Explore" (Raziskovanje). V polju "Dependent List" (Seznam odvisnih spremenljivk) morate določiti spremenljivke, ki jih želite preveriti. Nato v razdelku "Plots" (Grafi) izberite "Normality Plots with Tests" (Grafi normalnosti s testi). Test se šteje za uspešnega, če je stolpec Sig (*p-vrednost*) v rezultatih večji od 0,05, kar kaže na normalno porazdelitev. Če je *p-vrednost* manjša od 0,05, to kaže na nenormalno porazdelitev in test se šteje za neuspešnega. Ta test smo izvedli še enkrat, pri čemer smo uporabili isto podatkovno zbirko kot pri prejšnjih testih. Iz rezultatov lahko sklepamo, da je po standardu Kolmogorova-Smirnova test pozitiven, saj je *p-vrednost* večja od 0,05. Pri Shapiro-Wilkovem testu pa je *p-vrednost* nižja, kar kaže na negativen rezultat testa. Do teh različnih rezultatov pride, ker imata oba pristopa različne nastavitve občutljivosti in moč pri odkrivanju odstopanj (Ghasemi in Zahediasl, 2012). Ker smo že izvedli teste Q-Q Plots in teste histogramskega grafa, lahko test normalnosti na splošno štejemo za pozitivnega. S potrditvijo testov normalnosti lahko izvedemo glavne teste, kot je test enega vzorca.



Slika 7.11 Nastavitve in rezultati testa normalnosti.

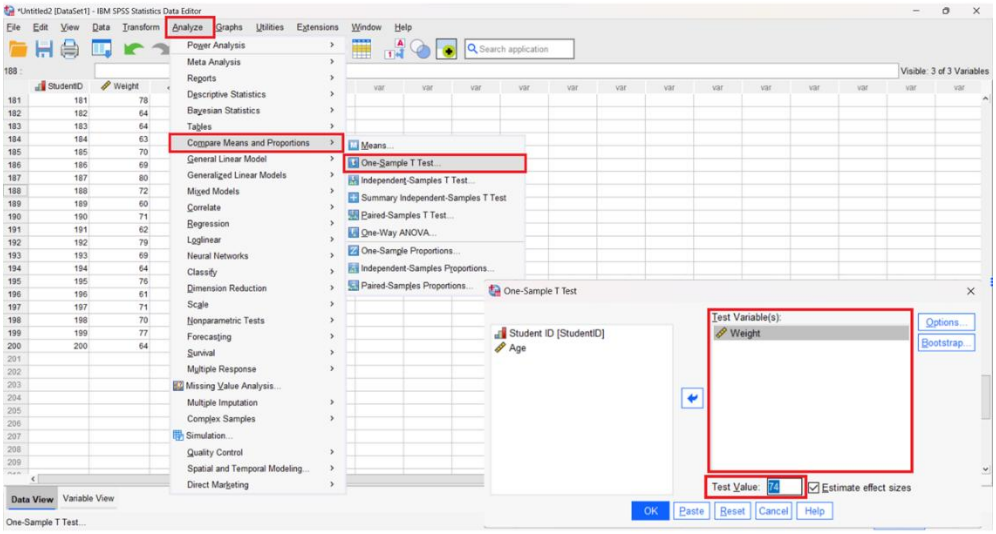
7.4 T-test z enim vzorcem

V prejšnjih poglavjih ste že obravnavali teorijo, ki stoji za T-testom enega vzorca, zato se bomo osredotočili predvsem na izvedbo testa s programsko opremo SPSS. Za naš Eno-vzorčni T-test smo pripravili podatkovno zbirko z vzorcem 200 vnosov, ki vključuje 1 kategorično spremenljivko (ID študenta) in 2 številčni spremenljivki (teža in starost) (Kim, 2015). Po navodilih iz prejšnjih podpoglavij izvedemo:

- Raziščite podatke, in sicer naše **spremenljivke** in **opisno statistiko**, ter določite naše **vprašanje**.
- Preverite **normalnost**, saj bi morala zadostovati le histogram ene spremenljivke in graf Q-Q.
- Postavite hipotezo, pri kateri se pri **ničelni** spremenljivki spremenljivka ne razlikuje od določene vrednosti, pri **alternativni** pa se razlikuje.
- Izvedite **študentski T-test**.
- Interpretirajte rezultate, osredotočite se na to, ali je **ničelni test zavrnjen** ali **ne**, odgovorite na vprašanje in napišite poročilo o našem testu.

V našem primeru smo se odločili, da se bo naše vprašanje glasilo: "Ali je povprečna teža učencev večja od 74 kilogramov?". Po vprašanju določimo hipotezo za vprašanje, ki se glasi: "Ničelna =

razlike ni" in "Alternativna = razlika obstaja". Izvedli smo histogram in Q-Q grafe, da bi preverili teste normalnosti, po njihovem zaključku pa je sledil T-test. T-test izvedemo tako, da kliknemo "Analyze" (Analiziraj) in nadaljujemo s "Compare Means" (Primerjaj sredine) in "One-Sample T-test" (Eno-vzorčni T-test). V polje za spremenljivko testa vpišemo ID našega študenta, vrednost testa nastavimo na 74 in zaženemo test (glej sliko 7.12).



Slika 7.12 Nastavitve T-testa z enim vzorcem.

Po potrditvi testa se prikaže drugo okno z rezultati analize (glejte sliko 7.13). To okno vsebuje več informacij o naši analizi. V tem primeru sta obe *p-vrednosti* nižji od 0,05, kar kaže na pomembnost testa. Poleg tega preverimo vrednosti *t* in *df*, ki sta v našem primeru -9,806 oziroma 199. Iz teh rezultatov lahko sklepamo, da je naša ničelna hipoteza zavrnjena. Zato je celotno poročilo o rezultatih naslednje: "Povprečna teža učencev je bistveno nižja (povprečje = 69,63) od vrednosti 74 kg (t-test za 1 vzorec, 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	95% Confidence Interval of the Difference Upper
Weight	-9,806	199	<,001	<,001	-4,370	-5,25	-3,49

One-Sample Effect Sizes

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

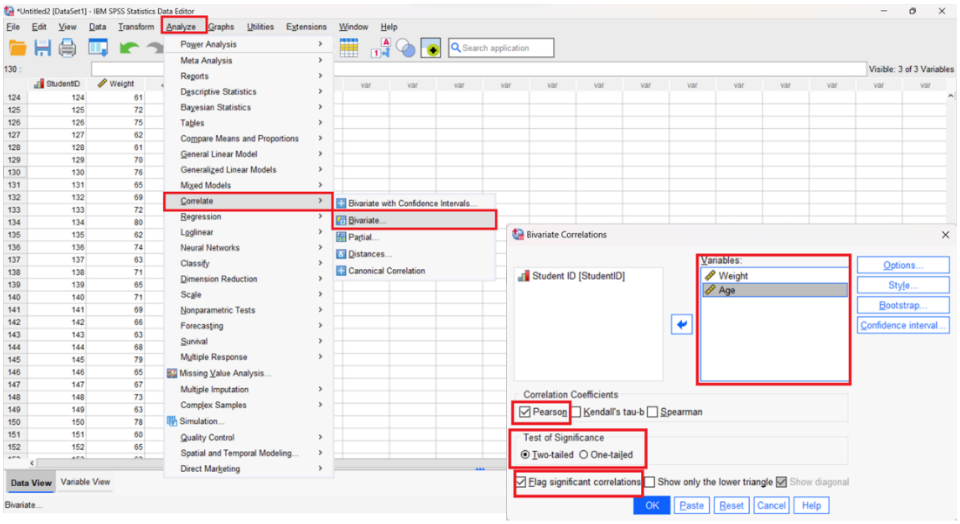
a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Slika 7.13 Rezultati T-testa z enim vzorcem.

7.5 Korelacija

Preidimo na drugi test, to je test korelacije. Izvedli ga bomo z isto podatkovno bazo kot v primeru t-testa. Podobno kot pri t-testu bomo upoštevali postopek z nekaj spremembami. Pri izvajanju korelacije med dvema spremenljivkama je pomembno določiti, katera je odvisna in katera neodvisna spremenljivka (Janse *et al.*, 2021; Mishra *et al.*, 2019). To izbiro lahko opravite na podlagi svojega raziskovalnega vprašanja. V našem primeru želimo raziskati "Ali obstaja povezanost med starostjo učenca in njegovo telesno težo?". V skladu z vprašanjem upoštevamo težo kot odvisno spremenljivko in starost kot neodvisno spremenljivko, saj želimo raziskati, ali so razlike v starosti povezane z razlikami v teži. Opredelimo ničelno in alternativno hipotezo (glej 7.3 in 7.4), nato pa izvedemo test s klikom na "Analyze", ki mu sledita "Correlate" in "Bivariate".

Obe spremenljivki je treba postaviti v polje "Variable" (Spremenljivka). Prepričajte se, da so izbrane ali nastavljene možnosti "Pearson", "Two-Tailed" in "Flag Significant" (glejte sliko 7.14). V tem primeru smo izbrali "Pearson", ker so naši podatki kazali normalno porazdelitev in jih je bilo mogoče analizirati s parametričnimi metodami. Če normalna porazdelitev ni nakazana, je treba uporabiti neparametrične metode (v tem primeru bi namesto Pearsonove izbrali Spearmanovo) (George in Mallery, 2022; McClure, 2005).



Slika 7.14 Nastavitve korelacijskega testa.

Rezultate ponovno dobimo v novem oknu (glej sliko 7.15). Iz rezultatov je razvidno, da je naša Pearsonova korelacija -0,038, *p-vrednost* pa 0,596. Pri korelacijski analizi velja, da je vrednost korelacije tem bližje ničli, čim šibkejša je korelacija med spremenljivkama. V našem primeru je korelacija zelo blizu nič, kar kaže na to, da med spremenljivkama ni pomembne korelacije (McClure, 2005). Poleg tega visoka *p-vrednost* (0,596) nakazuje, da ni bistvenih dokazov za sklepanje, da obstaja pomembna korelacija med izbranima spremenljivkama (Williamson, b. d.). Zato naša ničelna hipoteza ni zavrnjena. Na podlagi tega lahko poročamo, da "med starostjo in telesno težo učencev ni bilo korelacije".

Correlations

Correlations

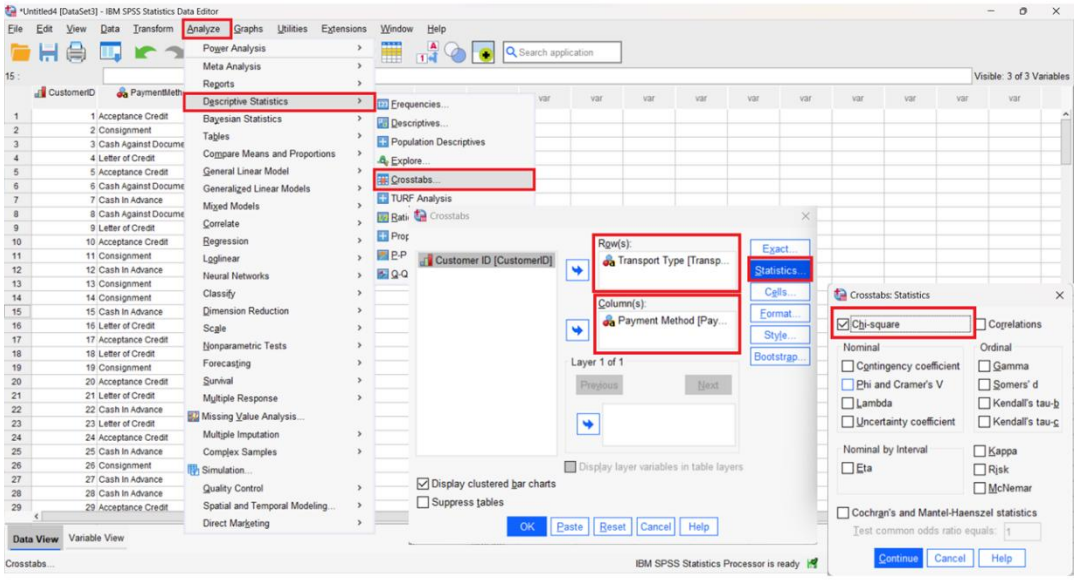
		Weight	Age
Weight	Pearson Correlation	1	-,038
	Sig. (2-tailed)		,596
	N	200	200
Age	Pearson Correlation	-,038	1
	Sig. (2-tailed)	,596	
	N	200	200

Slika 7.15 Rezultati korelacijskega preskusa.

7.6 Chi-kvadrat

Tretji test, ki ga bomo izvedli v programu SPSS, je test Chi-Square. Za razliko od prejšnjih dveh testov test Chi-Square primerja dve kategorični spremenljivki in ne številčnih spremenljivk (Turhan, 2020). Tako kot postopek v razdelkih 7.4 in 7.5 začnemo raziskovanje podatkov in oblikovanje raziskovalnega vprašanja. V našem primeru imamo logistično podjetje z 200 strankami in podatke

o vrsti plačila in vrsti prevoza, ki ga je izbrala vsaka stranka. Vprašanje, na katero želimo odgovoriti, se glasi: "Ali različne vrste plačil kažejo različne preference glede vrste prevoza?" Ker imamo opravka samo s kategoričnimi spremenljivkami, test normalnosti ni potreben. Določimo našo ničelno hipotezo (Preference za vrste prevoza so enake med vsemi vrstami plačil) in alternativno hipotezo. Za izvedbo analize Chi-Square kliknite "Analyze", nato "Descriptive Statistics" in izberite "Crosstabs". Ključno je, da spremenljivke, ki temeljijo na vašem raziskovalnem vprašanju, postavite v polje stolpcev ali vrstic (glejte sliko 7.16) (Garth, 2008).



Slika 7.16 Nastavitve testa Chi-Square.

Po analizi se v novem oknu prikažejo rezultati (glejte sliko 7.17). V tem oknu lahko opazite, da je vrednost Pearsonovega chi-kvadrata 11,614, vrednost *df* 12 in *p-vrednost* (asimptotska pomembnost) 0,477. Na podlagi teh rezultatov lahko sklepamo, da med spremenljivkama ni pomembne povezave in da ničelna hipoteza ni zavrnjena. Zato poročilo sledi: "Med različnimi vrstami plačil za različne vrste prevoza ni zaznati pomembnih preferenc (dvostranski test Chi-Square, 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

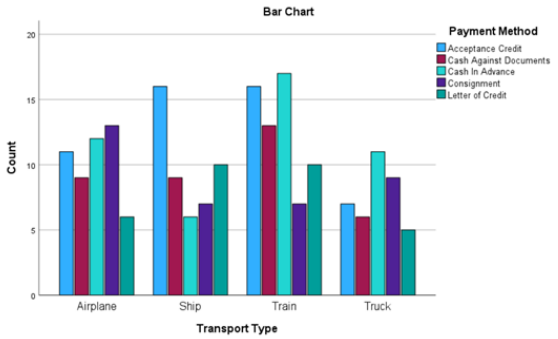
Count

		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		

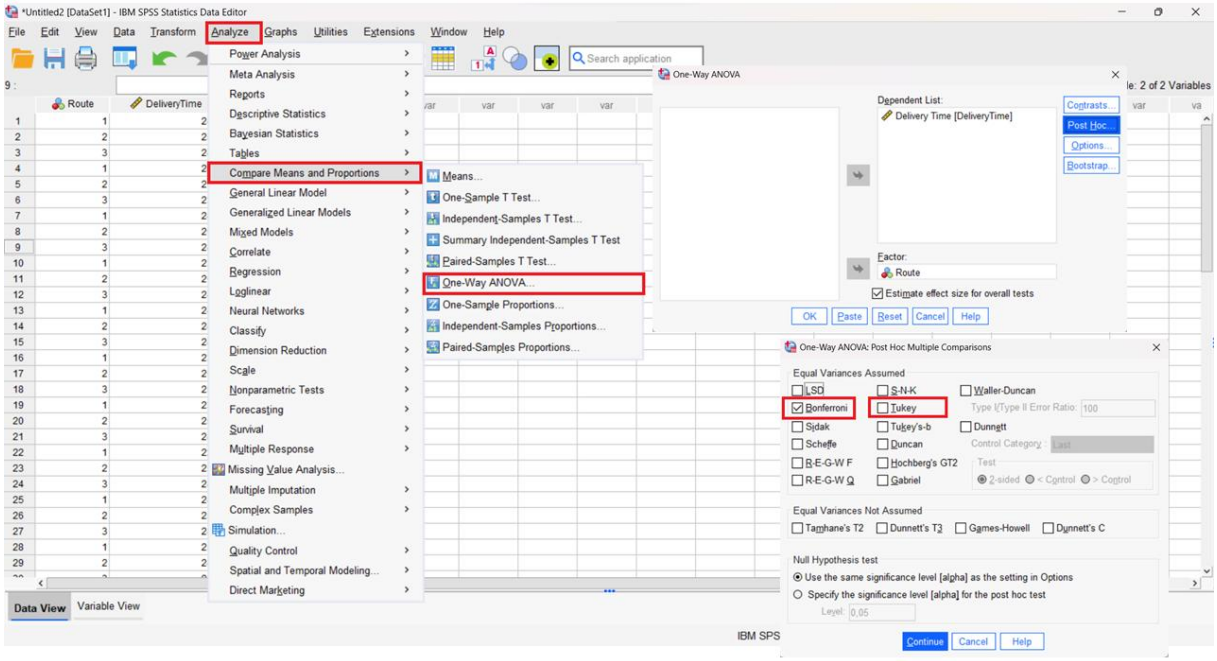
a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,89.



Slika 7.17 Rezultati testa Chi-Square.

7.7 ANOVA

Zadnji test, ki ga bomo obravnavali, je test ANOVA, pri čemer se bomo osredotočili na preprostejši model, znan kot enosmerna ANOVA, ki vključuje kategorično spremenljivko in številčno spremenljivko (Goeman in Solari, 2021). Tako kot pri T-testu bomo uporabili enak postopek: raziskali bomo podatke, oblikovali raziskovalno vprašanje, izvedli test normalnosti in postavili hipoteze. Obravnavajmo študijo primera prometnega dispečerja, zaposlenega v logističnem podjetju. Dispečer tesno sodeluje s partnerskim podjetjem in redno načrtuje tri različne poti za tovornjake za dostavo blaga. Zaradi politike "Just-in-time", ki poudarja hitrejše dobave, se postavlja vprašanje: "Ali izbira poti dostave vpliva na čas dostave za podjetje?" Če želite v programu SPSS izvesti test ANOVA, pojdite na "Analyze", ki mu sledi "Compare Means..." in nato "One-way ANOVA". Odvisno spremenljivko postavite v polje "Seznam odvisnih", spremenljivko dejavnik pa v polje "Dejavnik" (glejte sliko 7.18). Za temeljito analizo smo vključili tudi nastavitev Post Hoc. Pomembno je opozoriti, da je treba analizo Post Hoc izvesti le, če je začetni test ANOVA pozitiven. Z uporabo analize Post Hoc lahko določimo najbolj optimalno izbiro (v našem primeru pot). Najzanesljivejši metodi, ki ju je treba uporabiti za analizo Post Hoc, sta Bonferronijeva metoda ali metoda Tukey HSD (Goeman in Solari, 2021).



Slika 7.18 Nastavitve ANOVA.

Rezultati naše analize kažejo, da je vrednost *F-statistike* 11,173 (višje vrednosti kažejo na večje razlike med skupinami) in *p-vrednost* <0,001, kar pomeni, da je naša ničelna hipoteza zavrnjena (glej sliko 7.19). Ker med tremi potmi obstaja pomembna razlika (<0,001), je post hoc test veljaven tudi v našem primeru (George in Mallery, 2022). Po izvedbi Bonferronijevega korekcijskega testa lahko vidimo, da so najboljše *p-vrednosti* zabeležene v primeru poti 2 (glej sliko 7.19). V poročilu lahko ugotovimo, da "je bila ugotovljena pomembna razlika pri izbiri poti dostave v korelaciji s časom dostave (1-way ANOVA, *F = 11*,173, *df* = 47, *p-vrednost* = <0,001). Pot 2 je imela najboljše rezultate glede časa dostave."

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 7.19 Začetni rezultati ANOVA in rezultati testa Post Hoc.

To poglavje knjige zaključimo z zavedanjem, da smo v tem poglavju obravnavali nekaj najpogostejših testov. Obstajajo še drugi testi, kot so ANOVA s ponovljenimi meritvami, testi zanesljivosti in testi občutljivosti, ki jih je prav tako mogoče modelirati in analizirati s programsko opremo SPSS. Ti dodatni testi zagotavljajo širši nabor orodij za analizo podatkov in prinašajo smiseln vpogled v različne raziskave in praktične aplikacije.

Literatura poglavje 7

- 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, vključno z R in SQL

Kaj je poslovna analitika (BA)? Katere probleme rešuje in katera orodja uporablja? Kaj sta R in SQL? Kako je BA povezana z R in SQL? Ali obstajajo primeri dobrih praks, kjer se ti programi uporabljajo za reševanje poslovnih problemov v logistiki?

Na ta in podobna vprašanja bomo poskušali odgovoriti v naslednjem poglavju.

8.1 Kaj je poslovna analitika?

BA predstavlja celovit pristop k analizi podatkov in poslovnemu odločanju. Gre za okolje, ki temelji na podatkih, njegov cilj pa je izboljšati poslovno uspešnost podjetja z zagotavljanjem podlage za bolj premišljeno odločanje. Gre za sistematičen proces razmišljanja, ki uporablja kvalitativna, kvantitativna in statistična računalniška orodja in metode za analizo podatkov, pridobivanje vpogleda, informiranje in podporo odločanju. Vsaka posamezna analiza lahko uporablja različne tehnike, vključno z diagnostičnimi, napovednimi, preskriptivnimi in optimizacijskimi modeli (Power et al., 2018). Tako po Mikalef et al. (2019) BA zagotavlja načrt tako za akademsko raziskovanje kot za praktično izvajanje, pri čemer poudarja transformativni potencial analitike, če je ustrezno vključena v organizacijske procese. V skladu s tem avtorji navajajo, da BA od organizacij zahteva, da korenito preoblikujejo način pristopa, načrtovanja in izpopolnjevanja takšnih pobud, način načrtovanja in upravljanja virov ter strateško uskladitev, pa tudi ponovno ovrednotijo pričakovane rezultate delovanja, njihovo povezanost s strateškimi cilji in posledično razvijejo ustrezne KPI (Mikalef et al., 2019).

Glavna naloga BA je zagotoviti prenos znanja, da se zagotovi skladna povezava med neobdelanimi podatki in poslovnimi odločitvami. Splošni cilj je poslovna učinkovitost z "vertikalizacijo", uporabnostjo in integracijo z operativnimi sistemi (Kohavi, Rothleder in Simoudis, 2002). BA ima številna področja uporabe in povezane izpeljanke: Finančna analitika, analitika oskrbovalne verige, analitika kriz, analitika znanja, analitika trženja, analitika strank, analitika storitev, analitika človeških virov, analitika talentov, analitika procesov, analitika tveganja (Holsapple, Lee-Post in Pakath, 2014).

Obstajajo tri vrste platform za poslovno analitiko:

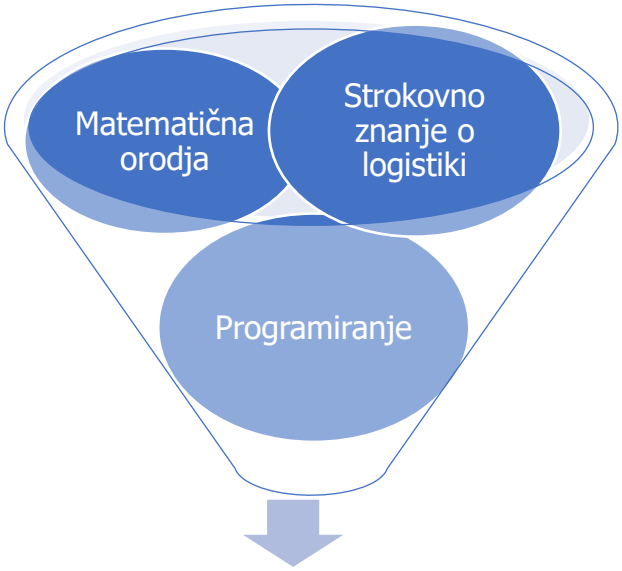
- **Opisni** - pregledajo obstoječe podatke ter zagotovijo zbirne statistične podatke in osnovno vizualizacijo.
- **Prediktivni** - na podlagi obstoječih podatkov ocenjujejo najverjetnejše scenarije za prihodnost.
- **Preskriptivni** - samodejno obdelujejo velike količine podatkov, poslovna pravila, tržne razmere itd. Te platforme uporabljajo metode strojnega učenja in umetne inteligence. Cilj je popolnoma avtomatizirano odločanje o tem, katere ukrepe naj podjetje sprejme glede na trenutne razmere, da bo doseglo želene poslovne cilje.

Zanimanje za velike podatke in poslovno analitiko se je v zadnjem desetletju eksponentno povečalo (Mikalef et al., 2019). Sodobni BA korenini v nenehnem napredku sistemov za podporo odločanju. Ti napredki vključujejo vse zmogljivejše mehanizme za pridobivanje, ustvarjanje, asimilacijo, izbiro in oddajanje znanja, pomembnega za sprejemanje odločitev. Glede na dediščino podpore odločanju je poslovna analitika nujno del teh mehanizmov in jih izkorišča. Znanje, ki ga je treba obdelati, sega od kvalitativnega do kvantitativnega, BA pa se ukvarja z delovanjem na obeh vrstah znanja, kot je primerno za zadevno odločitev (Kohavi, Rothleder in Simoudis, 2002). Razlog, zakaj naj bi nekatere organizacije uporabljale BA, je v problemih, ki jih rešuje. Težave, ki jih poudarja BA, so omejene težave za učinkovito upravljanje podjetja. V skladu s tem obstaja več razlogov za uporabo BA (Holsapple, Lee-Post in Pakath, 2014):

- Doseganje konkurenčne prednosti
- Podpiranje strateških in taktičnih ciljev organizacije
- Boljša organizacijska uspešnost
- Boljši rezultati odločitev
- Boljši ali bolj informirani procesi odločanja.
- Pridobivanje znanja
- Pridobivanje vrednosti iz podatkov

Ne glede na vrsto platforme, ki se uporablja v BA za reševanje in podporo procesa odločanja v posameznem podjetju, obstajajo trije ključni stebri vsake rešitve BA (slika 8.1). Na splošno BA zavzema mesto v spektru med računalništvom/matematiko/podatkovno znanostjo (na eni strani) ter poslovanjem in upravljanjem (na drugi strani). Poslovna analitika zahteva tako tehnično kot poslovno znanje. Glavna težava pri oblikovanju BA je, da meje niso jasno razmejene (Power et al.,

2018). V skladu s tem so za izvedbo BA potrebna matematična orodja, ki omogočajo prepoznavanje, pridobivanje in predstavitev vpogledov na ustrezen način prek tabel, grafikonov, formul itd. Poleg tega programska orodja služijo kot podpora tovrstnim dejavnostim ter omogočajo hitre in brezhibne izračune v primerjavi s tradicionalnim pristopom s papirjem in pisalom. Ne nazadnje je potrebno tudi strokovno znanje s področja logistike na določenem področju ali pri določenem poslovnem problemu, da se natančno določijo ključni vplivni dejavniki in z njimi povezan ekosistem.



Slika 8.1 Ključni stebri BA v okviru oskrbovalne verige in logistike.

Ključni uporabnik je poslovni uporabnik, čigar delo, po možnosti na področju prodaje, trženja ali prodaje, ni neposredno povezano z analitiko kot tako, vendar običajno uporablja analitična orodja za izboljšanje rezultatov nekega poslovnega procesa po eni ali več dimenzijah (kot sta dobiček in čas za vstop na trg). Poslovni uporabniki se ne želijo ukvarjati z naprednimi statističnimi koncepti; želijo preproste vizualizacije in rezultate, ki so pomembni za nalogo (Kohavi, Rothleder in Simoudis, 2002).

8.2 Kaj je R?

R je integriran paket programske opreme za manipulacijo s podatki, izračunavanje in grafično Prikaz (R Core Team, 2019). Med drugim ima:

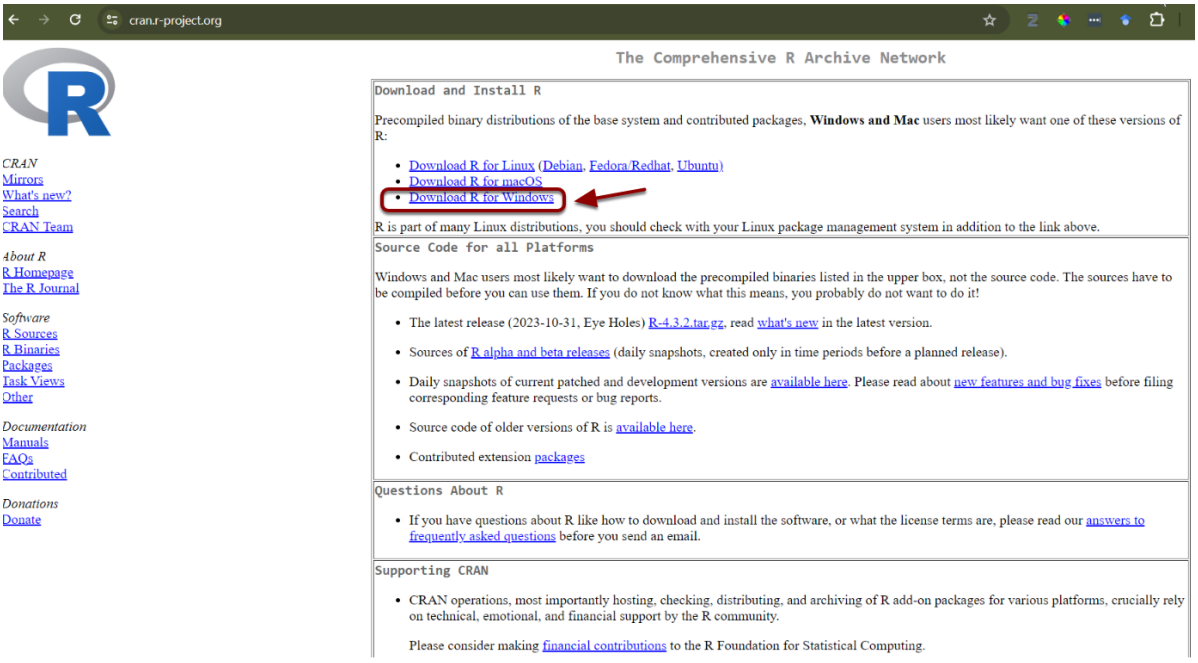
- učinkovito orodje za obdelavo in shranjevanje podatkov,

- nabor operatorjev za izračune na poljih, zlasti matrikah,
- velika, skladna in integrirana zbirka vmesnih orodij za analizo podatkov,
- Grafične možnosti za analizo in prikaz podatkov neposredno na računalniku ali na papirju ter dobro razvit, preprost in učinkovit programski jezik (imenovan "S"), ki vključuje pogoje, zanke, rekurzivne funkcije, ki jih določa uporabnik, ter vhodne in izhodne možnosti. (Večina funkcij, ki jih ponuja sistem, je dejansko napisana v jeziku S.)

Glavni prednosti programa R sta, da je brezplačen in da je na voljo veliko pomoči na spletu. Je precej podoben drugim programskim paketom, kot je MatLab (ni brezplačen), vendar je uporabniku prijaznejši od programskih jezikov, kot sta C++ ali Fortran (Torfs in Brauer, 2014). R je v veliki meri sredstvo za nov razvoj metod interaktivne analize podatkov. Hitro se je razvijal in bil razširjen z veliko zbirko paketov. Vendar je večina programov, napisanih v R, v bistvu efemernih, napisanih za eno samo analizo podatkov (R Core Team, 2019).

Namestitev programov R in R Studio

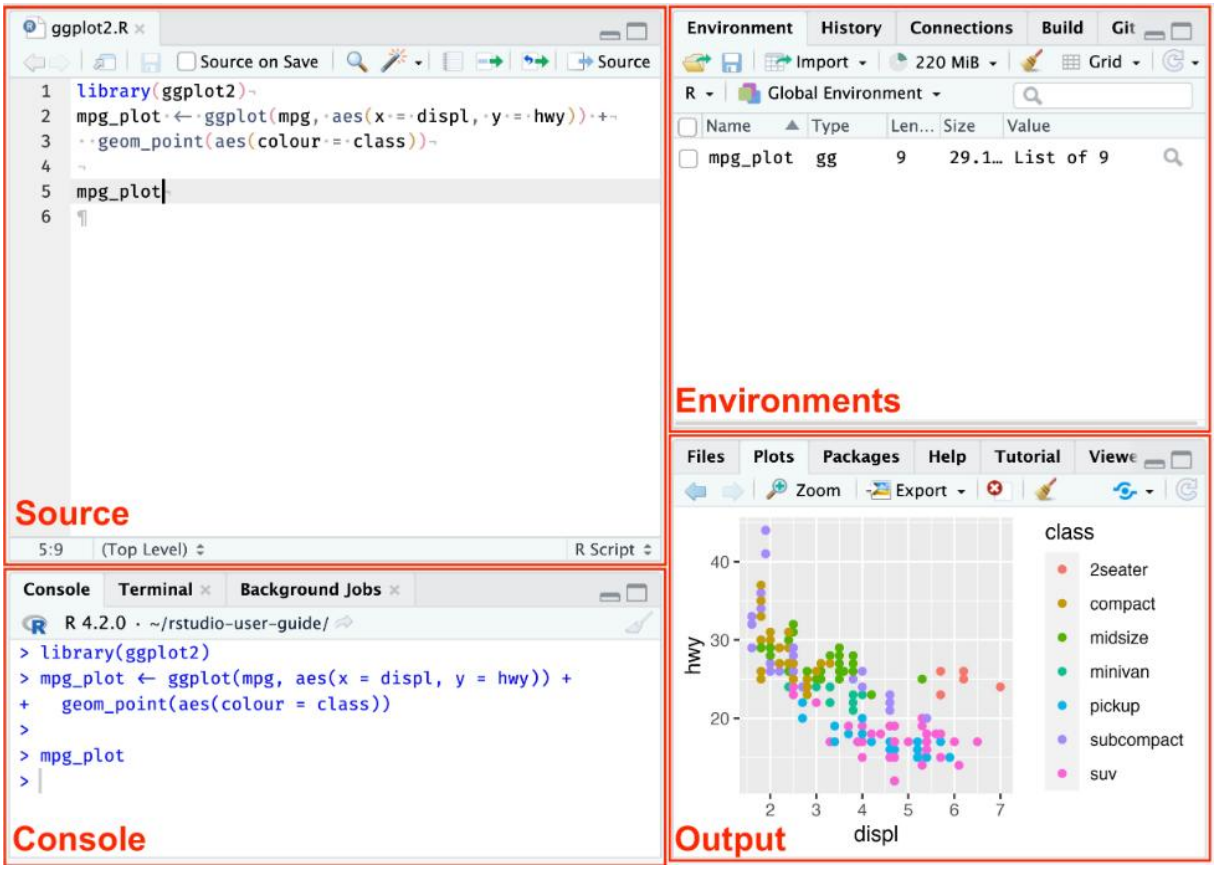
Če želite namestiti R, pojdite na stran r-project.org in kliknite spustni gumb R za določen operacijski sistem v računalniku (običajno Windows) (slika 8.2).



Slika 8.2 Stran za prenos programske opreme R.

Tako boste prenesli programsko opremo R, postopek namestitve pa je enak kot pri drugi programski opremi. Ko je programska oprema R nameščena, je na voljo brez naprednega

integriranega razvojnega okolja (IDE), ki je uporabnikom v pomoč in podporo pri izvajanju različnih analiz. Čeprav je mogoče opraviti katero koli analizo samo z nameščenim programom R, je bolje, da ga povežete s kakšnim sodobnim IDE, kot je RStudio, ki je eno izmed najbolj priljubljenih IDE. Postopek namestitve programa RStudio je podoben postopku namestitve osnovne programske opreme R. Pojdite na spletno stran https://posit.co/download/rstudio-desktop/, poiščite odprtokodno licenco RStudio Desktop, jo prenesite in namestite. Po namestitvi programov R in RStudio se uporabniku prikaže naslednji zaslon uporabniškega vmesnika (slika 8.3).



Slika 8.3 Uporabniški vmesnik in R & Rstudio (RStudio, 2024).

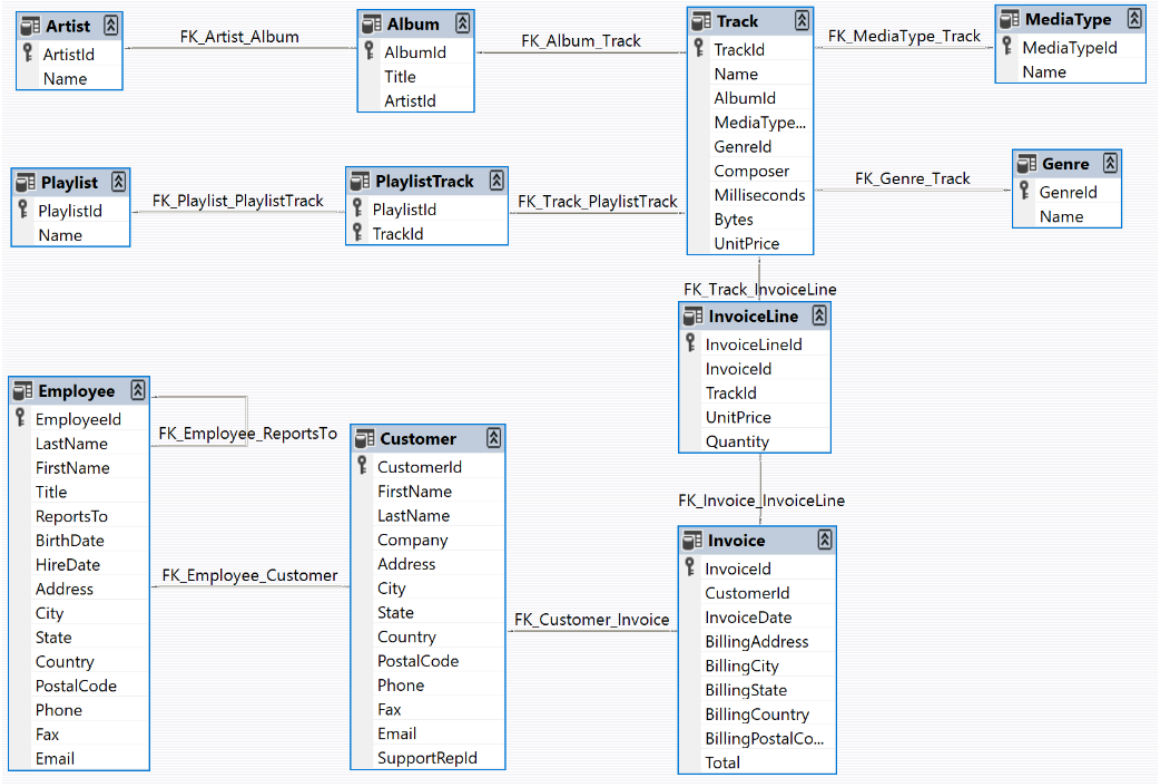
Uporabniški vmesnik programa RStudio ima štiri osnovne plošče (RStudio, 2024):

- Podokno z viri (Source);
- Podokno konzole (Console);
- Podokno Okolje, ki vsebuje zavihke Okolje, Zgodovina, Povezave, Sestavljanje, VCS in Učbenik (Environments);
- Podokno Izhod, ki vsebuje zavihke Datoteke, Sklopi, Paketi, Pomoč, Pregledovalnik in Predstavitev (Output).

Vsako od podoken ter pododdelek in možnosti v njem uporabnikom omogočajo izvajanje različnih operacij, nadzor nad nekaterimi analizami podatkov ali bolj strukturiran in jasen pregled nad procesom analize podatkov, ki se izvaja.

8.3 Kaj je SQL in kako je povezan z BA in R?

Podatkovna zbirka Chinook je vzorčna podatkovna zbirka, ki se uporablja za učenje in prikaz sistemov za upravljanje podatkovnih zbirk (DBMS) in poizvedb SQL. Zasnovana je kot trgovina z digitalnimi mediji, ki uporablja resnične podatke iz knjižnice iTunes, izmišljena imena/naslove za stranke in zaposlene ter naključne podatke za prodajne informacije. Podatkovna baza vsebuje različne tabele, ki predstavljajo podatke glasbene trgovine, vključno s podatki o izvajalcih, albumih, skladbah, strankah, računih in drugih (slika 8.4).



Slika 8.4 Podatkovni model podatkovne zbirke Chinook.

Slika 8.4 prikazuje podatkovni model podatkovne zbirke Chinook z različnimi podatkovnimi tabelami in njihovimi ključi (edinstvenimi identifikatorji) ter skupnimi tabelami, kot je tabela PlaylistTrack. Tabele posredujejo različne informacije o dani digitalni trgovini (preglednica 8.1).

Preglednica 8.1 Informacije v vseh tabelah podatkovne zbirke Chinook.

Table Name	Description
Artist	Contains information about music artists.
Album	Contains information about music albums, each associated with an artist.
Track	Contains information about individual music tracks, including references to albums, media types, and genres.
Genre	Contains information about music genres.
MediaType	Contains information about different types of media (e.g., audio, video).
Customer	Contains information about customers, including contact details and support representative information.
Employee	Contains information about employees, including their roles, reports-to relationships, and contact details.
Invoice	Contains information about invoices, including customer details, billing information, and total amounts.
InvoiceLine	Contains detailed information about each item in an invoice, including references to tracks and quantities.
Playlist	Contains information about playlists.
PlaylistTrack	Links tracks to playlists, indicating which tracks are included in which playlists.

8.4 Kako so povezani poslovna analitika, SQL in R?

Povezava med BA, SQL in R je naravna, saj morajo biti vsi poslovni podatki shranjeni v podatkovnih zbirkah SQL. To je še vedno idealističen cilj, saj je upravljanje podatkov v nekaterih delih malih in srednje velikih podjetij, ki še vedno ne razumejo v celoti moči podatkov, še vedno slabo. V velikih podjetjih so to že zdavnaj spoznali in podatki so ustrezno strukturirani v podatkovnih zbirkah (SQL ali drugih, vendar običajno v SQL). Po drugi strani je analizo podatkov mogoče opraviti v SQL, vendar je v ta namen bolje uporabiti statistično usmerjeno programsko opremo, kjer je v ospredju R kot ena najbolj priljubljenih statističnih platform za izvajanje analize podatkov.

Zato lahko SQL in R obravnavamo kot odlično orodje za sodelovanje, kadar gre za problem s področja BA. Glavnih razlogov je več, eden od njih pa je, da se podatki BA dnevno spreminjajo in posodabljajo glede na tržno realnost in dejavnosti podjetja: prodaja, zaposleni, prihodki itd. Podatkovne zbirke SQL so odlične za zajemanje teh sprememb in posodabljanje obstoječih podatkov, medtem ko so skripte R zelo dobre za avtomatizacijo nalog in oblikovanje novih paketov za analizo danih podatkov. Razlog za to je, da je jedro R bolj zgrajeno na konceptu analiziranja podatkov kot na splošnem programiranju, kot je na primer Python.

Poizvedovanje po zbirki podatkov SQL z R

Programski jezik R in podatkovne zbirke SQL sta naravno povezana, saj je programski jezik R namenjen predvsem statistični analizi podatkov, večina podatkov o transakcijah pa je v podatkovnih zbirkah. Način, kako R upravlja manipulacijo podatkov iz podatkovnih zbirk SQL, je prek programskih paketov DBI in RSQLite. Paket DBI zagotavlja standardiziran vmesnik za interakcijo z različnimi DBMS, ki uporabnikom omogoča povezovanje, poizvedovanje in dosledno upravljanje transakcij v različnih podatkovnih zbirkah. Paket RSQLite, ki se drži vmesnika DBI, posebej olajšuje interakcijo s podatkovnimi zbirkami SQLite in uporabnikom omogoča izvajanje poizvedb SQL, pridobivanje podatkov in izvajanje operacij s podatkovnimi zbirkami neposredno iz R. Ti paketi skupaj poenostavljajo postopek dela s podatkovnimi zbirkami v R ter ponujajo skladen in učinkovit potek dela.

Da bi učinkovito prikazali izvajanje operacije SQL iz programa R in ustvarjanje želenih vpogledov v podatke v zvezi z obravnavanim problemom, smo na slikah 8.5 in 8.6 prikazali več izsekov kode.

```
---
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 8.5 Korak kode za vzpostavitev povezave med SQL in R ter raziskovanje podatkovnih tabel, ki jih vsebuje SQL.

Prvi korak pri poizvedovanju po zbirki podatkov SQL prek programa R je vzpostavitev povezave (slika 8.5). Slika prikazuje uporabo paketov DBI in RSQlite, ki omogočata vzpostavitev povezave prek funkcije dbConect(). Rezultat povezave in podatkovne tabele, ki se odkrijejo prek omenjene

povezave, se nato izvozijo prek funkcij dbListTables(), ki izpišejo seznam vseh podatkov, najdenih prek povezave: Album, Izvajalec, Stranka, Zaposleni, Žanr, Račun, Linija računa MediaType, Seznam predvajanja, PlaylistTrack, Skladba.

Ko je povezava vzpostavljena, je mogoče izvesti številne analize, ki so odvisne od poslovnega cilja in prihodnje uporabe danih rezultatov. Na tem mestu bomo zaradi omejenega prostora prikazali le del možnih analiz podatkov z majhnim delom kode in naborom kodnih pravil, ki so potrebna za pridobivanje informacij iz SQL. Koda izvede poizvedbo po zbirki podatkov prek programa R in prikaže najbolje prodajane albume, njihove avtorje in število prodanih albumov (slika 8.6). Tabela 8.2 predstavlja rezultate poizvedovanja po podatkih prek odlomka kode na sliki 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 8.6 Delček kode za poizvedbo po SQL prek R in določitev 10 najbolje prodajanih albumov.

Preglednica 8.2 Deset najbolje prodajanih albumov v digitalni trgovini Chinook.

Album title	Artist name	Quantity sold
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
My Generation - The Very Best Of The Who	The Who	19
International Superhits	Green Day	18
Chronicle, Vol. 1	Creedence Clearwater Revival	18

Reference Poglavje 8

- 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.
- Introductory Statistics. Bentham Science Publishers, Kahl, A. (Publish 2023). DOI:10.2174/97898151231351230101

9. Napovedovanje povpraševanja, vizualizacija in inženiring časovnih vrst v oskrbovalnih verigah

Kaj je napovedovanje povpraševanja? Kako lahko učinkovito vizualiziramo in sklepamo o podatkih o strankah? Kako izvesti inženiring značilnosti časovnih vrst?

Na ta in podobna vprašanja bomo poskušali odgovoriti v naslednjem poglavju.

9.1 Kaj je povpraševanje kupcev in napovedovanje povpraševanja?

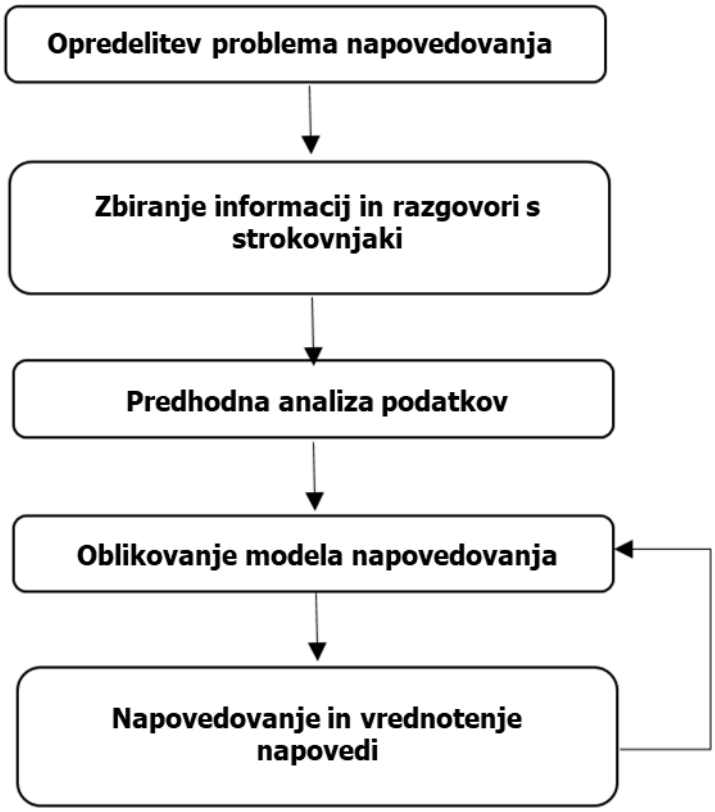
Povpraševanje končnega kupca spravi v tek celotno oskrbovalno verigo zainteresiranih strani (Syntetos et al., 2016). V skladu s tem je povpraševanje strank ključna sestavina za načrtovanje vseh logističnih procesov v oskrbovalni verigi, zato je določanje ravni povpraševanja strank zelo zanimivo za upravljavce oskrbovalne verige. Poleg tega je napovedovanje povpraševanja bistvena dejavnost za načrtovanje in razporeditev logističnih dejavnosti v opazovani oskrbovalni verigi (Mircetic et al., 2017). Natančni modeli napovedovanja povpraševanja neposredno vplivajo na zmanjšanje logističnih stroškov, saj zagotavljajo oceno povpraševanja strank (Mircetic et al., 2016). Napovedovanje v oskrbovalnih verigah presega operativno nalogo ekstrapolacije zahtev po povpraševanju na eni stopnji. Vključuje kompleksna vprašanja, kot sta usklajevanje oskrbovalne verige in izmenjava informacij med več deležniki (Syntetos et al., 2016).

Povpraševanje kupcev in spremljajoče napovedi so za podjetja ključnega pomena, saj zagotavljajo osnovne podatke za načrtovanje in nadzor vseh funkcionalnih področij, vključno z logistiko, trženjem, proizvodnjo itd. (Mircetic, 2018). Če bi bilo končno povpraševanje potrošnikov konstantno ali z gotovostjo znano precej vnaprej, potem bi bilo delovanje oskrbovalne verige preprosto (za nazaj) za načrtovati. Vendar pa povpraševanje ni znano, zato ga je treba napovedati. Prav zaradi negotovosti, povezane s tem povpraševanjem, je upravljanje oskrbne verige zelo zahtevno (Syntetos et al., 2016). Na učinkovitost napovedovanja povpraševanja vplivajo negotovosti, ki so neločljivo povezane s časovnimi vrstami povpraševanja, ki jih imajo oskrbovalne verige (Rostami-Tabar, 2013). Posledično sta obravnava in razumevanje teh negotovosti velik izziv za menedžerje pri usklajevanju in načrtovanju dejavnosti v oskrbovalnih verigah (Mircetic, 2018).

Negotovost povpraševanja je eden največjih izzivov za sodobne oskrbovalne verige. Nedavna pandemija COVID-19 je to vprašanje še dodatno poudarila, saj je povzročila obsežne motnje, ki so otežile načrtovanje in nadzor oskrbovalne verige (Nikolopoulos et al., 2020). Napovedovanje povpraševanja v oskrbovalnih verigah pogosto vključuje napovedovanje povpraševanja po številnih postavkah. Napovedovalci v oskrbovalnih verigah običajno ekstrapolirajo podatke časovnih vrst za vsako enoto za vzdrževanje zalog posebej. Trgovec na drobno lahko na primer uporabi podatke s prodajnih mest za izdelavo napovedi na ravni posamezne trgovine (Mircetic et al., 2022).

9.2 Koraki napovedovanja povpraševanja v oskrbovalnih verigah?

V skladu z zgoraj navedenimi trditvami in ugotovitvami glede pomena povpraševanja in napovedovanja povpraševanja za oskrbovalne verige je pri oblikovanju napovedi v oskrbovalnih verigah pomembno upoštevati posebne korake (slika 9.1).



Slika 9.1 Osnovni koraki za pravilno izvajanje napovedi v podjetju (Makridakis et al., 1998; Makridakis et al., 1983).

Vsak od korakov na sliki 9.1 ima svoje prednosti in prispeva k oblikovanju zanesljivih in uporabnih (poslovno usmerjenih) napovedi. V skladu s tem je opredelitev problema pogosto najtežji del

napovedovanja in zahteva razumevanje, kako se bodo napovedi uporabljale, ter vloge funkcij napovedovanja v opazovanem podjetju. Napovedovalec mora porabiti precej časa za komunikacijo z vsemi, ki sodelujejo pri zbiranju podatkov, vzdrževanju podatkovne zbirke in uporabi napovedi za načrtovanje prihodnosti. Eden od glavnih oteževalnih dejavnikov pri opredelitvi problema je, kako se bo končna napoved uporabljala pri vsakodnevnih logističnih dejavnostih (kakšna platforma, zasnova programske opreme, uporabniški vmesnik itd.)

Za korak zbiranja informacij sta vedno potrebni vsaj dve vrsti informacij: statistični podatki in strokovno znanje ljudi, ki zbirajo podatke in uporabljajo napovedi. V praksi je pogosto težko pridobiti zgodovinske podatke za oblikovanje dobrega statističnega modela. Prav tako obstaja veliko nerazumevanje, kaj so podatki o povpraševanju in kaj se lahko uporabi kot njihov približek. Obstaja slaba praksa uporabe podatkov o pošiljkah in dobavah kot približka podatkov o povpraševanju, kar samo poslabša postopek odločanja na podlagi napovedi, izdelanih na podlagi enostavne vrste podatkov. Podatki o prodaji so edini zanesljivi približek podatkov o povpraševanju (Syntetos et al., 2016), čeprav ta poenostavitev ni popolna, zlasti v oskrbovalnih verigah z veliko primeri izpada zalog.

V koraku predhodne analize podatkov je priporočljivo, da analizo podatkov vedno začnete z grafičnimi prikazi in tako odgovorite na naslednja vprašanja. Ali obstajajo dosledni vzorci? Ali obstaja pomemben trend? Ali je opazna sezonskost? Ali obstajajo dokazi o poslovnih ciklih? Kako močne so povezave med spremenljivkami? To so vprašanja, na katera lahko odgovori preprosta grafična ponazoritev, in omogočijo nadaljnjo analizo podatkov z zožitvijo fokusa, katere modele uporabiti za odkrite značilnosti povpraševanja. Običajno lahko tako preprost model premaga bolj zapletene in kompleksne modele (Rostami-Tabar in Mircetic, 2023).

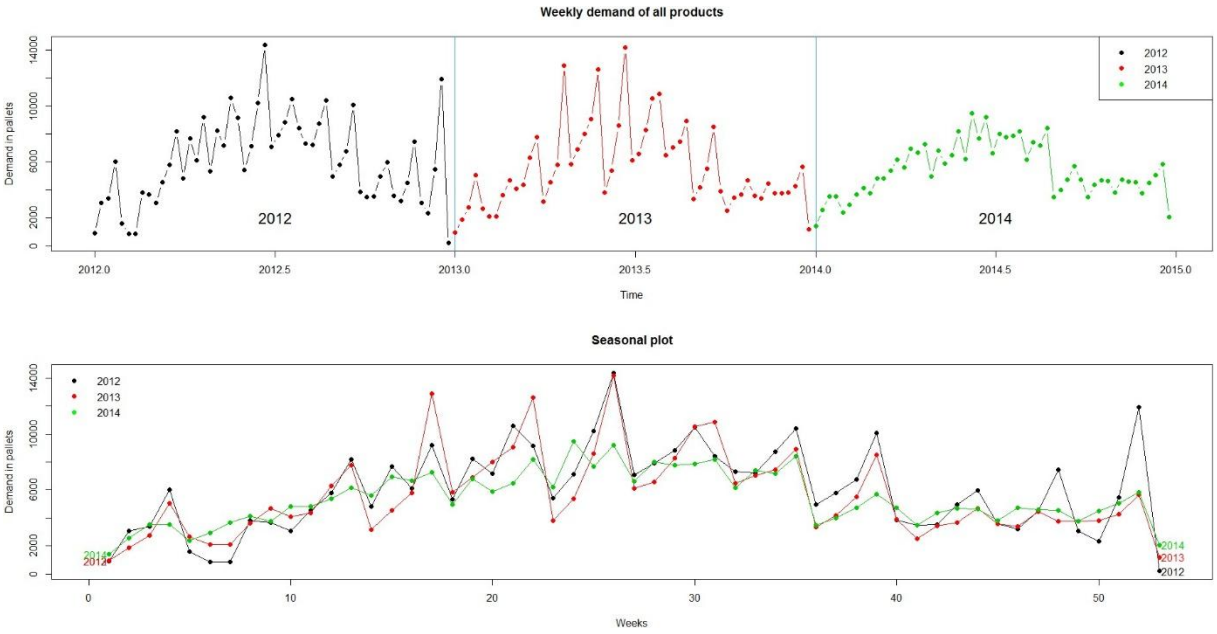
Izbira in oblikovanje modelov napovedovanja je najpomembnejši korak pri oblikovanju modela napovedovanja. Kateri model bomo uporabili, je odvisno od več dejavnikov, med katerimi sta najpomembnejša razpoložljivost preteklih podatkov in korelacija med odvisnimi in neodvisnimi spremenljivkami. Običajno se pri izbiri modela primerja dva ali tri možne modele. Vsak model je umetna konstrukcija, ki temelji na nizu predpostavk (eksplicitnih in implicitnih) in običajno vključuje enega ali več parametrov, ki jih je treba ustvariti z uporabo znanih preteklih podatkov. V naslednjem poglavju bo predstavljen postopek razvoja in uporabe modela ARIMA, ki je eden najboljših in najbolj priljubljenih modelov.

Vrednotenje modela napovedi je korak, ki meri uporabnost ustvarjenega modela. Po izbiri modela napovedovanja in oceni njegovih parametrov se model uporabi za izdelavo napovedi. Natančnost

modela se oceni z uporabo različnih statističnih podatkov, pomembno pa je tudi, da se napovedi preverijo z merami poslovnih implikacij (tj. merami uporabnosti).

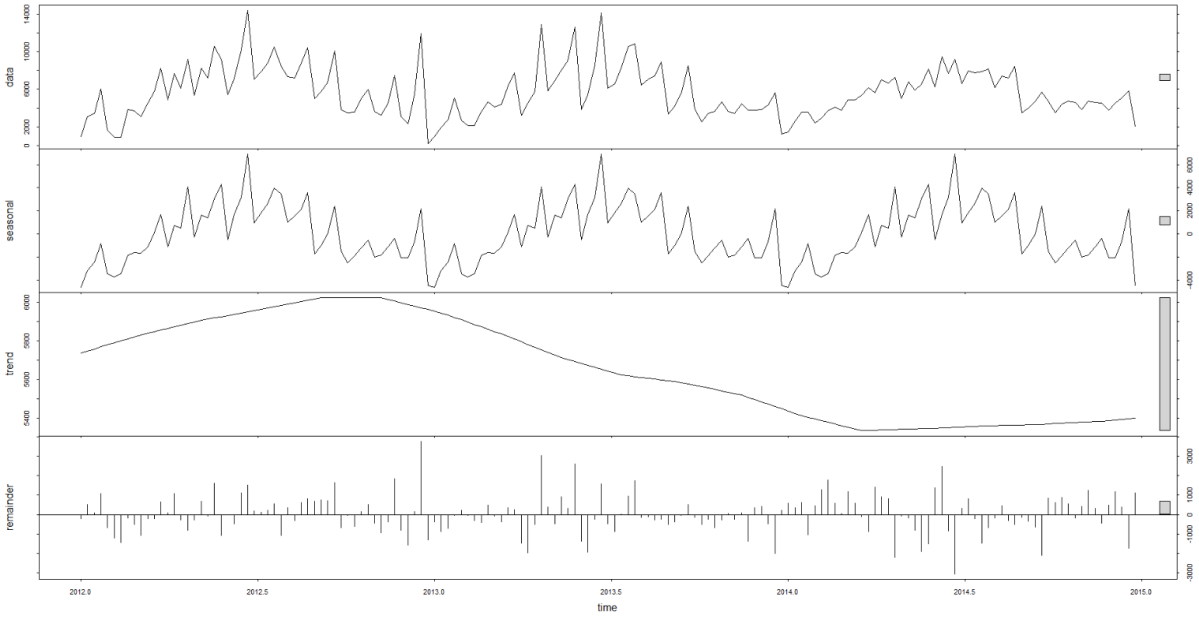
9.3 Napovedovanje povpraševanja v živilski industriji

Podatki o porabi vseh izdelkov opazovanega podjetja iz živilske industrije so predstavljeni kot tedensko povpraševanje od januarja 2012 do decembra 2014 (slika 9.2). Os x predstavlja čas, medtem ko so vrednosti povpraševanja predstavljene na osi y. Ti podatki so prikazani v tedenskih intervalih, saj to ustreza obdobju, v katerem poteka dobava na končna prodajna mesta. Posledično se vodstvo podjetja osredotoča na napovedovanje tedenske porabe na trgu.



Slika 9.2 Podatki o povpraševanju za opazovano živilsko podjetje.

Slika 9.2 prikazuje več pomembnih značilnosti in lastnosti danih podatkov. Prvič, podatki imajo izrazito sezonski značaj, saj je največja prodaja sredi leta (v poletnih mesecih). Drugič, posezonski graf (spodnji graf na sliki 9.2) prikazuje pomembno spremembo vzorca trenda upadanja v letu 2014. To sta zelo pomembni značilnosti za izbiro pravega modela napovedovanja in za višje vodstvene delavce v podjetju, saj razkrivata znaten upad porabe in izgubo trga. Za nadaljnjo raziskavo opaženih značilnosti je uporabljena metodologija razgradnje sezonskega trenda (STL) (slika 9.3). Dekompozicija STL razdeli izvirne vzorce povpraševanja na tri komponente: sezonsko komponento, komponento trenda in komponento preostanka.



Slika 9.3 Razčlenitev STL podatkov o povpraševanju.

STL je pokazala, da imajo opazovane časovne vrste aditivno naravo, kar pomeni, da se nihanja okoli krivulje trenda in cikla s časom ne povečujejo bistveno. Zato za neobdelane časovne vrste ni bilo treba uporabiti Box-Coxove transformacije. Dekompozicija je razkrila, da v opazovanih časovnih vrstah prevladuje sezonska komponenta, ki kaže velika nihanja znotraj enega leta. Glede na omejeno število let opazovanj je težko opredeliti poslovni cikel. Dekompozicija je tudi pokazala, da je trend v seriji minimalen, z upadajočim vzorcem od sredine leta 2013 naprej.

Te ugotovljene značilnosti so predstavljale pomemben vhodni podatek v postopku oblikovanja ustreznega modela napovedovanja. V ta namen je bil izbran model S-ARIMA (Seasonal Autoregressive Integrated Moving Average). Model S-ARIMA je zelo učinkovit za napovedovanje, saj združuje tako avtoregresijsko komponento kot komponento drsečega povprečja ter diferenciranje, da so podatki stacionarni. Ta model je še posebej dober pri zajemanju in modeliranju sezonskih vzorcev v podatkih časovnih vrst, zato je idealen za panoge s cikličnimi vzorci povpraševanja, kot je živilska industrija. Poleg tega zmožnost modela S-ARIMA, da obravnava kompleksne sezonske strukture in trende, omogoča natančnejše in zanesljivejše napovedi, ki so ključne za učinkovito načrtovanje oskrbovalne verige in upravljanje zalog. Strukturna oblika S-ARIMA je predstavljena v enačbi (1).

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

kjer sta $\Phi(z)$ in $\Theta(z)$ polinoma reda *P* oziroma *Q*, od katerih vsak nima korenov znotraj enotskega kroga. B je operator povratnega premika, ki se uporablja za opis procesa diferenciranja, tj. $By = y_{tt-1}$. Če je $c \neq 0$, je v napovedni funkciji impliciran polinom reda $d+D$. Ker je model S-ARIMA zelo parametriziran, je ključno vprašanje pri uporabi modela S-ARIMA izbira ustreznega reda modela, kar vključuje določitev vrednosti *p, q, P, Q, D* in *d*. Če sta *d* in *D* znana, lahko rede *p, q, P* in *Q* izberemo z uporabo informacijskega merila, kot je Akaikejev informacijski kriterij (AIC) ali Bayesov informacijski kriterij (BIC). Formuli za AIC in BIC sta naslednji:

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

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

kjer je $k=p+q+P+Q+1$, če je vključen konstantni člen, in 0 v nasprotnem primeru, *L* je maksimizirana verjetnost modela, prilagojenega diferenciranim podatkom, *SSE* je vsota kvadratnih napak, *N* je število opazovanj, uporabljenih za oceno, in *k* je število napovednikov v modelu.

Hyndman in Khandakar (2007) sta za določitev optimalnega nabora parametrov predlagala Canova-Hansenov in KPSS test korena enote v naslednjih korakih:

- Uporabite test Canova-Hansen za določitev *D* v okviru ARIMA.
- Izberite *d* z uporabo zaporednega KPSS testa enotnega korena na sezonsko diferenciranih podatkih (če je $D = 1$) ali na izvirnih podatkih (če je $D = 0$).
- Izberite optimalne vrednosti *p, q, P* in *Q* z zmanjšanjem AIC.

9.4 Razvoj modela napovedovanja S-ARIMA

Razvoj modela napovedovanja S-ARIMA

V skladu z zgoraj navedenim postopkom je bilo preizkušenih več nastavitev parametrov, njihova učinkovitost pa je predstavljena v preglednici 9.1. Za preizkušanje učinkovitosti različnih modelov je uporabljenih več meril: povprečna absolutna odstotna napaka (MAPE), povprečna kvadratna napaka (RMSE), povprečna absolutna skalirana napaka (MASE), AIC in BIC (enačbi 2 in 3).

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

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

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

kjer *so* e_i rezidualne vrednosti in . $q_j = \dfrac{e_j}{\dfrac{1}{T-m}\sum_{t=m+1}^{T}\left|y_t - y_{t-m}\right|}$

Najbolj priljubljeni meri za vodje v oskrbovalnih verigah sta RMSE in MAPE, saj zagotavljata "občutek", kako dobro deluje model v realnih številkah (RMSE) in odstotkih (MAPE).

Preglednica 9.1 Uspešnost modelov S-ARIMA z različnimi nastavitvami parametrov.

Modeli [a]	RMSE	MAPE	MASE	AIC	BIC
S-ARIMA $(5,0,1)(1,0,0)_{52}$[b]	1023	14.18 %	0.60 %	86.62	110.50
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	22.12 %	1.05 %	41.74	53.560
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	40.460

[a] Napake modela se izračunajo na testnem nizu podatkov.

[b] Podrobnosti o modelu S-ARIMA $(5,0,1)(1,0,0)_{52}$ so navedene v nadaljevanju.

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

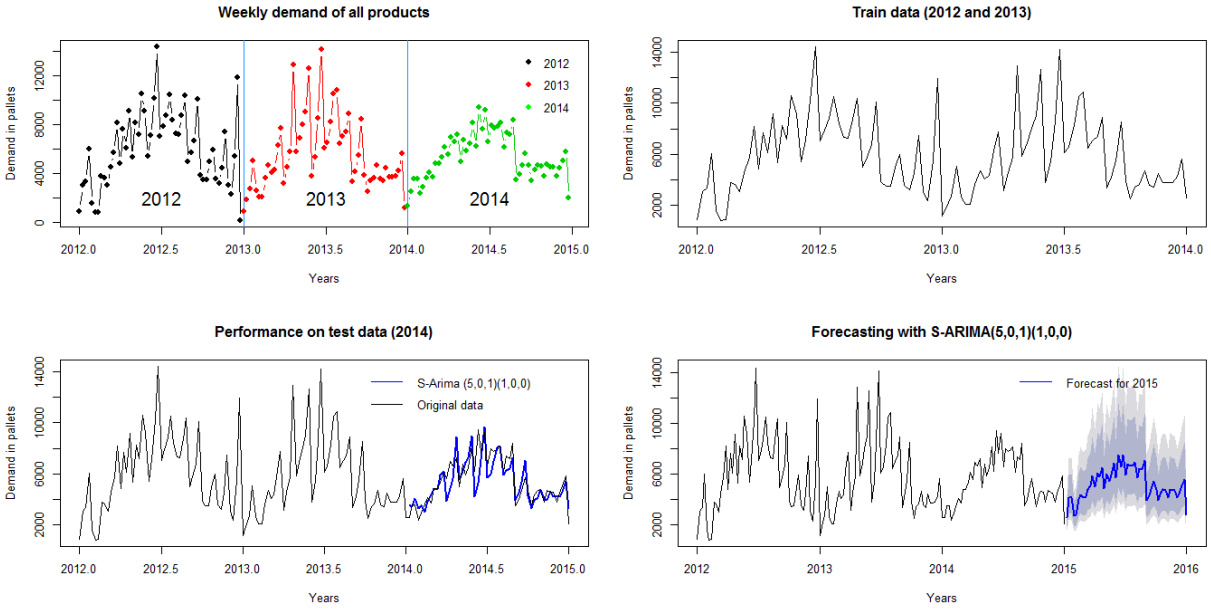
Iz preglednice 9.1 je razvidno, da je model S-ARIMA $(5,0,1)(1,0,0)_{52}$ presegel konkurenčne modele, saj je dosegel najnižje napake RMSE, MAPE in MASE. Oblika modela S-ARIMA $(5,0,1)(1,0,0)_{52}$ je predstavljena v enačbi (4), kjer je $\phi_1 = -\ 0{,}5421$, ϕ_2 = 0,2962, $\phi_3 = -\ 0{,}099$, ϕ_4 = 0,3974, ϕ_5 = 0,4994, Φ_1 = 0,9558, *c* = 8,523 in Θ_1 = 0,*6345*.

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

Podobno uspešnost sta imela modela S-ARIMA $(4,0,0)(1,0,0)_{52}$ in S-ARIMA $(4,0,1)(1,0,0)_{52}$. Pri izvajanju napovedi model S-ARIMA $(5,0,1)(1,0,0)_{52}$ v povprečju povzroči napako 1023 izdelkov, kar pomeni 14,18 %. Ti rezultati poudarjajo pomen skrbne izbire izrazov, ki jih je treba vključiti v model S-ARIMA, saj med uspešnostjo modela z različnimi parametri ni bistvenih razlik. Ocena je pokazala, da so bili modeli, ki so vključevali avtoregresijske (*p, P)* in drseče povprečne komponente (*q, Q*), učinkovitejši pri napovedovanju porabe pijač kot modeli, ki so vključevali sezonsko ali nesezonsko diferenciranje. Poleg tega je primerjalni pregled razkril nekaj nepričakovanih ugotovitev, na primer, da so modeli, ki vključujejo sezonsko diferenciranje S-ARIMA $(4,0,0)(0,1,1)_{52}$ in S-ARIMA $(0,0,1)(0,1,0)_{52}$, slabši od preprostega povprečnega naivnega modela napovedi, kar se kaže v tem, da so njihove napake MASE presegle ena.

9.5 Napovedi prihodnjega povpraševanja

Slika 9.4 prikazuje uspešnost metode S-ARIMA $(5,0,1)(1,0,0)_{52}$. . V zgornji levi plošči so podatki o začetnem povpraševanju, ki so obarvani zaradi lažjega razlikovanja različnih tržnih let. Zgornja desna plošča je vhodni podatek za S-ARIMA $(5,0,1)(1,0,0)_{52}$, ki predstavlja učne podatke, na podlagi katerih se parametri v S-ARIMA določijo po postopku, opisanem v podpoglavju 9.3. To je zelo pomembno, da razumemo, kako težko je delo modela za napovedovanje, saj ima v tem primeru kot vhodne podatke dve leti, napovedati pa mora prihodnje povpraševanje za eno leto vnaprej! To je precej pogost scenarij v oskrbovalnih verigah in logistiki, saj velja nenapisano pravilo, da podjetja hranijo zgodovino svojih podatkov tri leta, nato pa podatke zavržejo.



Slika 9.4 Usposabljanje, testiranje in napovedano povpraševanje modela S-ARIMA $(5,0,1)(1,0,0)_{52}$.

Spodnja leva plošča na sliki 9.4 prikazuje uspešnost opazovanega modela. Uspešnost modela je že predstavljena v preglednici 9.1 z različnimi statističnimi merami, vendar je za upravljavce običajno težko dobiti občutek, kako dober ali slab je model. V ta namen je na plošči grafično prikazana uspešnost modela. Lahko trdimo, da model precej dobro sledi testnim podatkom in v večini obdobij kaže odlično uspešnost. Za izdelavo prihodnjih napovedi je bil model S-ARIMA $(5,0,1)(1,0,0)_{52}$ ponovno oblikovan tako, da so bili podatkom iz usposabljanja dodani podatki iz leta 2014. Po tem modelu so bile izdelane 52-tedenske napovedi za leto 2015. Napovedi za leto 2015 so predstavljene v spodnji levi plošči. Napovedim sta priložena 80- in 95-odstotna napovedna intervala, ki kažeta na morebitno razpršenost prihodnjih napovedi od povprečnih napovedanih vrednosti. Model napoveduje nadaljnje zmanjševanje povpraševanja, ki se je začelo leta 2014. Vzroki za ta upad bi lahko bili različni, zato bi jih morali menedžerji na strateški ravni podjetja dodatno raziskati.

Literatura poglavje 9

- 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. Umetna inteligenca in strojno učenje v oskrbovalnih verigah

Kaj je umetna inteligenca (AI)? Ali je to res "živo" bitje, ki je sposobno razmišljati in sprejemati lastne odločitve na podlagi svojega uma, preteklih izkušenj, etike, prepričanj itd.? Kako je povezana s strojnim učenjem (ML)? Ali sta umetna inteligenca in računalniško učenje ena in ista stvar?

Katera orodja se uporabljajo pri umetni inteligenci in ML? Kakšno vlogo imata AI in ML v poslovnem kontekstu in kako ju je mogoče uporabiti pri vsakodnevnih poslovnih operacijah in optimizacijah? Ali obstajajo posebne arhitekture in primeri uporabe UI in ML v oskrbovalnih verigah?

Na ta in podobna vprašanja bomo poskušali odgovoriti v naslednjem poglavju, ki ga bomo zaključili s primerom študije primera uporabe algoritmov AI in ML v distribucijskem skladišču.

10.1 Kaj je umetna inteligenca?

Področje umetne inteligence se je začelo v petdesetih letih prejšnjega stoletja, ko so se računalniški znanstveniki spraševali: "Ali lahko računalniki razmišljajo kot ljudje?" Takratni raziskovalci so bili navdušeni nad možnostjo, da bi računalnike naučili opravljati zapletene naloge, in so v ta namen razvili vrsto različnih algoritmov. Opredelitev področja bi lahko navedli kot prizadevanje za avtomatizacijo intelektualnih nalog, ki jih običajno opravljajo ljudje (Chollet, 2021). Najznamenitejši algoritmi na področju umetne inteligence danes izhajajo iz ML in globokega učenja, ki sta podskupini umetne inteligence. Poleg ML in globokega učenja AI vključuje tudi veliko algoritmov, ki se ne učijo. Poleg tega lahko trdimo, da so te vrste algoritmov bolj prevladovale v zgodnji fazi razvoja UI. V skladu s tem je to del UI, znan kot simbolna UI, ki temelji na ideji, da je mogoče človeško raven zmogljivosti in inteligence doseči s programiranjem računalnikov z velikim naborom eksplicitnih pravil za reševanje opazovanega problema. Ta pristop je zagotavljal odlične rezultate pri logičnih problemih, ki so bili dobro opredeljeni, kot je na primer računalnik, ki igra šahovsko partijo, vendar se je pri bolj zapletenih problemih izkazal za zapletenega. Izkazalo se je, da je resnični svet veliko bolj zapleten, kot bi lahko v računalnik vstavili vsa eksplicitna programska pravila. Ta pristop je osredotočen na zamisel za določeno situacijo naredi to ali ono (pravila if-then). To je lahko razumljiv pristop, vendar po drugi strani zelo zamuden in včasih zelo težaven za določitev vseh možnih scenarijev, ki jih je treba vstaviti v program. Kot smešen primer, vendar

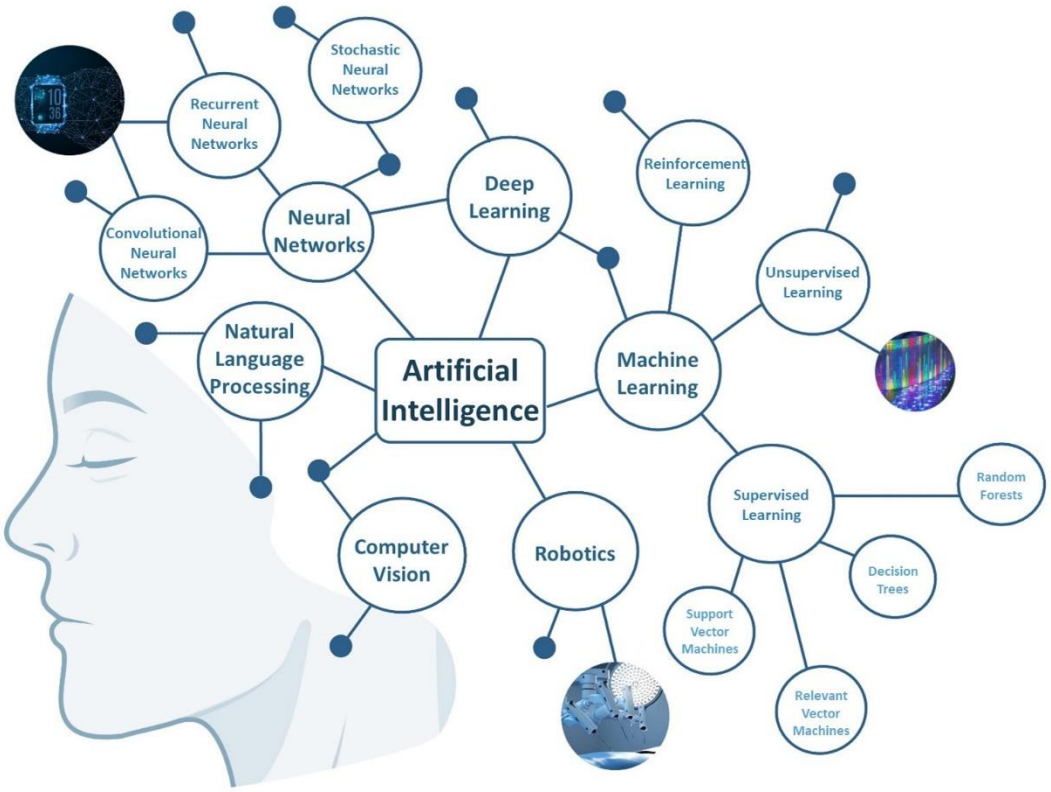
dobro ponazoritev določene teme, si oglejte sliko 10.1, ki prikazuje več scenarijev za določanje napovedi na podlagi stanja kamna.

Slika 10.1 Programska pravila If - Then (Gibbs, M. 2019).

Področje simbolne umetne inteligence je bilo najbolj priljubljeno v osemdesetih letih prejšnjega stoletja s pojavom ekspertnih sistemov (ES). ES predstavljajo podmnožico sistemov za podporo odločanju (DSS) (Turban, 1998), ki se osredotočajo na zagotavljanje računalniških zmožnosti odločanja, podobnih tistim, ki jih imajo človeški strokovnjaki na določenem področju. Ti sistemi so zasnovani za reševanje zapletenih problemov z uporabo vrste pravil ali algoritmov, ki simulirajo procese človeškega razmišljanja. Olson in Courtney opisujeta ekspertne sisteme (ES) kot računalniške programe, ki simulirajo človeške miselne procese za sprejemanje odločitev na določenem področju in vključujejo določeno stopnjo umetne inteligence, da bi ustrezali sklepom, do katerih bi prišel človeški strokovnjak (Olson in Courtney, 1992). Komponenta ES je še posebej uporabna za podporo odločevalcem na področjih, ki zahtevajo specializirano znanje (Turban et al., 2005). V bistvu ES zajame strokovno znanje človeškega strokovnjaka (ali drugega vira) in ga prenese na računalnik. Ta tehnologija lahko pomaga nosilcem odločanja ali pa jih v celoti nadomesti, zaradi česar je ena od najbolj razširjenih in komercialno uspešnih oblik umetne inteligence (Turban et al., 2005). Eden od ključnih razlogov za razvoj ES je razširjanje strokovnega znanja širšemu občinstvu (Jackson, 1999). V naslednjih podpoglavjih bomo prikazali uporabo ES, ki temelji na algoritmih AI & ML, v osrednjem skladišču kot delu celotnega DSS za upravljavce.

Danes področje umetne inteligence sestavljajo različni pristopi in algoritmi, vendar so najbolj uporabljeni opisani na sliki 10.2. Odmaknilo se je od sistemov ES, za ponovni vzpon področja pa so najbolj zaslužni algoritmi globokega učenja, ki so bili v zadnjih 12 letih zelo uspešni pri problemih prepoznavanja slik, govora, segmentacije slik, prepoznavanja obrazov itd. Globoko učenje

uporablja več plasti abstrakcije za prepoznavanje zapletenih vzorcev v visokodimenzionalnih podatkih. Ta pristop je dosegel pomemben napredek na področjih, kot so prepoznavanje govora in slik, odkrivanje zdravil in obdelava naravnega jezika. Globoko učenje ima sposobnost samodejnega odkrivanja ustreznih lastnosti in zmanjšuje potrebo po človeškem posredovanju pri oblikovanju lastnosti, zaradi česar je zelo učinkovito pri izkoriščanju velikih zbirk podatkov in računalniške moči (LeCun, Bengio in Hinton, 2015).



Slika 10.2 Glavna področja in podpodročja umetne inteligence (Athanasopoulou et al., 2022).

Velik korak k današnjemu stanju je bila raziskava klasifikacije številk, ki so jo opravili Hinton, Osindero in Teh (2006), ki jim je uspelo doseči več kot 98-odstotno natančnost pri klasifikaciji podatkovne baze MNIST (Modified National Institute of Standards and Technology). Eden od načinov razmišljanja o tem, kako je umetna inteligenca prešla iz simbolne v ML in kaj je bistvo ML, je, da si algoritme ML predstavljamo kot amorfno maso, ki se oblikuje glede na želene rezultate. Sistem pravil, ki od vhoda do izhoda sprejema podatke, se spreminja od problema do problema in se prilagaja obstoječim razmeram. Njegov cilj je poiskati pravila, ki bodo avtomatizirala nalogo z iskanjem statističnih vzorcev v podatkih. Zaradi tovrstnega pristopa k reševanju različnih problemov se je bistveno skrajšal čas za vzpostavitev sistema (v primerjavi s pravili če - potem), poleg tega pa je bolj univerzalen pristop za reševanje različnih problemov.

Bengio, Lecun in Hinton (2021) so poudarili, da je prihodnost umetne inteligence v globokem učenju ter revolucionarnem vplivu mehke pozornosti in transformatorskih arhitektur na umetno inteligenco. Te inovacije nevronskim mrežam omogočajo, da se dinamično osredotočajo na pomembne vhodne podatke in shranjujejo informacije v diferencialne pomnilnike, kar bistveno izboljša zaporedno obdelavo.

10.2 Kakšen je ekosistem umetne inteligence in računalništva na podlagi podatkov?

Ekosistem algoritmov AI in ML sestavljajo trije ključni stebri:

- Vhodni podatki;
- Izhodni podatki;
- Stroškovna funkcija.

Vhodni podatki so zapisi podatkov o določeni lastnosti ali lastnostih (glede na opazovano težavo). Natančnost vhodnih podatkov je ključnega pomena za izdelavo natančnih algoritmov. V resničnih aplikacijah pogosto ni tako, zato se običajno veliko časa in truda nameni zbiranju, čiščenju, urejanju, poenotenju, preverjanju napačnih vhodnih podatkov itd. Poleg natančnih podatkov je še en pomemben vidik značilnosti vhodnih podatkov njihova predstavitev in kodiranje. Različni načini kodiranja podatkov lahko razkrijejo različne značilnosti podatkov in znatno "pomagajo" modelom ML pri razkrivanju skritih vzorcev v podatkih. Tu vidimo Ahilovo peto algoritmov ML. Pogosto se preveč pozornosti posveča oblikovanju metod za pridobivanje informacij in inteligence iz podatkov (tj. samim algoritmom), premalo pozornosti pa se posveča vhodnim podatkom in njihovi povezavi z izhodnimi podatki. Na splošno se šteje za samoumevno, da so vhodni podatki vzročno povezani z izhodnimi podatki, kar pa včasih sploh ni tako. Zato mora biti naslednji velik korak pri razvoju ML iskanje boljših načinov za zbiranje, predstavitev in kodiranje podatkov.

Izhodni podatki predstavljajo meritve določenega problema, ki ga poskušamo rešiti. Pri klasifikacijskem problemu so to oznake razredov. Pri regresiji bo to realno število, ki ga poskušamo napovedati. V kontekstu specifičnega problema, pri problemih prepoznavanja govora, so lahko izhodni podatki človeško ustvarjeni prepisi zvočnih datotek. Pri težavah s prepoznavanjem slik so lahko izhodni podatki oznake razredov slik itd.

Stroškovna funkcija predstavlja način merjenja uspešnosti umetne inteligence in ML. Načeloma bi si želeli, da bi se odgovori algoritmov ujemali z izhodnimi podatki za dane vhodne podatke.

Stroškovna funkcija je tudi povratni signal za nabor parametrov, ki usmerjajo delo algoritma, tj. omogoča optimizacijo celotnega delovanja algoritma prek procesa učenja (iskanje optimalnega nabora parametrov). Proces učenja običajno vključuje nadzorovano učenje, pri katerem se model uči na označenih podatkih, da se s tehnikami, kot sta stohastično spuščanje po naklonu in povratno širjenje, čim bolj zmanjšajo napake pri napovedovanju. To modelu omogoča, da učinkovito prilagodi svoje notranje parametre, kar vodi do izboljšane učinkovitosti pri nalogah, kot sta odkrivanje in razvrščanje predmetov (LeCun, Bengio in Hinton, 2015).

10.3 Katera orodja se uporabljajo pri ML?

Na splošno lahko algoritme ML razdelimo v dve glavni kategoriji: nadzorovano in nenadzorovano učenje.

Pri nadzorovanem učenju se algoritmi usposabljajo na naboru označenih podatkov, pri čemer se vsaka vhodna podatkovna točka poveže s pravilnim izhodnim podatkom. Ta jasna "slika" o tem, kakšen naj bi bil pravilen odgovor za določen vhodni podatek, omogoča algoritmu, da se nauči funkcije preslikave med vhodnimi podatki in izhodnimi podatki. V skladu s tem so tako vhodni kot izhodni podatki znani (Athanasopoulou et al., 2022). Pogoste aplikacije nadzorovanega učenja vključujejo naloge razvrščanja (npr. določanje, ali je elektronsko sporočilo neželena pošta ali ne) in naloge regresije (npr. napovedovanje cen hiš na podlagi različnih značilnosti). Nekateri izmed najbolj priljubljenih algoritmov, ki so se izkazali s številnimi aplikacijami, so posplošeni aditivni modeli, naključni gozdovi, boosting, klasifikacijska in regresijska drevesa, podporni vektorski stroji, razširjena linearna regresija, logistična regresija, k-najbližji sosedje, linearna diskriminantna analiza, lasso, nevronske mreže, prilagodljivi nevrofuzzy inferenčni sistem itd. (Rostami-Tabar & Mircetic, 2023). Nadzorovano učenje je učinkovito, saj za doseganje visoke natančnosti napovedi uporablja podatke, ki jih je ustvaril človek. Vendar je njegova učinkovitost močno odvisna od kakovosti in količine označenih podatkov.

Nasprotno pa nenadzorovano učenje obravnava nabore podatkov, ki nimajo označenih odgovorov. Zato algoritmi strojnega učenja brez nadzora uporabljajo neoznačene podatkovne nize, ki vključujejo samo vhodne podatke (Athanasopoulou et al., 2022). Pri tem so algoritmu na voljo samo vhodni podatki, njegov cilj pa je poiskati osnovne vzorce, strukture ali odnose v podatkih. Pogoste tehnike nenadzorovanega učenja vključujejo grozdenje (npr. združevanje strank po nakupnem vedenju) in zmanjševanje razsežnosti (npr. zmanjševanje števila spremenljivk v podatkovnem nizu, pri čemer se ohranijo pomembne informacije). Nenadzorovano učenje je

koristno za raziskovalno analizo podatkov in odkrivanje skritih struktur v podatkih. Pogosto se uporablja, kadar je označenih podatkov malo ali niso na voljo.

Druga pomembna kategorija, ki se razlikuje od nadzorovanega in nenadzorovanega učenja, je učenje z okrepitvami. Pri tem se algoritem uči z interakcijo z okoljem in prejemanjem povratnih informacij v obliki nagrad ali kazni. Tak pristop poskusov in napak omogoča algoritmu učenje optimalnih dejanj za maksimiranje kumulativnih nagrad. Učenje z okrepitvami se pogosto uporablja na področjih, kot so robotika, igranje iger in avtonomni sistemi.

Eden izmed najbolj priljubljenih in uspešnih algoritmov za ML izhaja iz veje nevronskih mrež. Nevronske mreže obstajajo že od petdesetih let prejšnjega stoletja, vendar so postale priljubljene v osemdesetih letih prejšnjega stoletja in v zadnjih 12 letih. Zgrajene so na približku bioloških nevronov in načinu, kako si ti v možganih izmenjujejo delčke informacij, vendar razen tega med njimi ni pomembnih povezav. Danes se najpogosteje uporabljajo nevronske mreže v obliki globokega učenja, ki predstavljajo več skupkov skritih plasti med vhodnimi in izhodnimi značilnostmi, ki izvajajo več nelinearnih transformacij vhodnih značilnosti. Ker se je to izkazalo za zelo uspešno, je globoko učenje danes eno od najpomembnejših podpodročij ML (Chollet, 2021).

10.4 Študija primera?

Ugotovitve Wenzla, Smita in Sardesaija (2019) o ML pri upravljanju oskrbovalne verige kažejo na vse večjo integracijo aplikacij ML v različne naloge oskrbovalne verige. Skladno s tem opazovana študija primera predstavlja uporabo AI & ML, ki se izvaja v osrednjem skladišču tovarne hrane (Mirčetić et al., 2016; Mircetic et al., 2014). V tovarniškem kompleksu je 30 viličarjev. Viličarji so vključeni v različne operacije znotraj kompleksa, ki so ključne za logistične operacije v proizvodnji, skladiščenju in odpremi izdelkov. Osrednje skladišče ima zmogljivost 11 100 paletnih mest in letno proizvodnjo od 300 000 do 350 000 palet. Trenutno tovarna oskrbuje približno 20 000 supermarketov z neposredno dostavo.

Problem vključevanja viličarjev je povezan z dejstvom, da prevelika ali premajhna vključenost viličarjev v različne tovarniške procese vodi v velike finančne in tržne izgube. Trenutno proces odločanja o tem, kje in kaj bo delal posamezen viličar, temelji na odločitvah strokovnjakov (vodij). Odločitve strokovnjakov temeljijo na njihovih izkušnjah brez pomoči sistema za podporo odločanju (DSS). Številni empirični dokazi kažejo, da človekova intuitivna presoja in odločanje pogosto nista optimalna, zlasti v pogojih kompleksnosti in stresa (Druzdzel in Flynn, 2002). To poudarja pomen vključevanja sistemov za podporo odločanju (DSS), ki strokovnjakom pomagajo pri odločanju.

V tej aplikaciji smo izbrali več algoritmov ML za pomoč pri optimizaciji delovanja nakladalnega skladišča. Algoritmi ML so zbrani v edinstvenem odločitvenem okviru, ki služi kot DSS za menedžerje in strokovnjake v določenem podjetju. Poleg tega lahko celoten DSS za odločanje opazujemo kot platformo umetne inteligence, saj nenehno preračunava predloge iz več modelov ML (koliko viličarjev uporabiti in katere) ter samodejno izbere najboljše glede na zagotovljene vhodne podatke operaterjev.

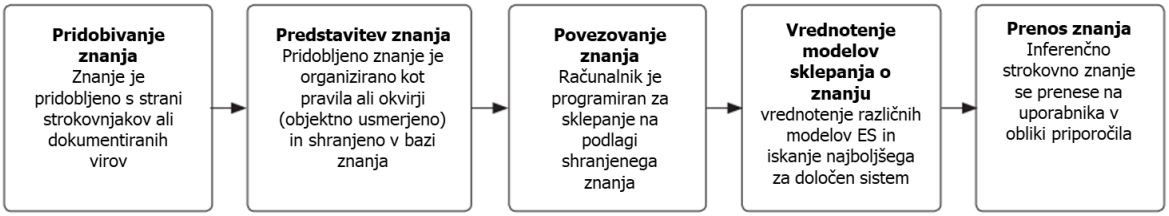
Opis težave

Postopek natovarjanja je ključnega pomena za logistiko v skladišču, saj vpliva na raven tržnih storitev. Med pošiljkami skladiščni strokovnjak določi število in izbor viličarjev za nakladanje, pri čemer ga vodijo trije dejavniki: (1) dokončanje nakladanja v določenem časovnem okviru, (2) zmanjšanje motenj pri drugih nalogah viličarjev in (3) uskladitev uporabe viličarjev z zmogljivostmi vzdrževanja, ki lahko hkrati opravijo dva generalna popravila. Vsak viličar opravi štiri do pet vzdrževalnih pregledov na leto.

Viličarji so bistveni za nakladanje, ki mora podpirati tržno strategijo podjetja in hkrati zagotavljati nemoteno izvajanje drugih dejavnosti. Napačna razporeditev viličarjev lahko privede do nezadostne uporabe virov ali škoduje ugledu podjetja in ravni storitev. Zamude pri nakladanju povzročajo kazni. Vodja mora uskladiti uporabo viličarjev v vseh dejavnostih, da bi se izognil hkratnim remontom in obvladal različne potrebe po vzdrževanju. Čeprav menedžerji običajno sprejemajo natančne odločitve, lahko v okoljih z visoko stopnjo stresa pride do napak. Zato je potreben DSS, ki bo povečal zanesljivost in zanesljivost odločanja.

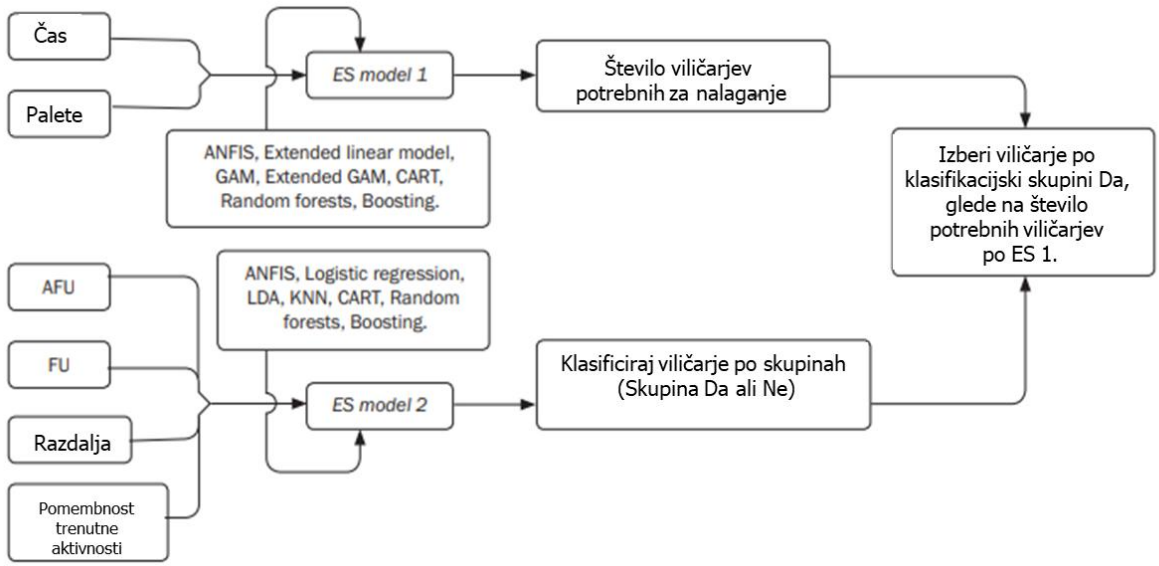
AI in ML kot DSS za centralno skladišče

Prvi korak pri ustvarjanju sistemov AI in ML je pridobitev stabilnega in pravilnega vira znanja (podatkovne zbirke) ter osvetlitev poslovnih vlog, ki jih mora podpirati. Zato je na sliki 10.3 predstavljena metodologija za izgradnjo sistema DSS AI & ML.



Slika 10.3 Metodološki koraki za izgradnjo sistema AI & ML DSS (Rainer & Turban, 2008; Turban, Aronson, & Liang, 2005).

Znanje je bilo pridobljeno z razgovori z vodji, opazovanjem njihovih postopkov odločanja in pregledom skladiščnih evidenc. Za razvoj DSS za dani problem sta bili vzpostavljeni dve bazi znanja. Prva baza znanja vključuje odločitve o številu viličarjev, razporejenih na območju nakladanja (434 strokovnih odločitev), druga pa zajema, kateri viličarji so bili uporabljeni (368 strokovnih odločitev) v različnih operativnih scenarijih. V fazi sklepanja o znanju je bilo uporabljenih več algoritmov ML z uporabo programske opreme Matlab: Prilagodljiv nevrofuzzy inferenčni sistem (ANFIS), posplošeni aditivni modeli (GAM), naključni gozdovi, pospeševanje, klasifikacijska in regresijska drevesa (CART), razširjena linearna regresija, logistična regresija, k-najbližji sosedje (KNN) in linearna diskriminantna analiza (LDA). Ocenjeni so bili različni ML modeli in določeni so bili tisti z najboljšo učinkovitostjo. ANFIS in CART sta pokazala najboljše rezultate in bila izbrana kot končna DSS za praktično uporabo v podjetju. Prenos znanja je bil olajšan z uporabniškim vmesnikom končnih modelov DSS. Struktura in logika DSS je prikazana na sliki 10.4.



Slika 10.4 Gradbena struktura skladiščnega DSS, ki temelji na algoritmih AI in ML.

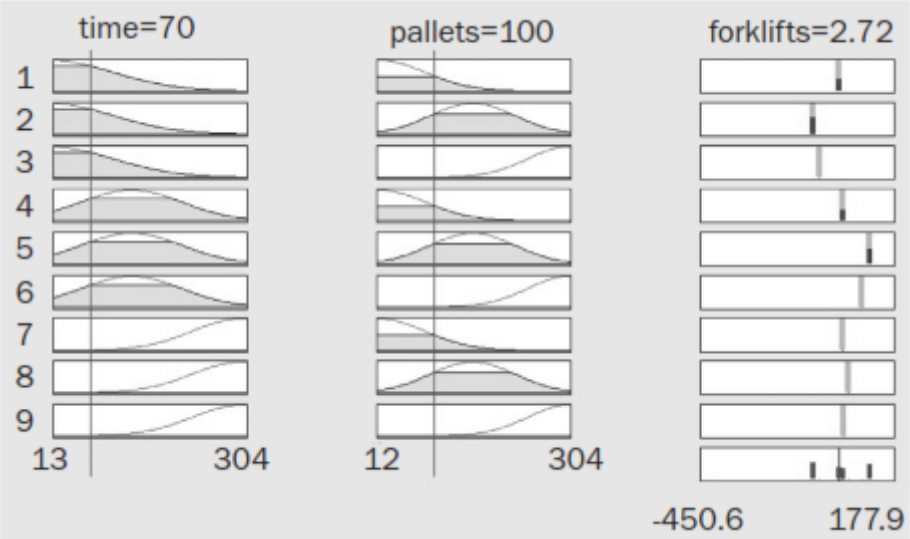
Okvir DSS je sestavljen iz vhodne plasti, ki jo sestavlja več ključnih dejavnikov, ki vplivajo na uporabo viličarjev. Sloj ML ima modele ML, ki preračunavajo predlog, koliko in katere viličarje je treba uporabiti v danem vhodnem scenariju. Najuspešnejši modeli so izbrani kot modeli ekspertnega sistema (ES modeli), saj je baza znanja, na podlagi katere so ustvarjeni ML modeli, pridobljena od strokovnjakov. Prvi model se osredotoča na določanje števila viličarjev, potrebnih na nakladalnem območju (model ES 1). Drugi model obravnava problem izbire, katere posebne viličarje je treba vključiti (model ES 2). Oba modela sta razvita z uporabo tehnik nadzorovanega strojnega učenja. Po navedbah Turbana, Aronsona in Lianga (2005) je strojno učenje pokazalo

odlične rezultate pri oblikovanju inteligentnih sistemov za podporo odločanju (DSS). Modela ES pošiljata signale (predloge in predloge ML) naprej v operacijo razvrščanja, pri čemer se lahko vsak viličar, ki je razvrščen v skupino razvrščanja "Da", vključi v določeno operacijo nakladanja.

Dejavniki, ki so vplivali na odločitve vodje, so bili ugotovljeni s posvetovanji. Pri določanju števila viličarjev, ki jih je treba namestiti na območju nakladanja, sta ključna dejavnika določen čas nakladanja (15 do 135 minut) in količina tovora (15 do 225 palet). Pri izbiri viličarjev, ki jih bo vključil, vodja upošteva pomembnost trenutne dejavnosti (v skladu s politiko podjetja je ocenjena od 1 do 9), stopnjo izkoriščenosti viličarja, njegovo oddaljenost od nakladalne rampe in povprečno stopnjo izkoriščenosti vseh viličarjev. Vsak viličar ima določeno število delovnih ur, preden je potreben remont, njegova uporaba pa je nad to mejo omejena. Izkoriščenost viličarja (FU) je odstotek delovnih ur, ki jih izkoristi posamezen viličar, medtem ko je povprečna izkoriščenost viličarja (AFU) povprečna izkoriščenost delovnih ur vseh viličarjev. Višja vrednost AFU pomeni, da bo večina viličarjev kmalu potrebovala remont.

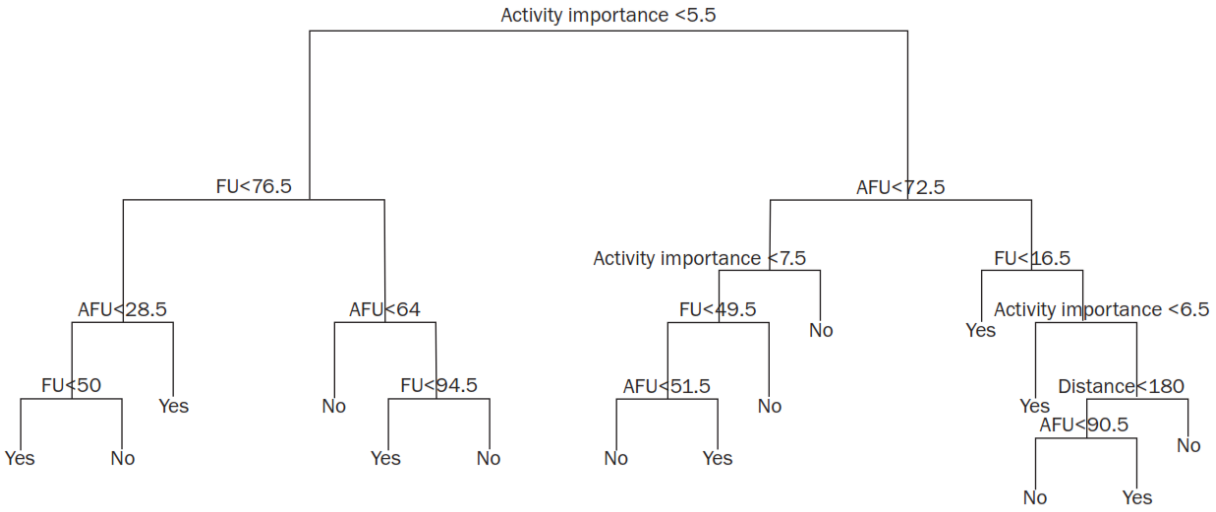
Uporabniški vmesnik sistema DSS

V večini vhodnih situacij sta se najbolje izkazala ANFIS in CART. Zato sta bila izbrana kot motorja danega DSS in njegovih ES. Uporabniški vmesnik modela ES 1 je predstavljen na sliki 10.5 in operaterjem omogoča hitro in enostavno odločanje o številu viličarjev, ki jih je treba uporabiti, tako da preprosto premaknejo navpično črto skozi področje vhodnih spremenljivk na podlagi določenega časa nakladanja in količine tovora.



Slika 10.5 Fuzzy inferenčni sistem modela ES 1.

Model ES 2 služi kot dopolnilno orodje k modelu ES 1, saj izboljša sprejemanje odločitev z zagotavljanjem informacij o tem, ali je treba na nakladalnem območju uporabiti določen viličar (slika 10.6). Z upoštevanjem položaja viličarja (oddaljenost od območja nakladanja), njegove trenutne dejavnosti (pomembnost dejavnosti), izkoriščenosti delovnih ur (FU) in povprečne izkoriščenosti vseh viličarjev (AFU) lahko uporabniki enostavno ugotovijo, ali je določen viličar primeren za nakladanje ali je treba izbrati drugega. Odločitveno drevo CART je preprosto za razlago, saj ni treba vnašati vrednosti v programsko opremo. Namesto tega lahko drevo s slike 10.6 natisnete in ga za hitro sklicevanje postavite na vidno mesto v skladišču.



Slika 10.6 Odločitveno drevo modela ES 2 v zvezi z vključitvijo viličarja.

Upravljavci lahko vsakodnevno uporabljajo predstavljeni sistem DSS, kar pripomore k večji odzivnosti oskrbovalne verige na zahteve strank in zagotavlja visoko verjetnost pravočasne dobave. Predlagani DSS z umetno inteligenco in ML je pokazal uspešne rezultate pri pridobivanju znanja "know-how" strokovnjakov in zajemanju njihove "logike sklepanja". Z uporabo te metode je mogoče pridobiti strokovno znanje vodij in ga uporabiti pri drugih skladiščnih operacijah. To je še posebej dragoceno za praktike, saj je najemanje skladiščnih strokovnjakov pogosto drago. Poleg tega lahko DSS služi tudi kot orodje za usposabljanje vodij začetnikov, saj jim pomaga pridobivati izkušnje in sčasoma izboljšati njihove sposobnosti odločanja. Zato so sistemi umetne inteligence in ML, ki lahko simulirajo odločitve vodje, bistvena orodja, ki omogočajo znatne prihranke pri stroških in večjo učinkovitost skladiščnih operacij.

Literatura poglavje 10

- 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

SEZNAM ŠTEVILKE

SEZNAM TABEL

Sanja Bojić, PhD

Dr. Sanja Bojić je izredna profesorica na Fakulteti tehničnih znanosti Univerze v Novem Sadu, kjer deluje na Katedri za strojegradnjo, transportne sisteme in logistiko. Njeno raziskovalno delo obsega transportno logistiko, intralogistiko, lokacijske probleme, optimizacijo oskrbovalnih verig ter uporabo simulacij in genetskih algoritmov v logističnem načrtovanju. Aktivno sodeluje pri nacionalnih in mednarodnih projektih, kot so Interreg projekt "DANUrB+", "Danube DNA" in pobude za trajnostno logistiko v pristaniščih. Rezultate svojega dela redno objavlja v uglednih znanstvenih revijah, kot so *Journal of Cleaner Production*, *Renewable and Sustainable Energy Reviews* ter *Transport Policy*, s poudarkom na trajnostnih rešitvah v logistiki in energetski učinkovitosti.

ORCID: 0000-0003-0030-4525

Kristijan Brglez

Kristijan Brglez je asistent na Fakulteti za logistiko Univerze v Mariboru in mlad raziskovalec na Fakulteti za naravoslovje in matematiko Univerze v Mariboru. Njegovo raziskovalno delo se osredotoča na krožno gospodarstvo, okoljsko inženirstvo, analizo materialnih tokov, oceno življenjskega cikla ter trajnostni razvoj v urbanem in logističnem kontekstu. Je soavtor več znanstvenih člankov v mednarodnih revijah z visokim faktorjem vpliva ter aktiven udeleženec mednarodnih znanstvenih konferenc. Sodeluje pri projektih, kot so BAS4SC, Erasmus+ in Digital Europe. Na pedagoškem področju poučuje predmete s področja krožne logistike ter sodeluje pri mentorstvu zaključnih del na prvi in drugi stopnji študija.

ORCID: 0000-0001-6433-8471

Maja Fošner, PhD

Redna profesorica za področje logistike (naravoslovje in matematika) na Fakulteti za logistiko Univerze v Mariboru, kjer je med letoma 2021 in 2025 opravljala funkcijo dekanice. Diplomirala je iz matematike na Pedagoški fakulteti Univerze v Mariboru in prejela Perlachovo nagrado. Magistrirala in doktorirala je s področja super algeber. Ima dolgoletne pedagoške in vodstvene izkušnje – bila je vodja več kateder, predsednica akademskega zbora, prodekanica, ter članica senata FL UM in UM. Vodi raziskovalni inštitut FL UM v Celju ter je aktivna v številnih projektih, vključno z Erasmus+, Digital Europe, COST in Green Campuses. Bila je vodja projekta "Modern logistics learning" in aktivno sodeluje pri razvoju mednarodnih akreditacij. Od leta 2018 je članica Evropske akademije znanosti in umetnosti v Salzburgu ter upravnega odbora ECBE. Njeno raziskovalno delo vključuje logistično modeliranje, e-izobraževanje, kakovost visokošolskega izobraževanja in mednarodno sodelovanje. Objavila je več kot 50 znanstvenih člankov, od tega 45 v revijah s faktorjem vpliva, in je recenzentka pri številnih uglednih revijah.

ORCID: 0009-0006-9272-0962

Roman Gumzej, PhD

Roman Gumzej je leta 1999 doktoriral s področja računalništva in informatike na Univerzi v Mariboru. Leta 2004 je bil izvoljen v naziv docent na Fakulteti za elektrotehniko in računalništvo, leta 2014 pa v naziv izrednega profesorja na Fakulteti za logistiko Univerze v Mariboru. Izvedel je več mednarodnih raziskovalnih projektov kot gostujoči raziskovalec na INSA de Lyon v Franciji leta 1998 in ifak e.V. Magdeburg v Nemčiji leta 2002. Kot gostujoči profesor je

predaval na Fakulteti za matematiko in računalništvo na Fernuniversität v Hagenu v Nemčiji v letih 2011, 2012 in 2013, prav tako na Fakulteti za pomorske študije, Univerza v Reki na Hrvaškem leta 2018. Njegova raziskovalna področja zajemajo vsa glavna področja računalništva v realnem času s posebno poudarkom na integracijah informacijskih sistemov v logistiki ter njihovi zanesljivosti in varnosti. Je član več mednarodnih strokovnih organizacij (npr. BVL, ELA, VDI/GI itd.) ter programskih in uredniških odborov mednarodnih konferenc in revij.

ORCID: 0000-0002-2646-217X

Rebeka Kovačič Lukman, PhD

Redna profesorica Fakultete za Logistiko Univerze v Mariboru z doktoratom s področja kemijskega inženirstva. Posebno se posveča trajnostnemu razvoju, krožnemu gospodarstvu, industrijski ekologiji ter oceni vplivov življenjskega cikla izdelkov in sistemov. Ves čas svoje kariere je koordinirala več kot 25 mednarodnih projektov, vključno s programom H2020 na področju krožnega gospodarstva. V reviji *Journal of Cleaner Production* je bila uvrščena med 13 globalno najbolj produktivnih raziskovalk na področju trajnosti – in najmlajša med njimi. S h-indeksom 14 in več kot 2 500 citati spada med 1 % najbolj citiranih raziskovalcev na svojem področju. Bila je gostujoča profesorica na TU Graz, Aalborgu in Osijeku, obiskala Manchester in Yale ter sodelovala na NTNU Trondheim. Je gostujoča urednica v *Journal of Cleaner Production* in *Sustainability*, članica uredniških odborov revij *LogForum* in *Logistics, Supply Chain, Sustainability and Global Challenges*, ter vodja mednarodnih odborov, kot so ERSCP in PREPARE. Izobraževalno se osredotoča na vključevanje trajnosti,

industrijske ekologije in krožne ekonomije v visokošolske kurikulume.

ORCID: 0000-0001-6733-9584

Benjamin Marcen, PhD

Doc. dr. Benjamin Marcen je doktor znanosti s področja matematike, svojo doktorsko disertacijo je opravil na Fakulteti za naravoslovje in matematiko Univerze v Mariboru. Zaposlen je kot visokošolski učitelj na Fakulteti za logistiko Univerze v Mariboru, kjer predava predmete s področij matematičnih metod in modelov v logistiki, statistike ter teorije optimizacije in načrtovanja v pametnih logističnih sistemih. Dejaven je na raziskovalnem področju, kjer sodeluje pri številnih projektih, povezanih z logistiko. Je soavtor znanstvenih in strokovnih člankov s področja algebre in logistike ter avtor več publikacij, v katerih se osredotoča na uporabo matematičnih pristopov za reševanje logističnih izzivov.

ORCID: 0000-0001-8436-6672

Marinko Maslarić, PhD

Prof. dr. Marinko Maslarić je redni profesor na Fakulteti za tehnične vede Univerze v Novem Sadu, kjer deluje na Katedri za logistiko in intermodalni transport. Diplomiral je iz prometnega inženirstva leta 2002, magistriral leta 2008 in doktoriral leta 2014. Na fakulteti je zaposlen od leta 2005, od leta 2024 kot redni profesor. Njegovo raziskovalno delo je osredotočeno na upravljanje oskrbovalnih verig, planiranje in nadzor zalog, digitalno transformacijo logističnih sistemov ter intermodalni transport. Je avtor več kot 100 znanstvenih in strokovnih prispevkov ter aktiven udeleženec raziskovalnih projektov, ki jih financirajo ministrstva Republike

Srbije in evropski programi. V svoji pedagoški in raziskovalni karieri tesno sodeluje z industrijo ter mednarodnim raziskovalnim okoljem. Poleg tekočega znanja angleščine ima tudi delovno znanje nemščine in ruščine.

ORCID: 0000-0002-4696-5188

Boško Matović, PhD

Doc. dr. Boško Matović je docent za prometno inženirstvo na Univerzi v Črni gori, kjer vodi študijski program cestnega prometa. Diplomiral, magistriral in doktoriral je na Fakulteti za tehnične vede Univerze v Novem Sadu. Njegovo raziskovalno delo se osredotoča na prometno varnost, pri čemer je sodeloval pri več kot 30 nacionalnih in mednarodnih projektih v regiji Zahodnega Balkana. Je član strokovnih in znanstvenih teles, med njimi Sveta za varnost prometa Republike Srbske ter programskih odborov mednarodnih konferenc. Je avtor številnih znanstvenih člankov v visoko rangiranih revijah. Ima certifikate za pregled in revizijo prometne varnosti ter za prevoz nevarnega blaga. Poleg pedagoškega dela deluje tudi kot svetovalec pri prometnovarnostnih strategijah in analizah. Odlikuje ga odlična usposobljenost za delo s strokovno programsko opremo in močna zavezanost k izboljšanju prometne varnosti in izobraževanja v regiji.

ORCID: 0000-0002-8336-8809

Dejan Mirčetić, PhD

Raziskovalni sodelavec na Inštitutu za raziskave in razvoj umetne inteligence Srbije ter asistent profesor na Fakulteti za tehnične znanosti Univerze v Novem Sadu. Dr. Dejan Mirčetić je doktoriral na področju napovedovanja povpraševanja in analitike oskrbovalne verige ter je avtor več kot 60 znanstvenih člankov. Njegove raziskave se osredotočajo predvsem na reševanje realnih poslovnih in industrijskih izzivov. Na Inštitutu za umetno inteligenco Srbije je odgovoren za oblikovanje in razvoj rešitev umetne inteligence za podjetja in industrijo. Trenutno vodi projekt ISIOP, ki je del platforme AIPlan4EU v okviru programa za raziskave in inovacije Obzorje 2020. Poleg dela na področju upravljanja oskrbovalne verige dr. Mirčetić pomembno prispeva k uporabi umetne inteligence v zdravstvu in prehrambni industriji, kjer se osredotoča na inovativne pristope k optimizaciji procesov in povečanju učinkovitosti.

ORCID: 0000-0002-1602-7908

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS -