Primjena programskog jezika R u inženjerstvu kvalitete

Jukić, Marko

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture / Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:235:334151

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

Repository of Faculty of Mechanical Engineering and Naval Architecture University of Zagreb

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

DIPLOMSKI RAD

Marko Jukić

Zagreb, 2021. godina

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

DIPLOMSKI RAD

Mentorice:

prof. dr. sc. Biserka Runje
dr. sc. Amalija Horvatić Novak

Student:

Marko Jukić

Zagreb, 2021. godina

Izjavljujem da sam ovaj rad izradio samostalno koristeći znanja stečena tijekom studija i navedenu literaturu.

Zahvaljujem mentorici prof. dr. sc. Biserki Runje, komentorici dr. sc. Amaliji Horvatić Novak i asistentu Andreju Razumiću, mag. ing. mech. na stručnom vodstvu, pomoći i korisnim savjetima pruženim za vrijeme izrade ovog rada.

Marko Jukić

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE
Središnje povjerenstvo za završne i diplomske ispite
Povjerenstvo za diplomske radove studija strojarstva za smjerove:
proizvodno inženjerstvo, računalno inženjerstvo, industrijsko inženjerstvo i menadžment,
inženjerstvo materijala te mehatronika i robotika

Sveučilište u Zagrebu Fakultet strojarstva i brodogradnje	
Datum:	Prilog:
Klasa: 602-04/21-6/1	
Ur. broj: 15-1703-21	

DIPLOMSKI ZADATAK

Student: **MARKO JUKIĆ** Mat. br.: 0035200479

Naslov rada na hrvatskom jeziku: **Primjena programskog jezika R u inženjerstvu kvalitete**

Naslov rada na engleskom jeziku: **Application of R programming language in quality engineering**

Opis zadatka:

Inženjerstvo kvalitete je inženjerska disciplina koja osigurava da se usluge ili proizvodi dizajniraju, razvijaju i proizvode kako bi udovoljili ili premašili očekivanja kupca. Za razliku od tradicionalne kontrole kvalitete inženjerstvo kvalitete uključuje razvojne sustave i napredne metode analize. Inženjerstvo kvalitete obuhvaća sve zadatke, od ideje do realizacije što uključuje analizu dizajna proizvoda, njegov razvoj i konačno proizvodne procese. Fokus je na poboljšanju kvalitete konačnog proizvoda i učinkovitom proizvodnom procesu uz prepoznavanje i smanjenje otpada u različitim oblicima. Alati i metode inženjerstva kvalitete često se razvijaju i primjenjuju pomoću višefunkcionalnog pristupa. R je danas jedan od vodećih programskih jezika s kojim se na učinkovit način mogu provesti složene analize podataka, statističko modeliranje te strojno učenje. U radu je potrebno dati prijedlog i primijeniti odabrane R-funkcije i pakete u tijeku procesa osmišljavanja, razvoja i implementacije proizvoda ili usluge.

U radu je potrebno navesti korištenu literaturu i eventualno dobivenu pomoć.

Zadatak zadan:	Rok predaje rada:	Predviđeni datum obrane:
4. ožujka 2021.	6. svibnja 2021.	10. svibnja do 14. svibnja 2021.
Zadatak zadao:	Komentor:	Predsjednica Povjerenstva:
prof. dr. sc. Biserka Runje	dr. sc. Amalija Horvatić Novak	prof. dr. sc. Biserka Runje

SADRŽAJ

SADRŽAJ I
POPIS SLIKA III
POPIS TABLICA V
POPIS OZNAKA VI
POPIS KRATICA VIII
SAŽETAK IX
1. UVOD 1
2. INŽENJERSTVO KVALITETE 2
2.1. Aktivnosti koje osiguravaju kvalitetu 4
2.2. Odnos između karakteristike proizvoda i utjecajnih faktora 5
2.3. Robusno konstruiranje 7
2.4. Taguchijeva funkcija gubitka 8
2.5. Alati inženjerstva kvalitete 9
2.6. Alati i metode za procjenu rizika 11
2.7. Vrste troškova kvalitete 12
3. R PROGRAMSKI JEZIK 15
3.1. Razvoj 15
3.2. Prednosti i nedostaci 15
3.3. Instalacija 18
3.4. Korisničko sučelje 20
3.5. Paketi nadogradnje 21
4. ALATI KONTROLE KVALITETE 23
4.1. Kontrolne karte (engl. *Control Chart*) 23
4.2. Prikaz kutija-brkovi (engl. *Box-whisker plot*) 24
4.3. Dijagram uzorka i posljedica (*engl. Cause and effect diagram*) 25
4.4. Kontrolni list (engl. *Check Sheet*) 27
4.5. Histogram 28
4.6. Pareto dijagram 28
4.7. Dijagram rasipanja (engl. *Scatter Plot*) 30
5. PRILAGODBA TEORIJSKE RASPODJELE EMPIRIJSKIM PODACIMA 32
5.1. Faze prilagodbe određene teorijske raspodjele empirijskim podacima 32
5.1.1. Priprema podataka 32
5.1.2. Odabir teorijske raspodjele i metode procjene 32
5.1.2.1. Metoda najveće vjerodostojnosti (engl. Maximum likelihood estimation – MLE) [17] 33
5.1.2.2. Metoda momenata (engl. Moment matching estimation – MME) [17] 33
5.1.2.3. Metoda kvantila (engl. Quantile matching estimation – QME) [18] 33
5.1.2.4. Metoda temeljena na procjeni dobrote modela (engl. Maximizing goodness-of-fit estimation – MGE) [18] 33
5.1.3. Grafička analiza 33

5.1.4. Numerička analiza ... 34
5.1.4.1. Kolmogorov-Smirnov test [19] ... 34
5.1.4.2. Cramer-von Mises test [21] ... 35
5.1.4.3. Anderson-Darling test [22] ... 36
5.1.4.4. Akaikeov i Bayesov informacijski kriterij [24] ... 37
5.1.5. Prihvaćanje ili odbacivanje teorijske raspodjele ... 38
5.2. Prilagodba trokutaste teorijske raspodjele empirijskim podacima ... 38
5.2.1. Grafička analiza ... 39
5.2.2. Numerička analiza ... 41
5.3. Prilagodba pravokutne teorijske raspodjele empirijskim podacima ... 43
5.3.1. Grafička analiza ... 43
5.3.2. Numerička analiza ... 46
6. SPOSOBNOST PROCESA ... 48
7. UTJECAJ OCJENE SUKLADNOSTI PROIZVODA NA PROCJENU RIZIKA PROIZVOĐAČA I POTROŠAČA ... 55
7.1. Pravila za donošenje odluka odluke ... 55
7.1.1. Pravilo jednostavnog prihvaćanja ... 55
7.1.2. Binarno pravilo ... 56
7.2. Primjer ... 57
8. PROCJENA MJERNE NESIGURNOSTI ... 65
9. ZAKLJUČAK ... 79

POPIS SLIKA

Slika 1. Razvoj postupaka utvrđivanja kvalitete 3
Slika 2. Varijacija procesa [7] 5
Slika 3. P-dijagram [6] 6
Slika 4. Robusno konstruiranje [6] 7
Slika 5. Tradicionalna funkcija gubitka kvalitete [6] 8
Slika 6. Taguchijeva funkcija gubitka kvalitete [6] 9
Slika 7. Dijagram "cijena kvalitete - sposobnost organizacije" [1] 14
Slika 8. Krivulja učenja R programskog jezika [4] 18
Slika 9. R softver - službena web stranica [14] 19
Slika 10. Razlika između R-a i RStudija [14] 19
Slika 11. Korisničko sučelje softvera RStudio 21
Slika 12. $\boldsymbol{x}$-$\boldsymbol{R}$ kontrolna karta 24
Slika 13. Prikaz kutija-brkovi 25
Slika 14. Dijagram uzroka i posljedica 26
Slika 15. Kontrolni list 27
Slika 16. Histogram 28
Slika 17. Pareto dijagram 29
Slika 18. Dijagram rasipanja 31
Slika 19. Tablica graničnih vrijednosti - Kolmogorov-Smirnov test [20] 35
Slika 20. Tablica graničnih vrijednosti - Cramer-von Mises test[21] 36
Slika 21. Tablica graničnih vrijednosti - Anderson-Darling test [23] 37
Slika 22. Histogram empirijskih podataka za trokutastu raspodjelu 39
Slika 23. Usporedba histograma podataka s trokutastom raspodjelom 40
Slika 24. Usporedba empirijskih i teorijskih kvantila 40
Slika 25. Usporedba empirijske i teorijske kumulativne funkcije 41
Slika 26. Usporedba histograma podataka s pravokutnom raspodjelom 44
Slika 27. Usporedba empirijskih i teorijskih kvantila 45
Slika 28. Usporedba empirijske i teorijske kumulativne funkcije 45
Slika 29. $\boldsymbol{X}$ kontrolna karta 49
Slika 30. R kontrolna karta 49
Slika 31. Sposobnost procesa 50
Slika 32. Cullen-Frey graf 51
Slika 33. Sposobnost procesa primjenom Weibullove raspodjele 52
Slika 34. Sposobnost procesa primjenom log-normalne raspodjele 53
Slika 35. Sposobnost procesa primjenom gamma raspodjele 53
Slika 36. Sposobnost procesa primjenom Box-Cox transformacije 54
Slika 37. Pravilo jednostavnog prihvaćanja s gornjom i donjom granicom tolerancije [27] 56
Slika 38. Binarno pravilo ocjene sukladnosti [25] 56
Slika 39. Vjerojatnost sukladnosti 57
Slika 40. Funkcije povezane s normalnom raspodjelom podataka unutar R programskog jezika [28] 59
Slika 41. Graf ovisnosti mjerne nesigurnosti o iznosu rizika proizvođača i kupca 63
Slika 42. Graf ovisnosti mjerne nesigurnosti o iznosu rizika proizvođača i kupca - uvećano 64
Slika 43. Skalarni odnos između ulaznih veličina i mjerene veličine [29] 65
Slika 44. Vektorski odnos između ulaznih veličina i mjerene veličine [29] 65

Slika 45. Pravokutna razdioba [24] 67
Slika 46. Koeficijent osjetljivosti [26] 68
Slika 47. Simetrična trapezna razdioba [27] 69
Slika 48. Funkcija gustoće vjerojatnosti izlazne veličine (Trapezna razdioba) 72
Slika 49. Shema strujnog kruga 75

POPIS TABLICA

Tablica 1. Alati upravljanja kvalitetom [8] 9
Tablica 2. Alati inženjerstva kvalitete [6] 11
Tablica 3. Metode i alati za procjenu rizika [9] 12
Tablica 4. Doprinosi mjerne nesigurnosti 67
Tablica 5. Ulazne veličine (x_i) definirane su funkcijama gustoće vjerojatnosti $g(x_i)$ za izračun pogreške pokazivanja E_x 70
Tablica 6. Ulazne veličine (x_i) definirane su funkcijama gustoće vjerojatnosti $g(x_i)$ za izračun karakteristika kvalitete strujnog kruga 75

POPIS OZNAKA

Oznaka	Mjerna jedinica	Opis oznake
$1 - \alpha$	%	interval vjerodostojnosti
A_L	A	donja granica prihvaćanja
a_p		granica intervala pravokutne raspodjele
a_t, b_t, c_t, d_t	mm	koordinate vrhova trapeza
A_U	A	gornja granica prihvaćanja
c_i		koeficijent osjetljivosti
C_{pk}		demonstrirana izvrsnost procesa
E_x	mm	pogreška pokazivanja
$F(z)$		standardna normalna kumulativna funkcija raspodjele s varijablom z
$g(x_i)$		funkcije gustoće vjerojatnosti
H_0		nulta hipoteza
H_1		alternativna hipoteza
k		koeficijent gubitka
k_p		faktor pokrivanja
$L(y)$		funkcija koja opisuje gubitak kvalitete
l_{ix}	mm	pokazivanje pomičnog mjerila
l_S	mm	duljina stvarne granične mjerke
L_S	mm	nazivna duljina stvarne granične mjerke
M		signalni faktor
m		ciljana vrijednost karakteristike kvalitete
M		broj simulacija
N		faktor šuma
N, n		veličina uzorka
p, P	%	vjerojatnost
p_s	%	vjerojatnost sukladnih strojeva
$\bar{ps}$	%	vjerojatnost nesukladnih strojeva
R_C	%	rizik potrošača
R_P	%	rizik proizvođača
T	A	tolerancijsko područje
T_L	A	donja granica specifikacije
T_U	A	gornja granica specifikacija
U	A	mjerna nesigurnost
u_0	A	mjerna nesigurnost

u_m	A	mjerna nesigurnost opreme za kontrolu
X		kontrolni faktor
X_n		ulazna veličina
y		vrijednost karakteristike kvalitete
Y		izlazna veličina
y_0	A	srednja vrijednost
z		varijabla standardizirane normalne razdiobe
α	%	vjerojatnost pogreške
$\bar{\alpha}$	$°C^{-1}$	prosječni koeficijent toplinskog širenja pomičnog mjerila i granične mjerke
β		parametar brida
Δ		granica tolerancije
δL_{ix}	mm	ispravak zbog konačnog razlučivanja pomičnog mjerila
δL_M	mm	ispravak zbog mehaničkih posljedica kao što su primijenjena sila za mjerenja, Abbeove pogreške, pogreške ravnine i paralelizma mjernih ploha
Δt	°C	razlika temperatura između mjerila promjera i granične mjerke
$\rho_0(z)$		standardna normalna funkcija gustoće vjerojatnosti s varijablom *z*
σ	A	standardno odstupanje
a	mm	donja osnovica trapeza
b	mm	gornja osnovica trapeza
$u(x_i)$		standardna nesigurnost
ω	A	duljina sigurnosnog područja

POPIS KRATICA

Kratica	**Opis**
AD	*Anderson-Darling test* – Anderson-Darling test
AIC	*Akaike information criterion* – Akaikeov informacijski kriterij
BIC	*Bayesian Information Criterion* – Bayesov informacijski kriterij
CRAN	*Comprehensive R Archive Network* – sveobuhvatna arhivska mreža
CvM	*Cramer-von Mises test* – Cramer-von Mises test
DFMEA	*Design failure mode and effect analysis* – Konstruiranje i analiza mogućih grešaka i njihovih posljedica
DPMO	*defects per million opportunities* – greške na milijun mogućnosti
FMEA	*Failure Mode and Effect Analysis* – Analiza mogućih grešaka i njihovih posljedica
FOSS	*Free and Open Source Software* – besplatni softver otvorenog koda
FTA	*Fault tree analysis* – Stablo rizika
GA	*Genetic algorithm* – Genetski ili generički algoritam
IDE	*integrated development environment* – integrirano razvojno okruženje
ISO	*International organization for standardization* – Međunarodna organizacija za standardizaciju
JCGM	*Joint Committee for Guides in Metrology* – Zajednički odbor za upute u mjeriteljstvu
KLI	*Kullback-Leibler information* – Kullback-Leibler kriterij
KS	*Kolmogorov-Smirnov test* – Kolmogorov-Smirnov test
MGE	*Maximizing goodness-of-fit estimation* – metoda temeljena na procjeni dobrote modela
MLE	*Maximum likelihood estimation* – metoda najveće vjerodostojnosti
MME	*Moment matching estimation* – metoda momenata
MTS	*Mahalanobis-Taguchi system* – Mahalanobis-Taguchi
PDCA	*Plan, Do, Check, Act* – *D*emingov krug poboljšanja kvalitete
QE	*Quality Enginering* – Inženjerstvo kvalitete
QME	*Quantile matching estimation* – metoda kvantila
RSM	*response surface methodology* – metoda odzivnih površina
SPC	*Statistical Process Control* – statistička kontrola podataka

SAŽETAK

Inženjerstvo kvalitete obuhvaća sve aktivnosti koje osiguravaju kvalitetu proizvoda kroz njegov životni ciklus, a to su planiranje proizvoda, konstruiranje proizvoda, razvoj proizvodnog procesa sam proizvodni proces te usluge nakon prodaje. Za razliku od tradicionalne kontrole kvalitete u inženjerstvu kvalitete koriste se složene analize podataka, statističko modeliranje te strojno učenje koji su potpomognuti raznim računalnim programima. R programski jezik je jedan od vodećih programa u tom području te je u ovom radu prikazano njegovo široko područje primjene. Uvodno su prikazani jednostavni alati za kontrolu kvalitete kao što su histogram, kontrolne karte, Paretov dijagram, dijagram uzroka i posljedica te slični. Nakon toga prikazan je postupak prilagodbe teorijske raspodjele podataka empirijskim podacima te je prikazan način izračuna sposobnosti procesa. U nastavku je napravljena analiza kako ocjena sukladnosti proizvoda utječe na podjelu rizika između kupca i proizvođača. Korištenjem Monte Carlo simulacije u posljednjem dijelu rada napravljena je procjena mjerne nesigurnosti i statistička analiza tolerancija.

Ključne riječi:

Inženjerstvo kvalitete, R, R programski jezik, prilagodba teorijske raspodjele, sposobnost procesa, statistička analiza tolerancija

SUMMARY

Quality engineering encompasses all activities that ensure the quality of a product throughout its life cycle, namely product planning, product design, process design, production, and service after purchase. Unlike traditional quality control, quality engineering uses complex data analysis, statistical modeling and machine learning that are supported by various computer programs. The R programming language is one of the leading programs in this field, and this paper presents its wide field of application. Introductory, this paper explains simple quality control tools such as histogram, control chart, pareto diagram, the cause and effect diagram etc. After that, the paper is focused on the procedure of fitting distribution to empirical data, as well as the method of calculating the process capability. Following that is an evaluation on how the conformity assessment of a product affects the division of risk between the customer and the producer. In the final part of the paper the focus is on the conformity assessment and statistical analysis of tolerances using the Monte Carlo simulation.

Key words:

Quality engineering, R, R programming language, fitting distribution, process capability, statistical tolerances analysis

1. UVOD

Riječ kvaliteta često se koristi za označavanje izvrsnosti proizvoda ili usluge [1]. Izvrsnost je nešto čemu je društvo težilo još od drevne ljudske civilizacije. To se može vidjeti na primjerima antičkih građevina stare Grčke ili egipatskih piramida. Razvojem civilizacije razvijala se i industrija te se definicija kvalitete mijenjala. Nekada se pojam kvalitete upotrebljavao da bi se označilo da je proizvod u skladu s određenim fizičkim karakteristikama koje su utvrđene granicama specifikacija. Razvojem tehnologije i globalizacijom tržišta koja omogućava laku dostupnost proizvoda sama sukladnost proizvoda nije dovoljna, već je potrebno da proizvod ispunjava sve zahtjeve kupaca. Stoga je danas kvaliteta jednostavno definirana kao udovoljavanje zahtjevima kupaca [1]. Kako bi se postigla tražena razina kvalitete, sama kontrola kvalitete gotovog proizvoda nije dovoljna, već je potrebno kvalitetu ugraditi u sve faze životnog vijeka proizvoda. Takav pristup koji uza sebe povlači i korištenje suvremenih računalnih alata i tehnologija naziva se inženjerstvo kvalitete. Veliki broj programskih jezika podržava računalne alate za inženjerstvo kvalitete, jedan od njih je i R programski jezik. R je jedan od najpopularnijih besplatnih programskih softvera otvorenog koda za statističko modeliranje i analizu te je korišten u nastavku za prikaz određenih statističkih alata i rješavanje problema.

2. INŽENJERSTVO KVALITETE

Kako bi se precizno definiralo inženjerstvo kvalitete, u nastavku je definiran pojam kvalitete i povijesni razvoj postupaka ostvarivanja kvalitete.

Razni izvori daju razne definicije kvalitete, u nastavku su navede neke od njih:

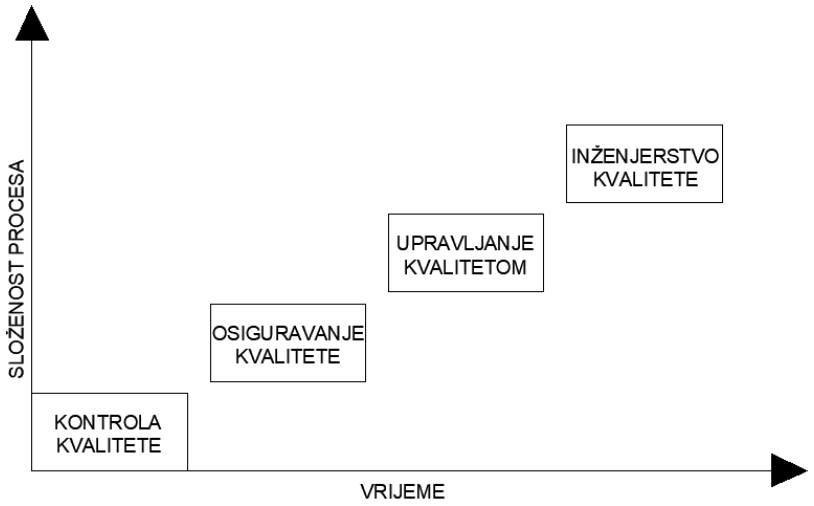
- Kvaliteta se definira kao stupanj do kojeg skupina svojstvenih karakteristika ispunjava zahtjeve [2].
- Kvaliteta se definira kao svojstvo, vrsnoća neke stvari, vrednota, odlika, značajka ili sposobnost [3].
- Kvaliteta je ono što označuje (određuje) neki predmet ili pojavu i razlikuje ih od ostalih predmeta ili pojava, svojstvo, osobina, odlika, obilježje, značajka; vrijednost, valjanost, vrsnoća neke stvari, njezina primjerenost određenim uzorima, paradigmama, normama (kvaliteta tehničkih i umjetničkih radova, prirodnih materija, proizvoda, tkanina, namirnica i sl.) [4].

Slika 1. prikazuje razvoj postupaka ostvarivanja kvalitete kroz povijest dok su definicije postupaka i njihova obilježja dana u nastavku.

1) Kontrola kvalitete usmjerena je na ispunjenje zahtjeva kvalitete [2]. Kontrola kvalitete označava postupak kontrole gotovog proizvoda. Karakterizira ju povijesno razdoblje kada je proizvodnja bila veća od potražnje proizvoda, proizvođači su diktirali izgled i svojstva proizvoda te kupac nije imao izbora već je morao prihvatiti ono što mu se nudi.
2) Osiguravanje kvalitete usmjereno je na stvaranje sigurnosti da su zahtjevi kvalitete ispunjeni [2]. U ovoj fazi proizvođači veću pozornost posvećuju izgledu proizvoda te kvaliteti. Osiguravanje kvalitete karakterizira statistička kontrola procesa (engl. *Statistical Process Control* – SPC). SPC je statistički alat koji se bavi prikupljanjem podataka o proizvodnom procesu i proizvodu, analizom prikupljenih podataka stvara se sustav koji čini osnovu za donošenje odluka, a sve u cilju održavanja stabilnosti proizvodnog procesa i ciljane kvalitete proizvoda.
3) Upravljanje kvalitetom razvija i formira politiku kvalitete i ciljeve kvalitete, te određene procese za formiranje tih ciljeva kao što su planiranje kvalitete, osiguravanje kvalitete, kontrola kvalitete i poboljšavanje kvalitete [2]. Konkurentnost među

proizvođačima kupca stavlja u mogućnost ne samo da diktira kvalitetu, već da zahtjeva proizvode po svojim željama. Upravljanje kvalitetom karakterizira Šest sigma metodologija (engl. *Six Sigma*) kojoj su težili proizvođači. Šest sigma predstavlja 99,9996 % uspješnosti. Ova razina uspješnosti (savršenstva) je ekvivalentna pojavi 3,4 greške na milijun mogućnosti (engl. *defects per million opportunities* – DPMO) [5].

4) Inženjerstvo kvalitete je inženjerska disciplina koja obuhvaća sve aktivnosti povezane s kvalitetom proizvoda ili usluga kroz cijeli njegov životni ciklus, a sve u cilju udovoljavanja zahtjevima kupaca. Aktivnosti povezane s kvalitetom su planiranje proizvoda, konstruiranje proizvoda, razvoj proizvodnog procesa sam proizvodni proces te usluge nakon prodaje. Kao što je rečeno, upravljanje kvalitetom generira samo nekoliko grešaka na milijun mogućnosti. Inženjerstvo kvalitete nastoji pravovremeno otkriti i spriječiti i te rijetke događaje, a glavni adut za to im je digitalizacija koja je omogućila razvoj Velikih podataka (*engl. Big data*) i umjetne inteligencije (engl. *Artificial Intelligence*). Veliki podaci kako i samo ime kaže označavaju veliki broj podataka s više izvora koje je moguće prikupiti u vrlo kratkom vremenu. Te podatke nije moguće obraditi konvencionalnim računalima već je potrebna primjena računalnih metoda kao što su neuronske mreže ili genetski algoritmi koji rade na bazi umjetne inteligencije.



Slika 1. Razvoj postupaka utvrđivanja kvalitete

2.1. Aktivnosti koje osiguravaju kvalitetu

Genichi Taguchi japanski je inženjer i statističar koji je poznat po svojim dostignućima vezanim za kvalitetu. Razvio je funkciju gubitka kvalitete (engl. *Quality loss function*), unaprijedio metodu planiranja pokusa (engl. *Design of experiments*) te razvio metodologiju za poboljšanje kvalitete i smanjenje troškova poznatu kao Taguchijeva metoda (engl. *Taguchi Method*). Osim navedenog, Taguchi je objavio ključne aktivnosti koje osiguravaju kvalitetu proizvoda kroz njegov ciklus. Aktivnosti su podijeljene u pet faza i navedene u nastavku [6]:

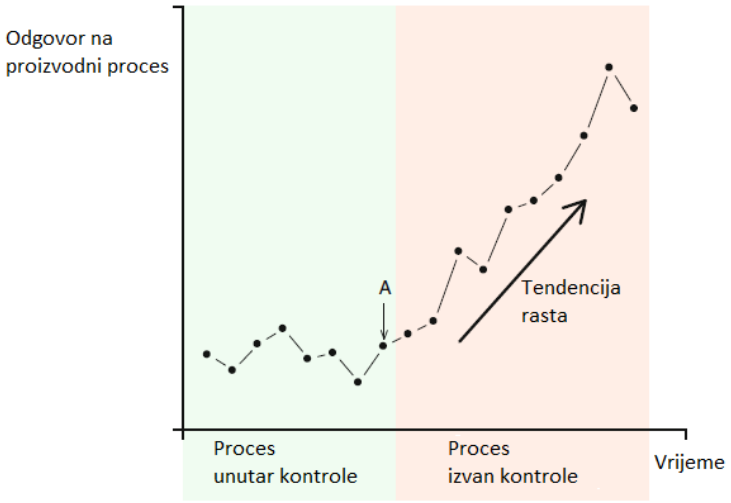
1) Planiranje proizvoda – cilj ove faze je da funkcija, cijena i životni ciklus proizvoda zadovoljavaju zahtjeve kupaca. Prije stavljanja proizvoda na tržište potrebno je prikupiti informacije u vidu želja, potreba i očekivanja kupaca.
2) Konstruiranje proizvoda – ova faza zahtjeva konstruiranje proizvoda koji ima funkcije određene u prethodnoj fazi. Potrebno je uzeti u obzir sve konstrukcijske aspekte koji utječu na odstupanje funkcionalnih karakteristika proizvoda od ciljanih vrijednosti. U ovoj se fazi obično primjenjuju tri koraka – konstruiranje sustava, odabir parametara i definiranje tolerancija.
3) Razvoj proizvodnog procesa – nakon što je konstruiran proizvod potrebno je konstruirati i razviti njemu odgovarajući proizvodni proces. Tri koraka korištena tijekom faze konstruiranja proizvoda mogu se primijeniti i na razvoj procesa.
4) Proizvodni proces (ili proizvodnja) – cilj ove faze je proizvodnja proizvoda zahtijevane kvalitete. Potrebno je pratiti i prilagođavati uvjete procesa promatrajući karakteristike kvalitete proizvoda i parametre procesa. Statistička kontrola procesa i povratna kontrola tipični su pristupi koji se mogu primijeniti u fazi proizvodnje.
5) Usluge nakon prodaje – usprkos tome što je kontrola kvalitete primijenjena tijekom faza konstruiranja proizvoda, razvoja proizvodnog procesa te same proizvodnje, neki neispravni proizvodi i dalje mogu završiti kod kupaca. Stoga je potrebno osigurati odgovarajuću uslugu nakon isporuke proizvoda kupcu. Usluga uključuje popravak ili zamjenu neispravnih proizvoda ili njihovih dijelova te nadoknađivanje pretrpljene štete.

Aktivnosti kontrole kvalitete provedene tijekom proizvodnje nazivaju se izravna kontrola kvalitete (engl. *on-line Quality Control*) [6].

Temelj izravne kontrole kvalitete je varijacija procesa. Varijacija se može definirati kao odstupanje kvalitete konačnog proizvoda. Prema uzroku varijacija se dijeli na [1]:

- običnе (slučajne) uzroke (engl. *Common Causes*)
- posebne (značajne) uzroke (engl. *Special Causes*)

Stabilan proces (ili proces unutar kontrole) pokazuje samo obične (slučajne) uzroke varijacija. To su varijacije nastale uslijed promjena u karakteristikama materijala, uvjetima rada i izvedbi radnika. Takve varijacije su neizbježne te je u tom slučaju ponašanje stabilnog procesa predvidljivo. Nasuprot tome, proizvodni proces može biti i izvan statističke kontrole. U određenim okolnostima neki uzroci postanu dovoljno snažni da uvode prepoznatljive uzorke ponašanja, oni se nazivaju posebni (značajni) uzroci varijacija te ih je potrebno eliminirati. Proces koji je izvan statističke kontrole može biti uzrokovan kvarom na stroju, istrošenosti alata, loše konstruiranim proizvodnim postupkom, itd. Zadatak izravne kontrole kvalitete je održavanja stabilnosti procesa te otkrivanja specijalnih uzroka varijacija i uklanjanje istih. Slika 2. prikazuje da se tendencija rasta jasno uočava u podacima nakon točke A, odnosno da se pojavljuje utjecaj značajnih uzroka.



Slika 2. Varijacija procesa [7]

Aktivnosti kontrole kvalitete u fazama planiranja proizvoda, konstruiranja proizvoda i razvoja proizvodnog procesa nazivaju se neizravna kontrola kvalitete (engl. *off-line Quality Control)* [6]. Primjeri aktivnosti neizravne kontrole kvalitete dani su u nastavku.

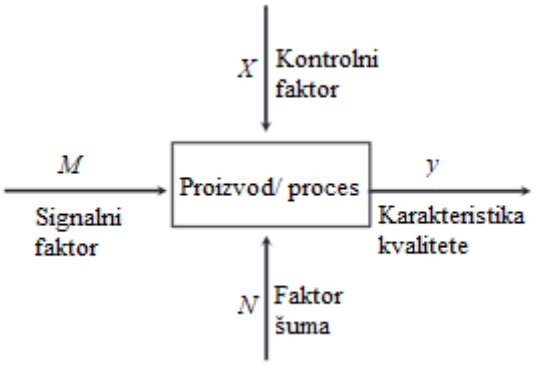
2.2. Odnos između karakteristike proizvoda i utjecajnih faktora

Odnosi između krajnjeg proizvoda i različitih faktora pri konstruiranju, razvoju i proizvodnji često su složeni i nelinearni. Inženjerstvo kvalitete uključuje pronalazak kombinacije vrijednosti parametara proizvoda koja dovodi do najmanjih odstupanja u karakteristikama

kvalitete proizvoda u odnosu na specificirane vrijednosti [6]. Kako bi se pronašla tražena kombinacija koristi se P-dijagram [Slika 3.].

Karakteristika kvalitete na P-dijagramu je označena *y* dok su parametri koji mogu utjecati na nju klasificirani u signalne faktore (*M*), kontrolne faktore (*X*) i faktor šuma (*N*). U nastavku će biti pobliže opisan svaki od njih [6]:

1) Signalni faktori (*M*) – ako se promijeni ciljana vrijednost karakteristike kvalitete (*y*), moguće je prilagoditi signalni faktor i na taj način postići traženu razinu kvalitete. Faktor signala određuje korisnik proizvoda ili operater kako bi postigao željeni odgovor. Primjerice, postavka brzine ventilatora može predstavljati signalni faktor kojim se može prilagoditi količina protoka zraka.
2) Kontrolni faktori (*X*) – konstruktor proizvoda u većini slučajeva određuje i specificira kontrolne faktore. On zapravo određuje razine kontrolnih faktora kako bi se umanjio gubitak *y*. Primjerice, broj defekata na površini poluvodičkih pločica (wafera) značajka je kvalitete tijekom postupka taloženja polisilicija, a kontrolni čimbenici koji utječu na ovu karakteristiku kvalitete su temperatura taloženja, tlak, protok dušika i protok silicija. Dizajner može odrediti potrebne postavke parametara, na primjer, postavljanjem temperature taloženja na 100 °C, 200 °C ili 300 °C.
3) Faktori šuma (*N*) – predstavljaju faktore koje konstruktor ne može kontrolirati. Određeni faktori se smatraju faktorima šuma sve dok je njihovu razinu teško ili preskupo kontrolirati. Utjecaj faktora šuma se mijenja s vremena na vrijeme, stoga ne postoji način da se u određenim uvjetima saznaju stvarne vrijednosti faktora šuma. Faktori šuma su ti koji mogu utjecati na odziv *y*, tj. mogu dovesti do odstupanja od ciljane izlazne vrijednosti i time postizanja gubitka kvalitete.

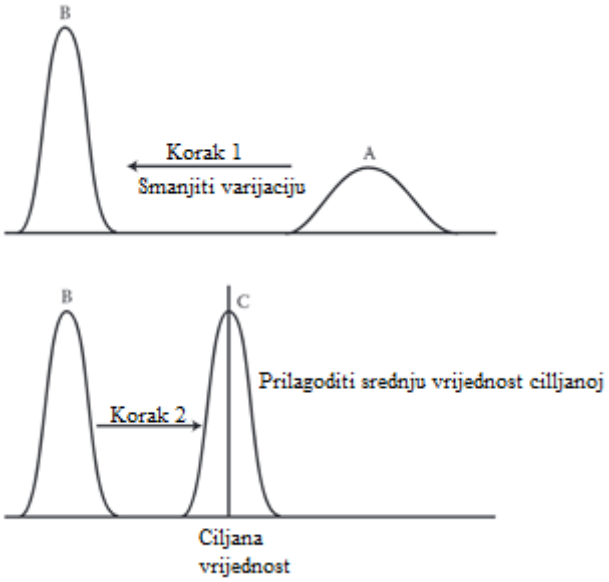


Slika 3. P-dijagram [6]

2.3. Robusno konstruiranje

Taguchi je zagovaratelj robusnog konstruiranja, odnosno proizvoda čija će kvaliteta biti jednaka kod različitih proizvodnih uvjeta. Glavno pitanje robusnog konstruiranja je mogu li tražene karakteristike kvalitete proizvoda zadovoljiti ciljanu vrijednost. Primjenom inženjerske analize i odabirom nekoliko kontrolnih faktora eksperimentalno je moguće odrediti je li srednja vrijednost tražene karakteristike blizu ciljane vrijednosti. Uz srednju vrijednost potrebno je i odrediti koliki je utjecaj faktora šuma odnosno kolika je varijacija karakteristike. Ako srednja vrijednost nije jednaka ciljanoj vrijednosti, potrebne su određene korekcije. Međutim, ako se prvo prilagodi srednja vrijednost, a zatim pokuša pronaći rješenje za smanjenje varijacija, teže je provesti samu korekciju te će u većini slučajeva cijena biti veća. Zbog toga Taguchi predlaže postupak optimizacije u dva koraka (engl. *two-step optimization procedure*) [Slika 4.].

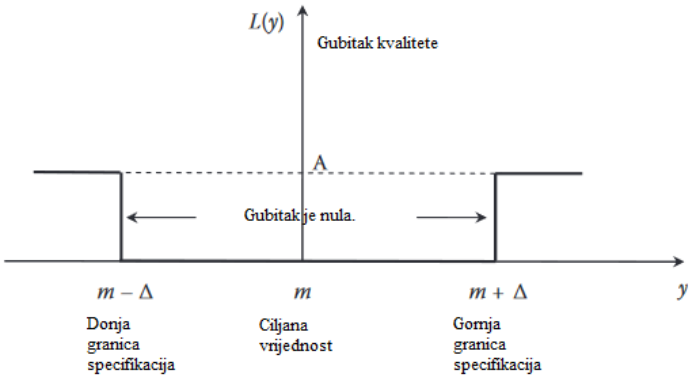
- Smanjenje varijacije – ovoj fazi se zanemaruje srednja vrijednost i odabire razina kontrolnih čimbenika kako bi se smanjila osjetljivost na šum.
- Prilagođavanje srednje vrijednosti ciljanoj vrijednosti – ovoj se fazi koristi faktor prilagodbe kako bi srednja vrijednost odgovarala ciljanoj. Kao faktor prilagodbe koristi se kontrolni faktor koji najmanje utječe na varijaciju, a najviše na srednju vrijednost.



Slika 4. Robusno konstruiranje [6]

2.4. Taguchijeva funkcija gubitka

Tradicionalno se većina proizvođača usredotočuje na postotak proizvoda koji ne ispunjava zahtjeve, odnosno na proizvode čije se karakteristike kvalitete ne nalaze u granicama specifikacija. Svi proizvodi smatraju se podjednako dobrima kada udovoljavaju specifikacijama, također svi se smatraju podjednako lošima kada se utvrdi da su izvan specifikacija. Slika 5. prikazuje takav pristup.

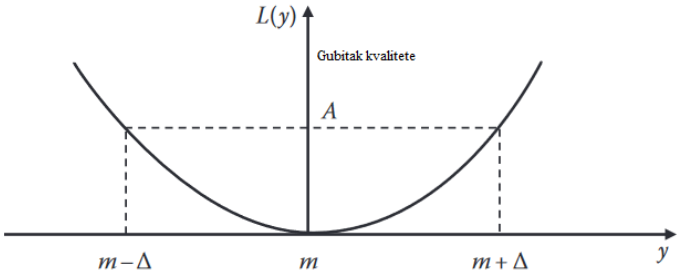


Slika 5. Tradicionalna funkcija gubitka kvalitete [6]

Vodoravna os predstavlja karakteristiku kvalitete proizvoda, a okomita os predstavlja gubitak kvalitete $L(y)$. Područje specifikacije zapisano je kao $m \pm \Delta$ te je gubitak u tom području jednak nuli. Proizvod koji se nalazi izvan granica specifikacije smatra se nesukladnim te je potrebna dorada ili će biti proglašen otpadom.

Nezadovoljan tradicionalnim pristupom Taguchi formira funkciju gubitka u kojoj naglašava gubitke povezane s proizvodom. Smatra da je kvaliteta gubitak koji proizvod nanosi društvu nakon što je isporučen, osim gubitaka koji su uzrokovani njegovim funkcijama. Smatra da porast varijacije određenog proizvoda rezultira porastom gubitka za kupca ili šire društvo.

Funkcija gubitka kvalitete može se učinkovito koristiti za procjenu kvalitete u ranoj fazi razvoja proizvoda ili procesa, pomažući tako u provođenju inženjerskih odluka (poput izbora materijala, dijelova ili konstrukcijskih rješenja). Slika 6. prikazuje Taguchijevu funkciju gubitka koja se može usporediti s ranije navedenom tradicionalnom funkcijom gubitka.



Slika 6. Taguchijeva funkcija gubitka kvalitete [6]

Pretpostavljeno je da karakteristike kvalitete imaju konačnu ciljanu vrijednost m ($m \neq 0$), a gubitak kvalitete simetričan je s obje strane cilja. Vidljivo je da se povećanjem varijacije karakteristike kvalitete povećava i gubitak kvalitete. Funkcija koja opisuje gubitak prikazana je u nastavku.

$$L(y) = k(y - m)^2 \quad \textbf{(1)}$$

Gdje je:

y	stvarna vrijednost karakteristike kvalitete
m	ciljana vrijednost karakteristike kvalitete
k	koeficijent gubitka

2.5. Alati inženjerstva kvalitete

Kako bi se inženjerstvo kvalitete primijenilo na određeni sustav, koristi se veliki broj alata koji imaju presudnu ulogu u mjerenju, opisivanju, analiziranju, tumačenju i modeliranju sustava [8]. Alati kvalitete bit će podijeljeni u dvije tablice. Prvu će tablicu činiti alati koji se koriste u upravljanju kvalitetom [Tablica 1.], dok će drugu činiti alati koji su razvijeni u inženjerstvu kvalitete [Tablica 2.].

Tablica 1. Alati upravljanja kvalitetom [8]

ALATI UPRAVLJANJA KVALITETOM		
LEAN alati	**Analitički alati**	**Alati za poticanje kreativnosti i razmišljanja**
Mapa toka vrijednosti (engl. *Value stream map*)	Statistička kontrola procesa (engl. *Statistical process control*)	Oluja mozgova (engl. *Brain Storming*)
5S	Pareto dijagram	Tehnika nominalne grupe (engl. *Nominal group*)

Poka Yoke	Dijagram uzroka i posljedica (engl. *Cause and Effect Diagrams*)	Praktične vježbe (engl. *Practical Technique*)
Kaizen	Histogram	Dijagram afiniteta (engl. *Affinity Diagram*)
Analiza toka rada (engl. *Work Flow Analysis*)	Dijagram raspršenosti (engl. *Scatter Diagrams*)	5 Zašto? (engl. *5 Why*)
	Mapa toka procesa (engl. *Process flow charting*)	Analiza sile polja (engl. *Force Field*)
Alati za planiranje	**Alati za praćenje procesa**	**Alati za razvoj i konstruiranje**
Matrični dijagrami (engl. *Matrix Diagrams*)	Statistička kontrola procesa (engl. *Statistical process control*)	Analiza mogućih grešaka i njihovih posljedica (engl. *Failure Mode and Effect Analysis* – FMEA)
Matrica odgovornosti (engl. *Responsibility Matrix*)	Dijagram trendova (engl. *Trend Charts*)	Konstruiranje i analiza mogućih grešaka i njihovih posljedica (engl. *Design failure mode and effect analysis* – DFMEA)
	Histogrami	Planiranje pokusa (engl. *Design of Experiments*)
	Gantogrami	Glas kupca (engl. *Voice of Customer*)

Kada ranije navedeni alati upravljanja kvalitetom nisu dovoljni za rješavanje određenog problema znanstvenici koriste nova rješenja kao što su razni računalni sustavi koji mogu biti potpomognuti umjetnom inteligencijom. Takve metode najvećim dijelom se koriste u ranije opisanoj neizravnoj kontroli kvalitete koja ima svoje tri faze [6]:

- Konstruiranje sustava (engl. *System design*) – odabir sustava kojim se može postići ciljana vrijednost

- Odabir parametara (engl. *Parameter design*) – određivanje vrijednosti parametara unutar sustava kako bi se smanjio utjecaj faktora šuma
- Definiranje tolerancija (engl. *Tolerance design*) – određivanje tolerancija proizvoda tako da se postigne željena ravnoteže između kvalitete i cijene.

U nastavku će biti navedene napredne tehnike i metode koje se koriste u inženjerstvu kvalitete te će biti podijeljene u dvije skupine. Prva skupina označava metode koje nisu potpomognute sustavom umjetne inteligencije, već su bazirane na određenom matematičkom ili analitičkom aparatu. Drugoj skupini pripadaju metode koje su bazirane na sustavu umjetne inteligencije odnosno metode strojnog učenja. Strojno učenje grana je umjetne inteligencije koja stvara algoritme učenjem iz određenog skupa podataka.

Tablica 2. Alati inženjerstva kvalitete [6]

Metode koje ne koriste umjetnu inteligenciju	Metode koje koriste umjetnu inteligenciju
Mahalanobis-Taguchi sustav (engl. *MTS – Mahalanobis-Taguchi system*)	Regresijske metode (engl. *Regression*)
Ortogonalni nizovi (engl. *Orthogonal arrays*)	Neuronske mreže (engl. *Artificial Neural Networks*)
Omjer faktora signala i šuma (engl. *Signal-to-noise ratio*)	Stabla odlučivanja (engl. *Decision Tree Learning*)
Konstruiranje dinamičkih parametara (engl. *dynamic parameter design*)	Genetski ili generički algoritam (engl. *Genetic algorithm* – GA)
Metoda odzivnih površina (engl. *RSM – response surface methodology*)	

2.6. Alati i metode za procjenu rizika

Rizik se definira kao učinak neizvjesnosti na postavljenje ciljeve, učinak koji predstavlja odstupanje od očekivanog stanja može biti pozitivan ili negativan [9]. Rizik se smatra kao glavni uzrok nesigurnosti u bilo kojoj organizaciji. Stoga, organizacije koje se više fokusiraju na procjenu rizika i upravljanje njime mogle imati veći potencijal za ostvarivanje postavljenih poslovnih ciljeva te na taj način osigurati napredak. Proces procjene rizika podijeljen je u tri dijela: identifikacija rizika, analiza rizika te ocjenjivanje rizika.

Ako se gleda prisutnost rizika u realnom sektoru proizvodnje ili u jednom životnom vijeku proizvoda može se reći da je u svakoj fazi prisutan rizik. Primjerice, nakon istraživanju tržišta i financijske analize procijenjeno je da će se prihvatiti određeni rizik te da će se krenuti u razvoj i konstruiranje proizvoda. Pri razvoju i konstruiranju, kako bi određeni parametri kvalitete bili zadovoljeni, koriste se alati i metode inženjerstva kvalitete navedeni u prethodnom poglavlju. Nakon što je proces razvoja gotov potrebno je procijeniti jesu li proces proizvodnje te sam proizvod dovoljno dobri da uspiju na tržištu i zadovolje potrebe kupca odnosno rizik početka proizvodnje. Kao što postoje alati i metode u inženjerstvu kvalitete, tako postoje i alati i metode u procjeni rizika, neki od alata se koriste u obje discipline. U nastavku će u tablici biti navedeni neki od alata i metoda za procjenu rizika te u kojoj se fazi procjene rizika mogu koristiti.

Tablica 3. Metode i alati za procjenu rizika [9]

Alati i metode za procjenu rizika	Identifikacija rizika	Analiza rizika	Ocjenjivanje rizika
Oluja mozgova (engl. *Brain Storming*)	■		
Kontrolna lista (engl. *Check-list*)	■		
Analiza mogućih grešaka i njihovih posljedica (engl. *FMEA – Failure Mode and Effect Analysis*)	■	■	■
Stablo rizika (*engl. FTA – Fault tree analysis*)	■	■	■
Leptir mašna metoda (*engl. Bow tie analysis*)		■	■
Monte Carlo simulacija			■
Stablo događaja (*engl. Event tree analysis*)	■	■	
Analiza troškova i koristi (engl. *Cost-benefit analysis*)	■	■	■

2.7. Vrste troškova kvalitete

Troškove kvalitete možemo podijeliti u dvije kategorije:

- troškovi kontrole kvalitete (troškovi ostvarivanja kvalitete)
- troškovi nedostatne kontrole kvalitete (troškovi odstupanja od kvalitete).

Troškovi kontrole kvalitete nastaju zbog osiguravanja da učinak (proizvod ili usluga) zadovoljava određene specifikacije, a troškovi nedostatka kontrole kvalitete nastaju kad proizvod ili usluga ne zadovoljava prethodno definirane specifikacije [10].

Gornja podjela troškova kvalitete može se dodatno podijeliti tako da u prvu kategoriju, troškova ostvarivanje kvalitete ulaze [1]:

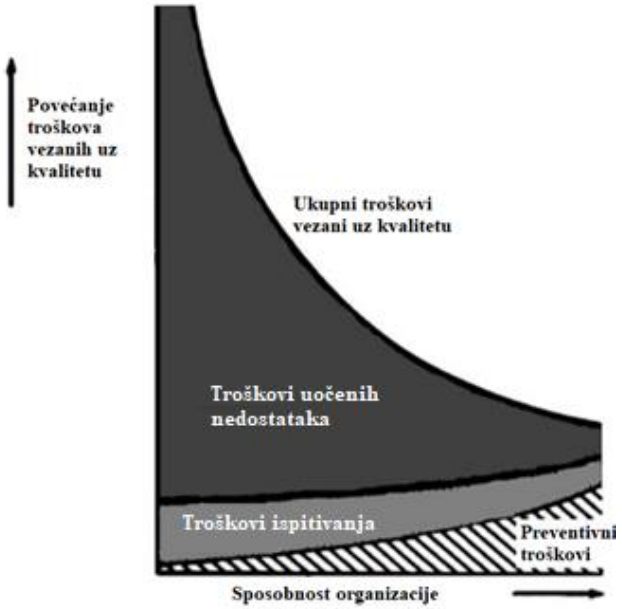
- Troškovi prevencije – troškovi koji nastaju u fazama planiranja proizvoda, konstruiranja proizvoda i konstruiranja proizvodnog procesa, tj. u fazama neizravne kontrole kvalitete. Primjeri preventivnih aktivnosti su: planiranje kvalitete, projektiranje unapređenja kvalitete, konstruiranje i razvoj proizvoda i/ili procesa, obuka i trening zaposlenika, itd.
- Troškovi ispitivanja (ocjene) – troškovi povezani s procjenom kvalitete kupljenog materijala, troškovi održavanja, nadzora i optimizacije proizvodnog procesa kojima se osigurava postizanje zadanih specifikacija kvalitete. U ovu vrstu troškova uključeni su troškovi koji nastaju: ispitivanje kupljenog materijala, preventivne kontrole kvalitete međuproizvoda tijekom proizvodnog procesa, umjeravanjem i ispitivanjem opreme rabljene u proizvodnom procesu (ili pružanja usluge), kontrole kvalitete konačnog proizvoda ili usluge, te njima slični troškovi.

Drugoj kategoriji troškova ostvarivanja kvalitete pripadaju:

- Troškovi interno uočenih nedostataka – troškovi koji se javljaju kada proizvodi ili usluge ne ispunjavaju specificirane granice kvalitete, a otkriju se prije isporuke kupcu. Primjeri ovih troškova su: škart, troškovi popravaka, troškovi ponovljenog ispitivanja, troškovi obavljanja neplaniranih ranije navedenih aktivnosti.
- Troškovi eksterno uočenih nedostataka – troškovi koji se javljaju kada proizvodi ili usluge ne ispunjavaju specificirane granice kvalitete, a otkriju se tek nakon isporuke kupcu. Nakon uočavanja nedostataka na proizvodu ili usluzi, kupac ih reklamira te se za proizvođača javljaju troškovi zamjene ili popravak proizvoda, troškovi povezani s aktivnostima obrade reklamacije kupca, troškovi koji proizlaze iz dane garancije, troškovi koji mogu nastati zbog reputacije.

Vrste troškova u stvarnosti vrlo dobro opisuje dijagram [Slika 7]. Apscisa dijagrama označava sposobnost organizacije da implementira alate inženjerstva kvalitete. Apscisa može označavati i vrijeme jer s vremenom dolazi i do razvoja kompanije i samim time razvoja i odjela za kvalitetu unutar nje. Ordinata predstavlja troškove vezane za kvalitetu proizvoda.

Ako se uzme u obzir da je sposobnost organizacije niska, zamišljena točka na grafu bi se nalazila negdje na početku. Takva organizacija ima male troškove prevencije, što znači da vrlo malo preventivno ulaže u kvalitetu te da provodi kvalitetu kontrole tek na gotovom proizvodu. Zbog toga ima visoke troškove interno i eksterno uočenih nedostataka i samim time i velike ukupne troškove vezane za kvalitetu. Razvojem organizacije i preventivnim ulaganjem, primjerice uvođenjem statističke kontrole procesa, osim kontrole krajnjeg proizvoda kontrolira se i proces proizvodnje. Time se smanjuju troškovi ispitivanja troškovi uočenih nedostataka te ukupni troškovi vezani uz kvalitetu. Daljnjim razvojem i preventivnim ugrađivanjem kvalitete u razvoj i dizajn procesa i proizvoda rastu preventivni troškovi, ali smanjuju se ostali i u sumi to daje ukupno smanjenje troškova vezanih uz kvalitetu.



Slika 7. Dijagram "cijena kvalitete - sposobnost organizacije" [1]

3. R PROGRAMSKI JEZIK

Tehnike i alati koji se primjenjuju u inženjerstvu kvalitete temelje se na podacima o određenim procesima. Strategija prikupljanja i obrade podataka trebala bi biti važan dio planiranja kontrole kvalitete, jer će se sve sljedeće aktivnosti temeljiti na takvim podacima. Kad su podaci prikupljeni, potrebno je pronaći odgovarajuće računalne alate za njihovu analizu. Primjena statističkih metoda u inženjerstvu kvalitete zahtijeva upotrebu specijaliziranog softvera. Naravno, proračunske tablice mogu se koristiti za neke zadatke, ali ako se radi o ozbiljnoj i kompleksnoj analizi podataka unutar inženjerstva kvalitete, potrebni su napredniji alati. Proračunske tablice i dalje mogu biti korisne za unos neobrađenih podataka, ispravljanje pogrešaka ili izvoz rezultata za daljnju upotrebu. Općenito postoji širok raspon softverskih paketa za statistiku. Većina ih uključuje specifične mogućnosti za kontrolu kvalitete, kao što su kontrolne karte ili analiza sposobnosti procesa. Pretraživanjem internetskih stranica može se vidjeti da je većina softvera koje posjeduju ranije navedene zahtjeve komercijalnog tipa, odnosno potrebno je kupiti licencu za njihovo korištenje. S druge strane, danas postoji sve više besplatnih softvera otvorenog koda (engl. *Free and Open Source Software* – FOSS), koji su dostupniji korisnicima. R softver je jedan od najpopularnijih FOSS-a za statističko modeliranje i analizu, koriste ga velike tvrtke poput Googlea, The New York Timesa i sl. [7].

3.1. Razvoj

Ross Ihaka i Robert Gentleman su 1991. godine na temeljima S programskog jezika koji je u osamdesetim godinama prošlog stoljeća korišten u laboratorijima Bell Telephonea napravili R programski jezik. Katedra za statistiku Sveučilišta u Aucklandu, čiji su Ihaka i Gentleman bili članovi 1993. godine prvi put je objavila da radi na razvoju R programskog jezika, razvoj je potrajao sve do 2000. godine kada je R verzija 1.0.0 puštena u javnost kao besplatni programski jezik. Na početku se R program uglavnom koristio u akademskim istraživanjima, a s vremenom se počeo koristiti i u privatnim tvrtkama i javnoj upravi [11].

3.2. Prednosti i nedostaci

Iako je jedan od najkorištenijih programskih jezika u području statistike, R softver uz veliki broj prednosti ima i nedostatke koji se pokušavaju ukloniti svakom novom verzijom. Kroz prednosti i nedostatke bit će prikazana obilježja samog softvera [12].

Prednosti R programskog jezika:

1) Besplatan softver otvorenog koda

   Prva prednost i možda najznačajnije obilježje R softvera je da se radi o besplatnom softveru otvorenog koda. To znači da svatko može raditi u R softveru bez potrebe za licencom ili plaćanjem naknade. Također, bilo tko može pridonijeti razvoju programa prilagođavanjem postojećih ili razvojem novih paketa, kao i rješavanjem problema.

2) Neovisnost o platformi

   R je programski jezik za više platformi. Kod se može izvoditi na svim operativnim sustavima, npr. Windows, Linux ili Mac. R omogućava programerima da razviju softver za nekoliko konkurentskih platformi pišući programski kod samo jednom.

3) Paketi nadogradnje (engl. *Packages*)

   Paketi su nizovi programskog koda koji dodaju nove funkcije unutar R programa, oni funkcioniraju poput malih softvera unutar glavnog softvera. Više o paketima u poglavlju 3.5.

4) Priprema podataka (engl. *Data wrangling*)

   Priprema podataka sastoji se od nekoliko faza: prikupljanje podataka, procjena podataka, konsolidacija i čišćenje, odabir podataka te transformacija. Ova faza unutar procesa rudarenja podataka (engl. *Data mining*) može oduzimati i do 90% od ukupnog vremena.[13] Pomoću programskih paketa poput `dplyr` i `readr` R softver svojim korisnicima omogućuje da u vrlo kratkom vremenu pretvore neuredne podatke u strukturirani oblik.

5) Operacije strojnog učenja

   R softver svojim korisnicima omogućuje izvedbu raznih operacija strojnog učenja kao što su klasifikacija i regresija. U tu svrhu, R podržava razne pakete nadogradnje za razvoj umjetne neuronske mreže.

6) Kompatibilnost

   R je vrlo kompatibilan i može se upariti s mnogim drugim programskim jezicima poput C, C ++, Jave i Pythona. Također se može integrirati s tehnologijama poput Hadoopom i raznim drugim sustavima za upravljanje bazama podataka.

7) Grafički prikaz i vizualizacija

Pomoću programskih paketa kao što su `ggplot2` i `plotly` R softver na vrlo jednostavan način vizualizira podatke u kvalitetne grafičke prikaze što ga izdvaja od konkurentskih programskih jezika.

Nedostaci R programskog jezika:

1) Rukovanje podacima

   R programski jezik nije idealno rješenje kada korisnici imaju posla s velikim brojem podataka jer koristi više memorije u usporedbi s konkurentskim programskim jezicima. Podaci se pohranjuju u fizičku memoriju te zahtijeva da svi podaci budu pohranjeni na jednom mjestu, što nije slučaj primjerice u Pythonu. Međutim, koristeći odgovarajuće paketima za upravljanje podacima i integracijom s već spomenutim Hadoop-om moguće je popraviti ovaj nedostatak.

2) Osnovna sigurnost

   R softver nema osnovni sustav za sigurnost što je još jedan od nedostataka u usporedbi s najbližom konkurencijom. Iz tog razloga postoje mnoga ograničenja za R softver, a najveće od njih je da programske kodove pisane u R softveru nije moguće ugraditi u web aplikacije.

3) Sporija brzina procesiranja podataka

   Programski jezik R mnogo je sporiji u procesiranju podataka u usporedbi s konkurentskim programskim jezicima kao što su Python i MATLAB.

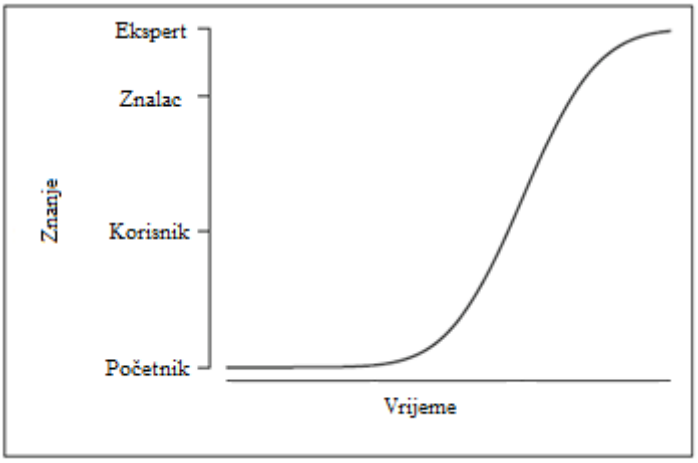
4) Loše podrijetlo

   Kao što je spomenuto R softver dijeli svoje podrijetlo s mnogo starijim programskim jezikom „S“. Iz tog razloga u osnovnom paketu nema podršku za dinamičke ili 3D grafičke prikaze, no i tom problemu je razvojni tim R softvera doskočio paketima za nadogradnju kao što su `ggplot2` ili `plotly`.

5) Kompliciran jezik

   R softver ima kompliciran programski jezik te je krivulja učenja na početku vrlo spora i potrebno je dosta vremena da se nauče osnove [Slika 8.]. To može biti obeshrabrujuće za početnike, posebno kada postoji vremenski okvir u kojem se mora nešto naučiti. No ovaj problem ima i drugo gledište. U odnosu na R softver, konkurentski programski jezici koji imaju više „user-friendly“ grafička korisnička sučelja ne tjeraju korisnika da razmišlja o tome što zapravo radi jer postoji mehanički

slijed klikova koji radi za korisnika. U R softveru da bi korisnici mogli napisati što žele da stroj radi, moraju znati napisati programski kod koji je u mogućnosti to i izvesti. Nakon što se nauče osnove jezika i prihvati drugačiji način rada (pisanje koda, a ne kliktanje ili „drag and drop“) krivulja učenja strmo raste i učenje postaje puno brže.



Slika 8. Krivulja učenja R programskog jezika [4]

3.3. Instalacija

Službena web stranica glavni je izvor informacija za početak rada [Slika 9.]. U središnjem dijelu početne stranice moguće je pronaći dva bloka informacija:

- Početak rada: sadrži poveznice do stranica za preuzimanje programa i do odgovora na često postavljana pitanja
- Vijesti: blok s najnovijim informacijama o R softveru kao što su nove verzije softvera, konferencije i izdanja R Journala. Na slici se može vidjeti da se trenutno koristi verzija softvera R version 4.0.2. (Taking Off Again) koja je korištena i u radu.

S lijeve strane nalazi se izbornik u kojem je dostupan veliki broj poveznica na kojima se mogu naći osnovne informacije o R softveru, također informacije o osnivanju, donatorima, konferencijama, alatima, itd.

Slika 9. R softver - službena web stranica [14]

Za postavljanje programa i početak rada u R softveru potrebno je:

1. preuzeti i instalirati R
2. preuzeti i instalirati RStudio
3. pokrenuti RStudio.

Slika 10. Razlika između R-a i RStudija [14]

Važno je razumjeti razliku između R-a i RStudija. R je programski jezik koji se koristi za statističko računanje, dok Rstudio je aplikacija bazirana na Javi koja koristi R programski jezik za razvoj statističkih programa. U R-u korisnik može napisati programski kod i pokrenuti ga neovisno o bilo kojem drugom računalnom programu, dok se RStudio mora koristiti uz R program kako bi pravilno funkcionirao. Za RStudio se kaže da je integrirano razvojno okruženje (engl. *integrated development environment* – IDE) jer omogućava

korisnicima da razvijaju i uređuju programe na R programskom jeziku koristeći veliki broj statističkih paketa, kvalitetniju grafiku i bolje korisničko sučelje. Zaključno, R i RStudio nisu različite verzije istog programa i ne mogu se međusobno zamijeniti. R se može koristiti bez RStudia, ali se RStudio ne može koristiti bez R programa [15].

3.4. Korisničko sučelje

Ranije objašnjena razlika između R programa i RStudio programa slikovito se može prikazati na način da R program predstavlja motor automobila, a RStudio volan i upravljačku ploču. Stoga će se u nastavku koristiti RStudio kao alat za statističko modeliranje i analizu.

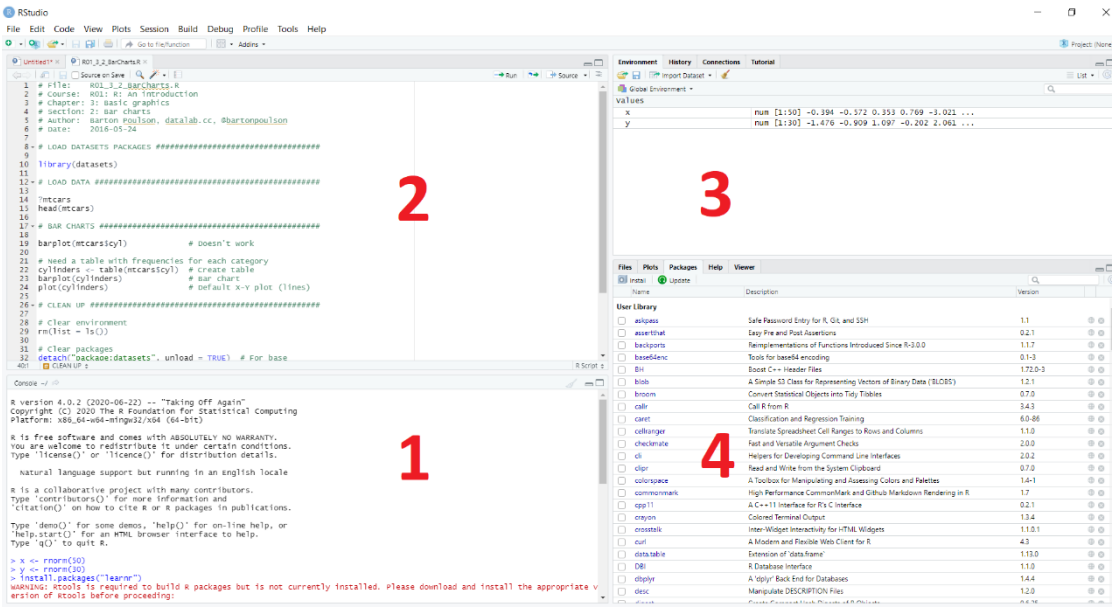
Korisničko sučelje ima 4 prozora [Slika 11.] čije se dimenzije mogu prilagođavati [7]:

1. Upravljački okvir (engl. *Console*) – upravljački okvir u RStudiju nalazi se prema zadanim postavkama u donjem lijevom prozoru. Simbol „>“ predstavlja početak upita, odnosno mjesto gdje korisnik upisuje izraz koji nakon toga RStudio procesira i daje rezultat. Ukoliko softver procjeni da je izraz nepotpun pojavljuje se simbol „+“, koji sugerira da se očekuje ostatak izraza.
2. Okvir za uređivanje koda (engl. *Source Editor*) – okvir za uređivanje koda u RStudiju nalazi se prema zadanim postavkama u gornjem lijevom prozoru. U tom okviru korisnik može stvarati tekstualne datoteke s R izrazima, odnosno pisati programski kod koji može biti dulji od jednog retka. Takve datoteke se nazivaju R Skripte. Programske naredbe odvajaju se znakom točka-zarez (;), a komentari se mogu uključiti pomoću znaka #. Razlog korištenja skripti je ponovna upotreba koda kojeg korisnik jednom napiše. Također u ovom prozoru se mogu otvoriti i druge vrste datoteka, primjerice tekstualne datoteke koje sadrže podatke, programski kod pisan u drugim jezicima (Python, C++, itd.) ili datoteke izvještaja.
3. Prema zadanim postavkama u gornjem desnom prozoru imamo četiri različite kartice:
    1) Okruženje (engl. *Environment*) – mjesto gdje se pohranjuju objekti koje stvaramo putem naredbi u programskom kodu.
    2) Povijest (engl. *History*) – sadrži sve izraze koji su procesuirani u prošlosti te se oni mogu slati u prozore Console i Source Editor. Povijesti se također može pristupiti putem R programskog koda funkcijama `loadhistory`, `savehistory`, `history.`
    3) Veze (engl. *Connections*) – kartica služi za povezivanje na baze podataka.

4) Upute za korištenje (engl. *Tutorial*) – u novijim verzijama softvera dostupni su pokazni primjeri koji početnicima pomažu pri učenju-

4. Prema zadanim postavkama u donjem desnom prozoru imamo pet različitih kartica:

1) Datoteke (engl. *Files*) – predstavlja sistemski pretraživač datoteka u kojemu se može odabrati mapa u koju će biti pohranjene datoteke napravljene u RStudiju.
2) Grafički podaci (engl. *Plots*) – predstavlja grafički uređaj RStudija, odnosno mjesto gdje se generiraju grafikoni
3) Paketi (engl. *Packages*) – prikazuje pakete dostupne u sustavu koji se mogu instalirati, deinstalirati ili ažurirati
4) Pomoć (engl. *Help*) – omogućuje pristup cijeloj R dokumentaciji, uključujući dokumentaciju instaliranih dodatnih paketa
5) Preglednik (engl. *Viewer*) – koristi se za razvoj web aplikacija



Slika 11. Korisničko sučelje softvera RStudio

3.5. Paketi nadogradnje

Za pakete već je spomenuto da su nizovi programskog koda koji dodaju nove funkcije unutar R programa te da funkcioniraju poput malih softvera unutar glavnog softvera. Slikovito se može reći da su to super moći koje korisniku RStudija omogućavaju gotovo sve mogućnosti u statističkoj analizi i modeliranju.

R ima širok spektar paketa koji se dijele u dvije kategorije:

- Osnovni – instalirani su tijekom prve instalacije R softvera, ali prema zadanim postavkama nisu učitani te ih je potrebno učitati
- Dodatni – potrebno ih je preuzeti, instalirati te učitati.

Dodatni paketi korisnicima daju više mogućnosti, a moguće ih je preuzeti sa sljedećih web stranica:

- CRAN (Comprehensive R Archive Network) – službena R web stranica za preuzimanje paketa nadogradnje sa svom popratnom dokumentacijom
- GitHub – sadrži različite popise po kojima korisnik pretražuje pakete nadogradnje (npr. lista najpopularnijih, lista posljednje objavljenih, itd.).

Neki od najpoznatijih dodatnih paketa za nadogradnju R softvera su:

- `tidyr` – čišćenje nepotrebnih podataka
- `dplyr` – manipuliranje sa različitim tipovima podataka
- `stringr` – za rad s tekstualnim podacima
- `lubridate` – za rad s raznim formatima datuma
- `httr` – za rad s podacima s web stranica
- `ggplot2` – najkorišteniji paket za kreiranje grafova i vizualizaciju podataka
- `rio` – koristi se za uvoz i izvoz podataka
- `qcc` – alat za kontrolne karte i statističku kontrolu procesa.

4. ALATI KONTROLE KVALITETE

Kaoru Ishikawa jedan je od najznačajnijih ličnosti u području kontrole kvalitete. On je izabrao set od sedam jednostavnih alata čijim korištenjem se može riješiti većina standardnih problema s kvalitetom i produktivnošću procesa [7]. Za svaki alat bit će prikazan primjer koristeći R programski jezik u RStudio programu.

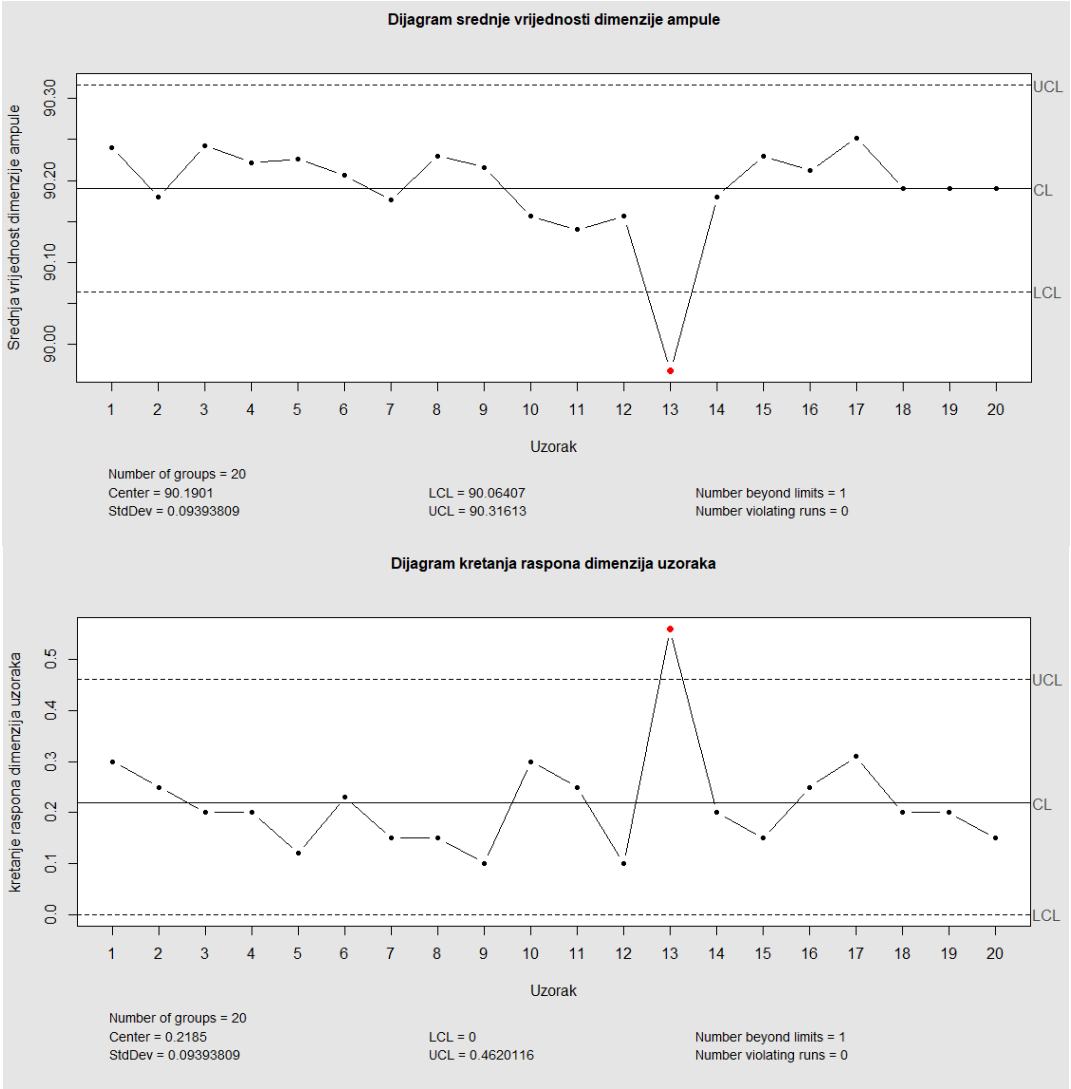
4.1. Kontrolne karte (engl. *Control Chart*)

Kontrolne karte su grafički prikazi uzorkovanih varijabli nekog procesa tijekom vremena. Nakon što izmjerene vrijednosti varijabli prijeđu kontrolne granice detektiraju se tzv. neuobičajene varijable čiji se značajni uzroci trebaju otkriti i eliminirati.

Historijski podaci koji su u nastavku učitani bit će korišteni za prikaz nekih od alat kontrole kvalitete. U ovom primjeru korišteni su podaci o dimenzijama ampule. Kontrola dimenzije ampule se provodi tako da se iz populacije izuzme uzorak te se mjere dimenzije. Svaki uzorak sadrži 5 ampula.

```
library(readxl)
Podaci_za_alate <- read_excel("C:/Users/Marko Jukic/Downloads/Podaci za
alate.xlsx",
    col_types = c("numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "text"))
library(qcc)

dimenzija_uzorak <- qcc.groups(
  data = Podaci_za_alate$dimenzija_ampule,
  sample = Podaci_za_alate$broj_uzorka)
X_graf <- qcc(data = dimenzija_uzorak,
              type = "xbar",
              xlab = "Uzorak",
              ylab = "Srednja vrijednost dimenzije ampule",
              title = "Dijagram srednje vrijednosti dimenzije ampule")

R_graf <- qcc(data = dimenzija_uzorak,
              type = "R",
              xlab = "Uzorak",
              ylab = "kretanje raspona dimenzija uzoraka",
              title = "Dijagram kretanja raspona dimenzija uzoraka")
```

Nakon učitavanja podataka aktiviran je ranije spomenuti programski paket `qcc`. Pomoću istoimene funkcije na osnovu podataka o dimenziji ampula i broju uzoraka program stvara kontrolnu kartu tipa `xbar` koji prikazuje srednje vrijednosti kontinuirane varijable te kontrolnu kartu tipa `R` koja prikazuje kretanje raspona dimenzija uzorka.



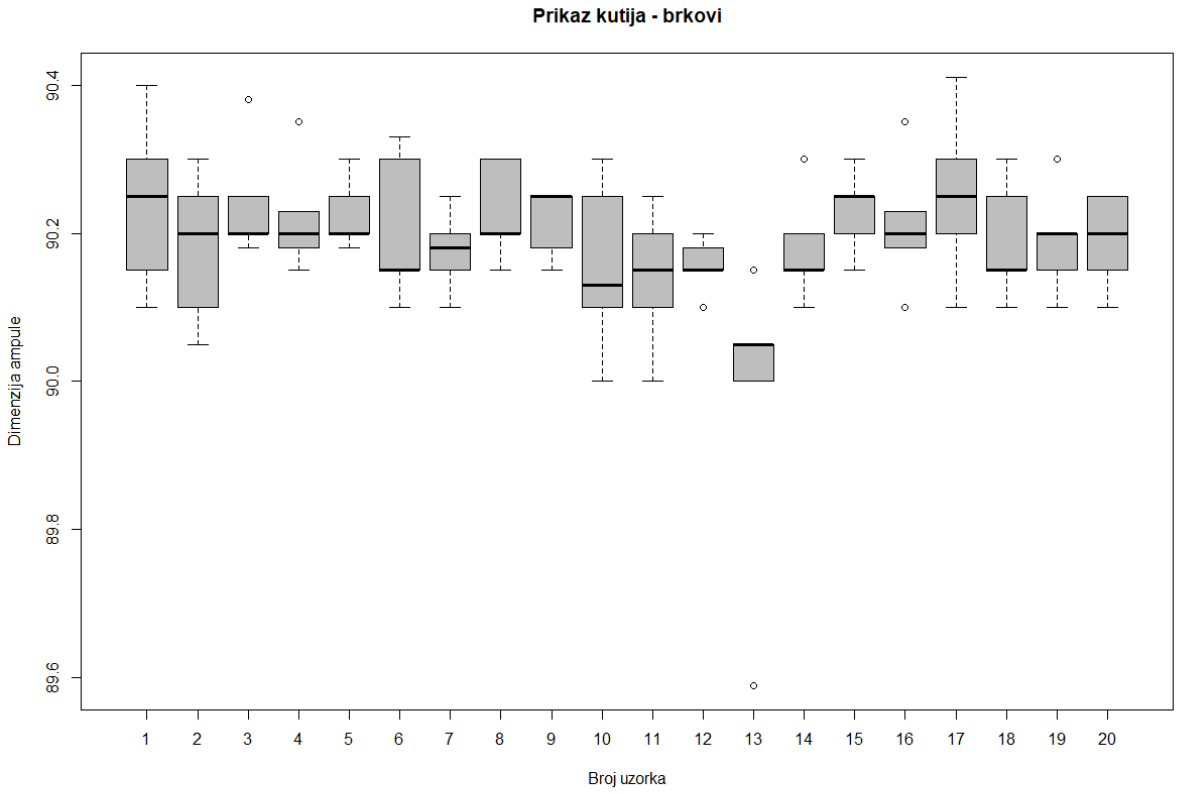
Slika 12. $\bar{x}$-*R* kontrolna karta

Iz kontrolne karte [Slika 12.] vidljivo je da postoji jedno mjerenje (broj 13) označeno crvenom bojom, koje se nalazi izvan kontrolnih granica. To ukazuje na mogućnost postojanja značajnih uzroka varijacija te je potrebno provesti daljnju analizu proizvodnog procesa.

4.2. Prikaz kutija-brkovi (engl. *Box-whisker plot*)

Prikaz kutija-brkovi grafički sažima raspodjelu uzorka kontinuirane varijable i omogućuje jednostavnu dijagnostiku problematičnih podataka (ekstrema, „outliera“). U primjeru su korišteni podaci o dimenzijama ampula kao i u prethodnom zadatku.

```
boxplot(Podaci_za_alate$dimenzija_ampule ~ Podaci_za_alate$broj_uzorka,
        col = "gray",
        xlab = "Broj uzorka",
        ylab = "Dimenzija ampule",
        main = "Prikaz kutija - brkovi")
```



Slika 13. Prikaz kutija-brkovi

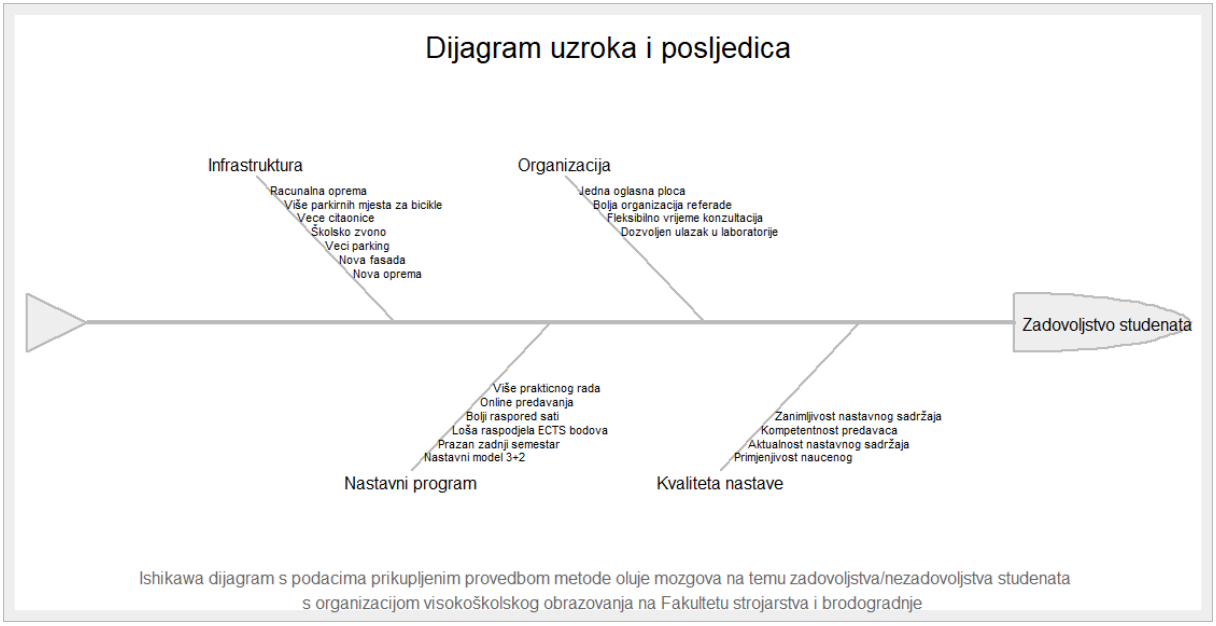
Slika 13. predstavlja prikaz kutija-brkovi. Gornja i donja strane kutija predstavljaju prvi i treći kvartil. Medijan je prikazan kao crna crta preko kutije. Ovakav dijagram otkriva je li raspodjela centrirana ili pristrana (položaj medijana u odnosu na ostale podatke), postoje li odstupanja (točke izvan brkova koje predstavljaju „outliere“) ili su podaci blizu središnjih vrijednosti (male kutije i brkovi).

4.3. Dijagram uzorka i posljedica (*engl. Cause and effect diagram*)

Dijagram uzroka i posljedica poznat i kao Ishikawa dijagram ili dijagram riblje kosti daje vizualni prikaz faktora koji najvjerojatnije pridonose promatranom problemu ili učinku na proces. Podaci koji su korišteni u nastavku prikupljeni su provedbom metode Oluje mozgova na temu zadovoljstva/nezadovoljstva studenata s organizacijom visokoškolskog obrazovanja

na Fakultetu strojarstva i brodogradnje. Dobiveni rezultati upisani su u programskom kodu te je pomoću programskog paketa `SixSigma` napravljen dijagram uzroka i posljedica [Slika 14.].

```
cInfrastruktura <- c("Računalna oprema","Više parkirnih mjesta za
bicikle", "Veće čitaonice","Školsko zvono","Veći parking", "Nova fasada",
"Nova oprema")
cOrganizaicja <- c("Jedna oglasna ploča", "Bolja organizacija referade",
"Fleksibilno vrijeme konzultacija", "Dozvoljen ulazak u laboratorije")
cKvaliteta_nastave <- c("Primjenjivost naučenog", "Aktualnost nastavnog
sadržaja", "Kompetentnost predavača","Zanimljivost nastavnog sadržaja")
cNastavni_program <- c("Nastavni model 3+2", "Prazan zadnji semestar",
"Loša raspodjela ECTS bodova", "Bolji raspored sati", "Online predavanja",
"Više praktičnog rada")
cEfekt <- "Zadovoljstvo studenata"
cGroup <- c("Infrastruktura", "Organizacija", "Kvaliteta nastave",
"Nastavni program")

library(SixSigma)
ss.ceDiag(
  effect = cEfekt,
  causes.gr <- cGroup,
  causes = list(cInfrastruktura, cOrganizaicja, cKvaliteta_nastave,
cNastavni_program),
  main = "Dijagram uzroka i posljedica",
  sub = "Ishikawa dijagram s podacima prikupljenim provedbom
Brainstorminga na temu zadovoljstva/nezadovoljstva studenata
  s organizacijom visokoškolskog obrazovanja na Fakultetu strojarstva i
brodogradnje")
```



Slika 14. Dijagram uzroka i posljedica

4.4. Kontrolni list (engl. *Check Sheet*)

Kontrolni list je važan alat za prikupljanje podataka. On omogućuje vertikalnu razmjenu informacija i podataka od zaposlenika koji s nalaze niže na hijerarhijskoj ljestvici i temeljito znaju pojedine dijelove procesa do zaposlenika koji upravljaju tim procesima. U praksi gledano, iako se većina podataka obnavlja automatski, neki se podaci prikupljaju ručno te se kasnije unose u bazu podataka ili proračunsku tablicu kako bi se mogli analizirati u idućem koraku.

```
cGlodanje <- c("Greška broj 1", "Greška broj 2", "Greška broj 3")
cBusenje <- c("Greška broj 1", "Greška broj 2", "Greška broj 3")
cTokarenje <- c("Greška broj 1", "Greška broj 2", "Greška broj 3")
data_checkSheet <- rbind(
  data.frame(Group = "Glodanje",
             Cause = cGlodanje),
  data.frame(Group = "Bušenje",
             Cause = cBusenje),
  data.frame(Group = "Tokarenje",
             Cause = cTokarenje))
data_checkSheet$Frekvencija <- NA

source ("kontrolni_list.R")
library (xtable)
print (xtable(data_checkSheet), type = "HTML",
       html.table.attributes =
         "border=1 width = 100% cellpadding = 10")
```

	Group	Cause	Frekvencija
1	Glodanje	Greska broj 1	
2	Glodanje	Greska broj 2	
3	Glodanje	Greska broj 3	
4	Busenje	Greska broj 1	
5	Busenje	Greska broj 2	
6	Busenje	Greska broj 3	
7	Tokarenje	Greska broj 1	
8	Tokarenje	Greska broj 2	
9	Tokarenje	Greska broj 3	

Datum **Operater** **Potpis**

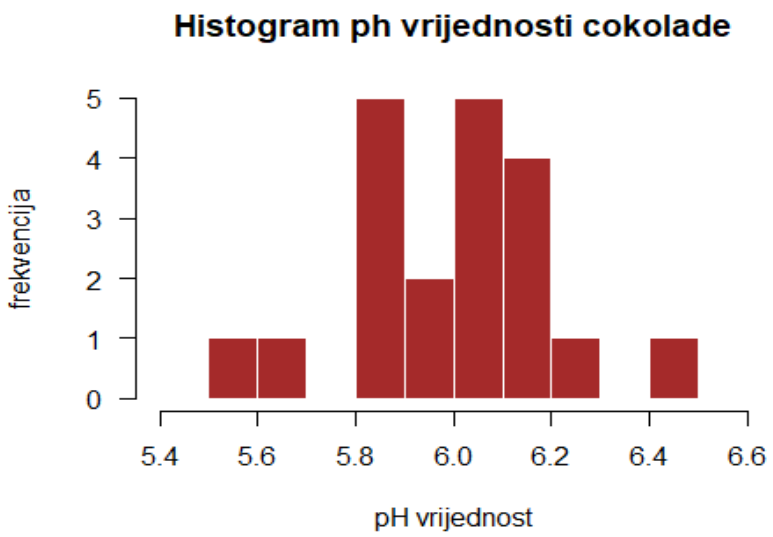
Slika 15. Kontrolni list

Funkcija `rbind` kombinira niz vektora, matrica ili drugih vrsta datoteka koje su znakovnog tipa (ne smiju biti izrazi) u stupce ili redove. Dodatni se stupci mogu dodati za registraciju događaja prema različitim kriterijima. U ovom slučaju radi se o kontroli kvalitete nakon tri faze proizvodnog procesa, svaka faza ima tri karakteristične greške te se na kontrolnoj karti označava frekvencija grešaka. Za formiranje tablica u .html i .pdf obliku koristi se `xtable` programski paket. Tablica [Slika 15.] je u .html obliku.

4.5. Histogram

Histogrami prikazuju frekvenciju ponavljanja određene varijacije u procesu. U histogramu se dokumentiraju varijacije tako da se podaci svrstavaju u stupce koji predstavljaju rezultate mjerenja podijeljene u jednake intervale. U ovom primjeru korišteni su historijski podaci koji su učitani na početku poglavlja. Napravljen je histogram [Slika 16.] prema mjerenjima pH vrijednosti čokolade.

```
hist(Podaci_za_alate$pH,
     breaks = c(5.4+0.1*(0:12)),
     main ="Histogram ph vrijednosti čokolade",
     xlab ="pH vrijednost",
     ylab = "frekvencija",
     las = 1,
     col = "brown",
     border = "white")
```

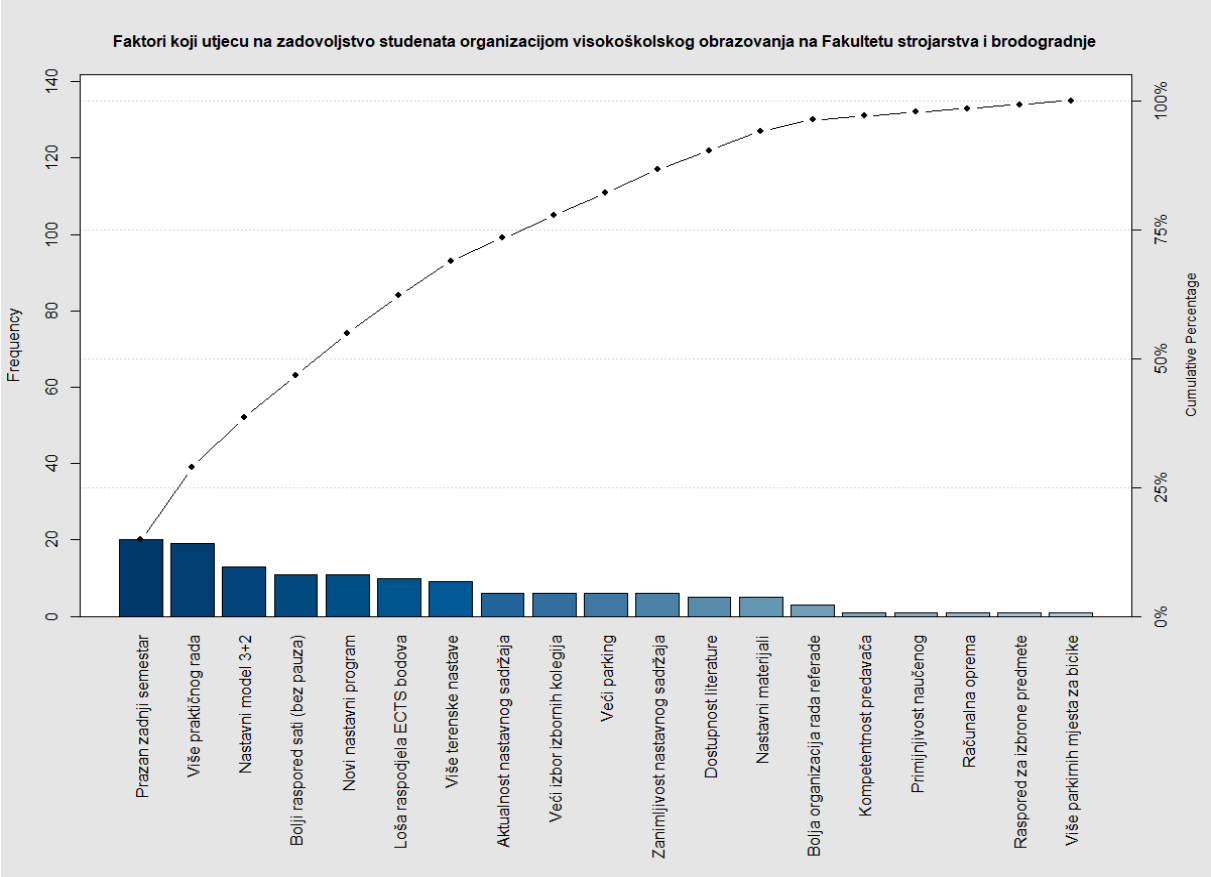


Slika 16. Histogram

4.6. Pareto dijagram

Pareto dijagrami se fokusiraju na čimbenike koji uzrokuju najviše problema u procesu. Poznat je Paretov princip ili pravilo 80-20 u kojem 20 % glavnih uzroka proizvodi 80 % problema. Iz Pareto dijagrama lako se mogu prepoznati faktori koji imaju najveći potencijal za poboljšanje. U ovom slučaju korišteni su podaci iz dijagrama uzroka posljedica [Slika 14.] kojima su dodijeljene vrijednosti odnosno bodovi. Stupci u histogramu [Slika 17.] koji predstavljaju čimbenike kreću se s lijeva na desno prema redoslijedu silazne važnosti, mjereno utjecajem čimbenika na zadovoljstvo studenta.

```
par(mar = c(9, 4, 4, 2 + 0.1))
Pareto2 <- Pareto$Bodovi
names(Pareto2) <- Pareto$Faktor
pareto.chart(data =  Pareto2,
             main = "Faktori koji utječu na zadovoljstvo studenata
organizacijom visokoškolskog obrazovanja na Fakultetu strojarstva i
brodogradnje")
```



Slika 17. Pareto dijagram

```
##
## Pareto chart analysis for Pareto2
##                                  Frequency   Cum.Freq.  Percentage
##   Prazan zadnji semestar         20.0000000  20.0000000  14.8148148
##   Više prakticnog rada           19.0000000  39.0000000  14.0740741
##   Nastavni model 3+2             13.0000000  52.0000000   9.6296296
##   Bolji raspored sati (bez pauza) 11.0000000  63.0000000   8.1481481
##   Novi nastavni program          11.0000000  74.0000000   8.1481481
##   Loša raspodjela ECTS bodova    10.0000000  84.0000000   7.4074074
##   Više terenske nastave           9.0000000  93.0000000   6.6666667
##   Aktualnost nastavnog sadržaja   6.0000000  99.0000000   4.4444444
##   Veci izbor izbornih kolegija    6.0000000 105.0000000   4.4444444
##   Veci parking                    6.0000000 111.0000000   4.4444444
##   Zanimljivost nastavnog sadržaja 6.0000000 117.0000000   4.4444444
##   Dostupnost literature           5.0000000 122.0000000   3.7037037
##   Nastavni materijali             5.0000000 127.0000000   3.7037037
##   Bolja organizacija rada referade 3.0000000 130.0000000   2.2222222
```

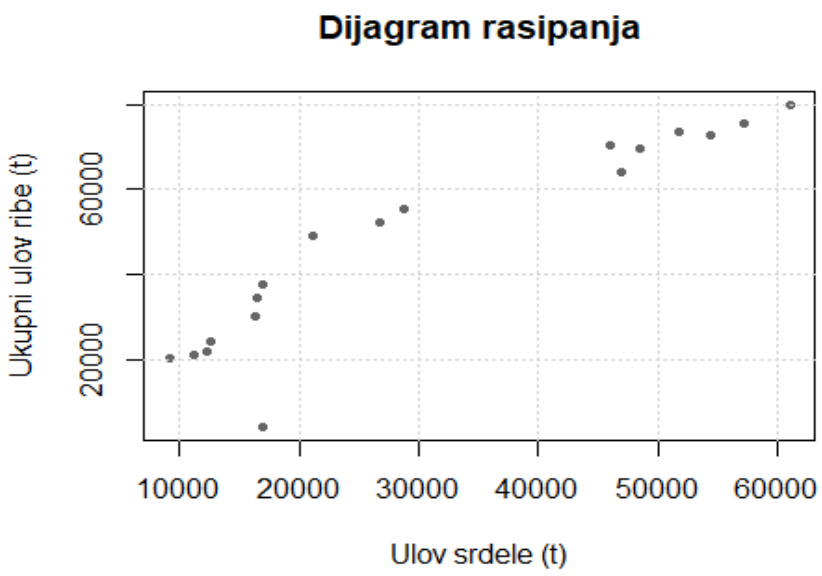
```
##   Kompetentnost predavaca             1.0000000 131.0000000   0.7407407
##   Primijnjivost naucenog              1.0000000 132.0000000   0.7407407
##   Racunalna oprema                    1.0000000 133.0000000   0.7407407
##   Raspored za izbrone predmete        1.0000000 134.0000000   0.7407407
##   Više parkirnih mjesta za bicike     1.0000000 135.0000000   0.7407407
##
## Pareto chart analysis for Pareto2
##                                    Cum.Percent.
##   Prazan zadnji semestar             14.8148148
##   Više prakticnog rada               28.8888889
##   Nastavni model 3+2                 38.5185185
##   Bolji raspored sati (bez pauza)    46.6666667
##   Novi nastavni program              54.8148148
##   Loša raspodjela ECTS bodova        62.2222222
##   Više terenske nastave              68.8888889
##   Aktualnost nastavnog sadržaja      73.3333333
##   Veci izbor izbornih kolegija       77.7777778
##   Veci parking                       82.2222222
##   Zanimljivost nastavnog sadržaja    86.6666667
##   Dostupnost literature              90.3703704
##   Nastavni materijali                94.0740741
##   Bolja organizacija rada referade   96.2962963
##   Kompetentnost predavaca            97.0370370
##   Primijnjivost naucenog             97.7777778
##   Racunalna oprema                   98.5185185
##   Raspored za izbrone predmete       99.2592593
##   Više parkirnih mjesta za bicike   100.0000000
```

Kako se u analizi koristi 19 faktora, približno 4 faktora iznosi 20 % od ukupnog broja faktora koji bi prema Paretovom principu trebali većinskim udjelom utjecati na zadovoljstvo studenata. Prema izvještaju iz R programskog jezika koji prikazuje kumulativne vrijednosti vidljivo je da 4 najfrekventnija faktora imaju utjecaj od 46,67 %, što je značajan utjecaj te bi se njihovim rješavanjem uvelike utjecalo na zadovoljstvo studenata. S druge strane, utjecajem na 4 faktora s najmanjom frekvencijom zadovoljstvo studenata bi se povećalo samo 3 %.

4.7. Dijagram rasipanja (engl. *Scatter Plot*)

Dijagrami rasipanja grafički prikazuju odnosa između dvije varijable kao niz točaka. Raspon mogućih vrijednosti za svaku varijablu predstavljen je apscisom i ordinatom. Točke predstavljaju odnos dviju varijabli. Dijagram rasipanja pokazuje postoji li korelacija između varijabli. Podaci za ovaj primjer preuzeti su s Državnog zavoda za statistiku Republike Hrvatske. Ukupni ulov i uzgoj ribe u tonama predstavlja jednu varijablu dijagrama, dok ukupni ulov ribe srdele u tonama predstavlja drugu varijablu u razdoblju od 2001. do 2017. godine.

```
library(readxl)
Uzgoj_ribe <- read_excel("C:/Users/Marko Jukic/Desktop/Uzgoj_ribe.xlsx",
    col_types = c("numeric", "numeric"))
View(Uzgoj_ribe)

plot(Uzgoj_ribe$UKUPNO ~ Uzgoj_ribe$Srdela,
     col = "gray40",
     pch = 20,
     main = "Dijagram rasipanja",
     xlab = "Ulov srdele (t)",
     ylab = "Ukupni ulov ribe (t)")
grid()
```



Slika 18. Dijagram rasipanja

Prema dijagramu [Slika 18.] je vidljivo da rastom vrijednosti jedne varijable raste vrijednost i druge, stoga možemo zaključiti da postoji pozitivna korelacija.

5. PRILAGODBA TEORIJSKE RASPODJELE EMPIRIJSKIM PODACIMA

Teorijske raspodjele su raspodjele koje se mogu očekivati u skladu s našim iskustvom. Pretpostavljamo ih u nekom statističkom modelu ili ih postavljamo kao hipotezu koju treba ispitati. Za njih su unaprijed poznate karakteristike kao što su aritmetička sredina, mod, medijan itd. [16]. Empirijski ili originalni podaci su podaci koji su nastali opažanjem ili istraživanjem.

Prilagodba određene teorijske raspodjele na empirijske podatke te testiranje adekvatnosti iste vrlo je čest zadatak u statistici. Ovoj metodi pribjegava se kada postoji određeni set podataka, te se želi znati koja se raspodjela podataka može koristiti za opisivanje tog seta.

5.1. Faze prilagodbe određene teorijske raspodjele empirijskim podacima

Postupak prilagodbe sastoji se od nekoliko faza koje će biti prikazane u nastavku i prema kojima će biti napravljena dva primjera u programskom jeziku R.

5.1.1. Priprema podataka

Priprema podataka može se sastojati od prikupljanja podataka, procjene podataka, konsolidacije i čišćenja, odabira podataka te transformacija. Sam unosa podataka u R može se obaviti direktnim unosom podataka ili uvozom podataka nekih drugih formata kao što su .csv ili .xlsx. Također, postoje razni R programski paketi koji omogućuju učitavanje podataka napravljenih u drugim statističkim softverima. Primjerice paket `foreign` omogućuje uvoz podataka napravljenih u Minitabu, S-u, SAS-u ili Systatu, dok paket `XML` omogućava čitanje .xml podataka, itd.

5.1.2. Odabir teorijske raspodjele i metode procjene

Programski paket nadogradnje R programskog jezika koji je namijenjen prilagodbi određene teorijske raspodjele empirijskim podacima naziva se `fitdistrplus`. On pruža funkcije za prilagođavanje raznih univarijantnih raspodjela različitim vrstama podataka (diskretne i kontinuirane varijable). Korisnik može odabrati jednu od brojnih teorijski raspodjela kao što su normalna, log-normalna, Posionova, Weibullova, Gumbelova, gamma, trokutasta, pravokutna, itd. Također može izabrati i različite metode procjene koje su opisane u nastavku:

5.1.2.1. Metoda najveće vjerodostojnosti (engl. *Maximum likelihood estimation* – MLE) [17]

Metoda najveće vjerojatnosti jedna je od najboljih metoda za procjenjivanja parametara jer se može primijeniti na najveći dio teorijskih raspodjela. Ova metodu razvio je Sir R. A. Fisher početkom 20. stoljeća. Metoda najveće vjerodostojnosti temelji se na pretpostavci da je poznat oblik funkcije raspodjele osnovnog skupa. Ona se sastoji u traženju one procjene nepoznatog parametra za koji zajednička funkcija vjerodostojnosti uzorka, odnosno realizacije statističkog procesa doseže maksimum.

5.1.2.2. Metoda momenata (engl. *Moment matching estimation* – MME) [17]

Metoda momenata je jedna od prvih metoda za procjenu parametara koju je uveo K. Paerson. Metoda se temelji na tome da su izračunate vrijednosti statističkih momenata na temelju slučajnog uzorka $X_1, X_2, \dots, X_n$ bliske teorijskim vrijednostima momenata vjerojatnosne raspodjele koja je pretpostavljena u danom statističkom modelu.

5.1.2.3. Metoda kvantila (engl. *Quantile matching estimation* – QME) [18]

Funkcija kvantila je definirana kao inverz kumulativne funkcije gustoće. Metodu kvantila predložili su Carmody, Eubank i LaRiccia 1984. godine. To je metoda kojom se procjenjuje udaljenost teorijskih i empirijskih kvantila. Numerička optimizacija se provodi kako bi se minimizirala suma kvadrata razlika između teorijskih i empirijskih kvantila.

5.1.2.4. Metoda temeljena na procjeni dobrote modela (engl. *Maximizing goodness-of-fit estimation* – MGE) [18]

Korištenje ove metode zahtijeva dodatni argument `gof` kodiranja kojim se odabire na koji će način biti računata udaljenost između dobrote empirijskog i teorijskog modela. Najčešće korišteni su Cramer-von Mises ("CvM"), Kolmogorov-Smirnov ("KS") i Anderson-Darling ("AD"). Ova metoda nije prikladna za diskretne raspodjele.

Nakon što je odabrana teorijska raspodjela i metoda procjene, pokreće se programski kod koji kao izlaznu vrijednost daje parametre odabrane raspodjele npr. srednju vrijednost i standardnu devijaciju.

5.1.3. Grafička analiza

Grafička analiza podijeljena je u tri dijela:

1) Usporedba teorijske funkcije gustoće vjerojatnosti s histogramom i empirijskom funkcijom gustoće vjerojatnosti
2) Usporedba teorijskih i empirijskih kvantila
3) Usporedba teorijske i empirijske kumulativne funkcije raspodjele

5.1.4. Numerička analiza

Naredba `gofstat` ranije spomenutog programskog paketa `fitdistrplus` predložena je za numeričko testiranje adekvatnosti odabrane raspodjele. Rezultati ove naredbe sastoje se od dva dijela.

Prvi dio je baziran na procjeni koliko je dobro prilagođena raspodjela podacima (eng. *goodness-of-fit*) i sastoji se od tri testa. To su Kolmogorov-Smirnov, Anderson-Darling i Cramer-von Mises statistički testovi.

Drugi dio se odnosi na kriterij odabira raspodjele i sastoji se od Akaikeovog informacijskog kriterija (engl. *Akaike information criterion* – AIC) i Bayesovog informacijskog kriterija (engl. *Bayesian Information Criterion – BIC*).

U nastavku će ukratko biti opisan svaki od testova i kriterija.

5.1.4.1. Kolmogorov-Smirnov test [19]

Kolmogorov-Smirnov test je neparametarski test koji se temelji na maksimalnoj udaljenosti kumulativnih raspodjela podataka. Postoje dvije vrste Kolmogorov-Smirnov testa, prvi se odnosi na prikladnost uzorka, a drugi na test dvaju uzoraka. Kako će u radu biti potrebno utvrditi pripadaju li empirijski podaci određenoj raspodjeli, koristit će se prva vrsta testa.

Kolmogorov-Smirnov test ima određena ograničenja:

- koristi se samo u slučajevima kontinuirane raspodjele
- osjetljiviji je oko centra raspodjele nego na krajevima
- primjenjiv je samo u slučajevima kada je raspodjela potpuno definirana.

Kolmogorov-Smirnov test je definiran na sljedeći način:

1) H_0: Podaci prate određenu raspodjelu.
2) H_1: Podaci ne prate određenu raspodjelu.
3) Granične vrijednosti [Slika 19.] – hipoteza vezana za vrstu raspodjele se odbacuje ukoliko je parametar testa veći od granične vrijednosti dobivene iz tablice.

U nastavku se nalazi tablica prema kojoj se očitava i izračunava granična vrijednost za Kolmogorov-Smirnov test. Gdje $1-\alpha$ predstavlja interval vjerodostojnosti, a n veličinu uzorka.

n \ α	0.001	0.01	0.02	0.05	0.1	0.15	0.2
1		0.99500	0.99000	0.97500	0.95000	0.92500	0.90000
2	0.97764	0.92930	0.90000	0.84189	0.77639	0.72614	0.68377
3	0.92063	0.82900	0.78456	0.70760	0.63604	0.59582	0.56481
4	0.85046	0.73421	0.68887	0.62394	0.56522	0.52476	0.49265
5	0.78137	0.66855	0.62718	0.56327	0.50945	0.47439	0.44697
6	0.72479	0.61660	0.57741	0.51926	0.46799	0.43526	0.41035
7	0.67930	0.57580	0.53844	0.48343	0.43607	0.40497	0.38145
8	0.64098	0.54180	0.50654	0.45427	0.40962	0.38062	0.35828
9	0.60846	0.51330	0.47960	0.43001	0.38746	0.36006	0.33907
10	0.58042	0.48895	0.45662	0.40925	0.36866	0.34250	0.32257
11	0.55588	0.46770	0.43670	0.39122	0.35242	0.32734	0.30826
12	0.53422	0.44905	0.41918	0.37543	0.33815	0.31408	0.29573
13	0.51490	0.43246	0.40362	0.36143	0.32548	0.30233	0.28466
14	0.49753	0.41760	0.38970	0.34890	0.31417	0.29181	0.27477
15	0.48182	0.40420	0.37713	0.33760	0.30397	0.28233	0.26585
16	0.46750	0.39200	0.36571	0.32733	0.29471	0.27372	0.25774
17	0.45440	0.38085	0.35528	0.31796	0.28627	0.26587	0.25035
18	0.44234	0.37063	0.34569	0.30936	0.27851	0.25867	0.24356
19	0.43119	0.36116	0.33685	0.30142	0.27135	0.25202	0.23731
20	0.42085	0.35240	0.32866	0.29407	0.26473	0.24587	0.23152
25	0.37843	0.31656	0.30349	0.26404	0.23767	0.22074	0.20786
30	0.34672	0.28988	0.27704	0.24170	0.21756	0.20207	0.19029
35	0.32187	0.26898	0.25649	0.22424	0.20184	0.18748	0.17655
40	0.30169	0.25188	0.23993	0.21017	0.18939	0.17610	0.16601
45	0.28482	0.23780	0.22621	0.19842	0.17881	0.16626	0.15673
50	0.27051	0.22585	0.21460	0.18845	0.16982	0.15790	0.14886
OVER 50	$\frac{1.94947}{\sqrt{n}}$	$\frac{1.62762}{\sqrt{n}}$	$\frac{1.51743}{\sqrt{n}}$	$\frac{1.35810}{\sqrt{n}}$	$\frac{1.22385}{\sqrt{n}}$	$\frac{1.13795}{\sqrt{n}}$	$\frac{1.07275}{\sqrt{n}}$

Slika 19. Tablica graničnih vrijednosti - Kolmogorov-Smirnov test [20]

5.1.4.2. Cramer-von Mises test [21]

Cramer-von Mises je neparametarski test koji je usavršeni oblik Kolmogorov-Smirnov testa, također temelji se na usporedbi teorijske i empirijske kumulativne raspodjele podataka. Smatra se preciznijim od Kolmogorov-Smirnov testa te veliki broj analitičara preferira njegovo korištenje.

Cramer-von Mises test definiran je na sljedeći način :

1) H_0: Podaci prate određenu raspodjelu.
2) H_1: Podaci ne prate određenu raspodjelu.
3) Granične vrijednosti [Slika 20.] – hipoteza vezana za vrstu raspodjele se odbacuje ukoliko je parametar testa veći od granične vrijednosti dobivene iz tablice.

U nastavku se nalazi tablica prema kojoj se očitava granična vrijednost za Cramer-von Mises test. Gdje $1 - \alpha$ predstavlja interval vjerodostojnosti, a N veličinu uzorka. Za veličine uzorka veće od 100 uzima se granična vrijednost za $N = 100$.

$N \setminus \alpha$	0.2	0.15	0.1	0.05	0.01
2	0.138	0.149	0.162	0.175	0.186
3	0.121	0.135	0.154	0.184	0.23
4	0.121	0.134	0.155	0.191	0.28
5	0.121	0.137	0.16	0.199	0.3
6	0.123	0.139	0.162	0.204	0.31
7	0.124	0.14	0.165	0.208	0.32
8	0.124	0.141	0.165	0.21	0.32
9	0.125	0.142	0.167	0.212	0.32
10	0.125	0.142	0.167	0.212	0.32
11	0.126	0.143	0.169	0.214	0.32
12	0.126	0.144	0.169	0.214	0.32
13	0.126	0.144	0.169	0.214	0.33
14	0.126	0.144	0.169	0.214	0.33
15	0.126	0.144	0.169	0.215	0.33
16	0.127	0.145	0.171	0.216	0.33
17	0.127	0.145	0.171	0.217	0.33
18	0.127	0.146	0.171	0.217	0.33
19	0.127	0.146	0.171	0.217	0.33
20	0.128	0.146	0.172	0.217	0.33
30	0.128	0.146	0.172	0.218	0.33
60	0.128	0.147	0.173	0.22	0.33
100	0.129	0.147	0.173	0.22	0.34

Slika 20. Tablica graničnih vrijednosti - Cramer-von Mises test[21]

5.1.4.3. Anderson-Darling test [22]

Anderson-Darling test proizlazi iz Kolmogorov-Smirnov testa, s tim da veću važnost pridodaje krajevima raspodjele. Kritične vrijednosti Kolmogorov-Smirnov testa ne ovise o teorijskoj raspodjeli koja se ispituje, dok Anderson-Darling test uzima u obzir o kojoj se teorijskoj raspodjeli radi prilikom računanja kritičnih vrijednosti. To za posljedicu ima veću osjetljivost testa, zbog čega kritične vrijednosti moraju biti izračunate za svaku distribuciju zasebno. Anderson-Darling test je definiran na sljedeći način:

1) H_0: Podaci prate određenu raspodjelu.
2) H_1: Podaci ne prati određenu raspodjelu.
3) Granične vrijednosti [Slika 21.] – granične vrijednosti vezane za Anderson-Darling test ovise o razdiobi koja se testira.

U nastavku se nalazi tablica prema kojoj se izračunava granična vrijednost za Anderson-Darling test. Gdje $1 - \alpha$ predstavlja interval vjerodostojnosti, a n veličinu uzorka.

Distribution [Ref]	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
Normal & lognormal [43]	0.631	0.752	1.035
Weibull [43]	0.637	0.757	1.038
Generalized extreme value [44]	-	-	-
$n = 10$	0.236	0.276	0.370
$n = 20$	0.232	0.274	0.375
$n = 30$	0.232	0.276	0.379
$n = 40$	0.233	0.277	0.381
$n = 50$	0.233	0.277	0.383
$n = 100$	0.234	0.279	0.387
Generalized logistic [44]	-	-	-
$n = 10$	0.223	0.266	0.374
$n = 20$	0.241	0.290	0.413
$n = 30$	0.220	0.301	0.429
$n = 40$	0.254	0.306	0.435
$n = 50$	0.258	0.311	0.442
$n = 100$	0.267	0.323	0.461
Uniform [52] *	1.936	2.499	3.903

Slika 21. Tablica graničnih vrijednosti - Anderson-Darling test [23]

5.1.4.4. Akaikeov i Bayesov informacijski kriterij [24]

Model koji se dobije za neke podatke ne može savršeno precizno opisati stvarnost. Među svim modelima koji se promatraju, tražimo one koji najbolje aproksimiraju stvarnost, tj. pokušavamo minimizirati gubitak informacija. Kullback i Leibler (1951.) razvili su kriterij koji predstavlja gubitak informacija prilikom aproksimiranja stvarnosti. Nazivamo ga Kullback-Leibler kriterij (engl. *Kullback-Leibler information* – KLI). Desetak godina kasnije, Akaike (1973.) predložio je da se KLI koristi kod izbora modela. Akaikov informacijski kriterij (engl. *Akaike Information Criterion* – AIC) radi na principu kažnjavanja modela s većim brojem parametara u modelu. Poznato je da AIC ima sklonost voditi prekomjernom broju parametara (engl. *overfitting*) stoga je 1978. Gideon E. Schwarz dao izmijenjenu inačicu Akaikeovog informacijskog kriterija u vidu Bayesovog informacijskog kriterija strože kažnjava dodatne parametre.

Sam za sebe, broj AIC ili BIC parametra koji se dobije za pojedini model, ne znači ništa dok ga ne usporedi s vrijednostima koje se dobiju za druge modele. Pomoću AIC-a i BIC-a može se odrediti model s najmanjim gubitkom informacija među unaprijed odabranim modelima.

Kako u niže navedenim primjerima nema referentnih modela niti će biti više sličnih modela za iste ulazne podatke, numerička analiza biti će bazirana na ranije navedenim testovima.

5.1.5. Prihvaćanje ili odbacivanje teorijske raspodjele

Zaključno, nakon grafičke i numeričke analize određuje se je li valjana prilagodba određene teorijske raspodjele empirijski podacima odnosno, pripadaju li empirijski podaci pretpostavljenoj teorijskoj raspodjeli.

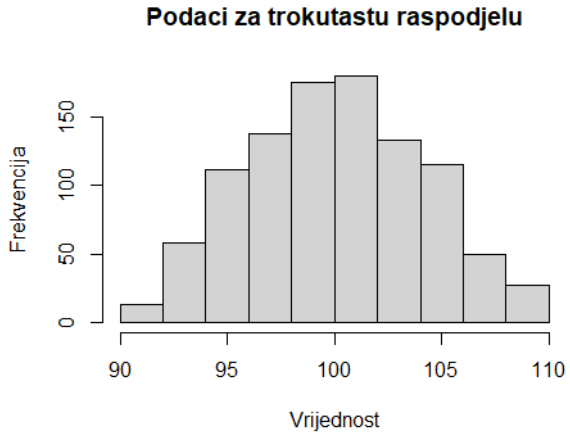
5.2. Prilagodba trokutaste teorijske raspodjele empirijskim podacima

U nastavku će biti prikazana prilagodba trokutaste raspodjele na set empirijskih podataka te će prema ranije navedenim smjernicama biti ispitano opisuje li odabrana teorijska raspodjela empirijske podatke.

Prva faza je učitavanje podataka. U programskom kodu vidljivo je da su učitani podaci .xlsx odnosno Excel formata te nakon toga spremljeni u R format kao tablica. Uz to, prethodno je napravljena aktivacija programskih paketa koji će biti korišteni u nastavku primjera.

```
library(fitdistrplus)

library(ggplot2)
library(mc2d)

library(readxl)
Podaci <- read_excel("C:/Users/Marko Jukic/Documents/MARKO/Podaci.xlsx",
    col_types = c("numeric", "numeric"))

hist(Podaci$trokutasta,
     main = "Podaci za trokutastu raspodjelu",
     xlab = "Vrijednost",
     ylab = "Frekvencija")

trok_mle <- fitdist(Podaci$trokutasta,
                    "triang",
                    method="mle",
                    start = list(min=90, mode=100, max=110))
```

Funkcijom `fitdist` napravljena je prilagodba trokutaste teorijske raspodjele empirijskim podacima prema metodi najveće vjerojatnosti. Za trokutastu raspodjelu navedena programska funkcija zahtijeva dodatne parametre. Potrebno je pretpostaviti vrijednost moda te minimalnu i maksimalnu vrijednost. Kako bi se te vrijednosti procijenile napravljen je histogram [Slika 22.] podataka i prema njemu izabrane vrijednosti koje su navedene iznad u programskom kodu.



Slika 22. Histogram empirijskih podataka za trokutastu raspodjelu

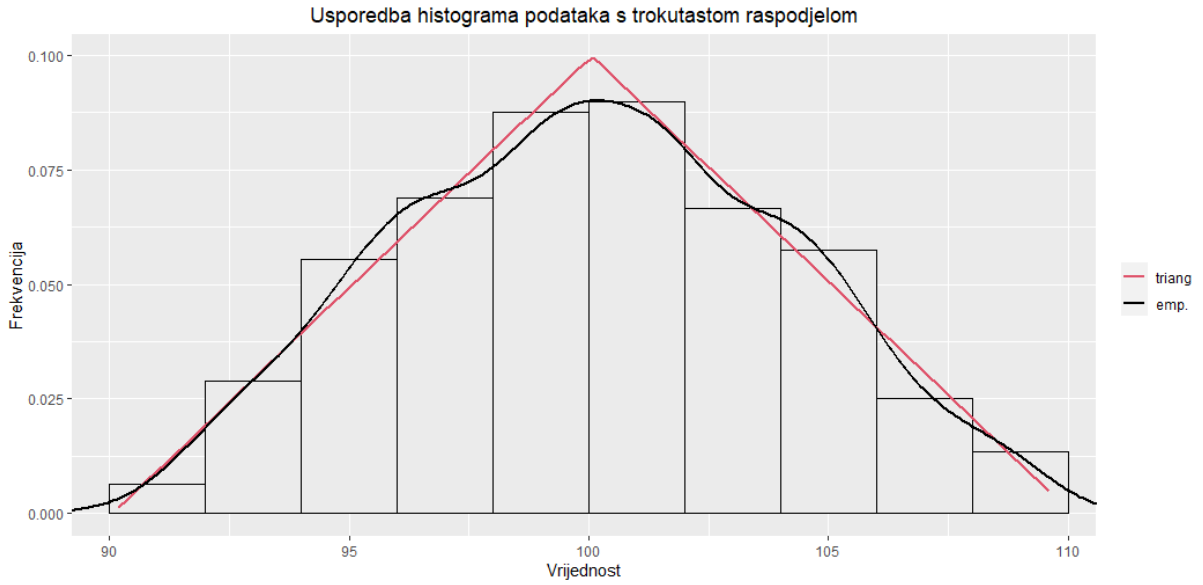
Kao izlazna vrijednost naredbe `print` dobiju se parametri trokutaste raspodjele za empirijske podatke koji su prikazani u nastavku.

```
print(trok_mle)

## Fitting of the distribution ' triang ' by maximum likelihood
## Parameters:
##       estimate Std. Error
## min   90.06415  0.1148962
## mode 100.05985  0.1111872
## max  110.09006  0.1762956
```

5.2.1. Grafička analiza

U nastavku će biti napravljena grafička analiza. Usporedba teorijske funkcije gustoće vjerojatnosti sa histogramom i empirijskom funkcijom gustoće vjerojatnosti biti će pokrenuta naredbom `denscomp`.

```
denscomp(trok_mle, demp = TRUE,
         plotstyle = "ggplot",
         xlab = "Vrijednost",
         ylab = "Frekvencija",
         main = "Usporedba histograma podataka s trokutastom raspodjelom")
```

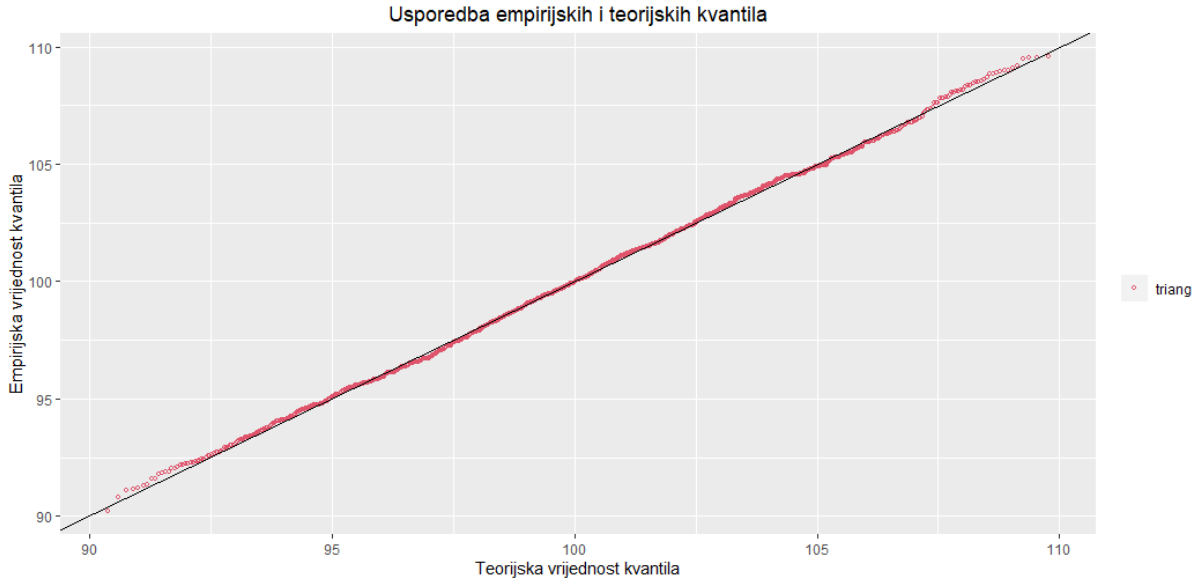


Slika 23. Usporedba histograma podataka s trokutastom raspodjelom

Crvena linija označava teorijsku funkciju gustoće vjerojatnosti trokutaste raspodjele dok crna označava empirijsku funkciju gustoće [Slika 23.]. Vidljivo je da empirijska linija prati teorijsku te da se na nekim dijelovima u potpunosti poklapaju.

Usporedbu teorijskih i empirijskih kvantila pokreće naredba `qqcomp`.

```
qqcomp(trok_mle,
       plotstyle = "ggplot",
       xlab = "Teorijska vrijednost kvantila",
       ylab = "Empirijska vrijednost kvantila",
       main = "Usporedba empirijskih i teorijskih kvantila")
```

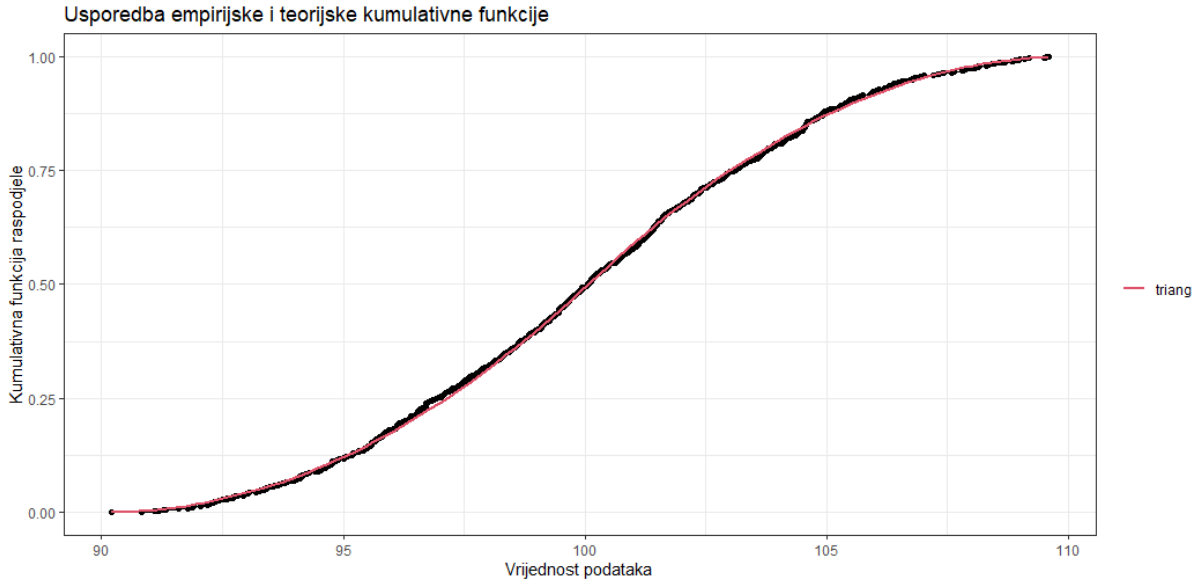


Slika 24. Usporedba empirijskih i teorijskih kvantila

Udaljenost crvenih točkica od crnog pravca označava razliku između vrijednosti kvantila empirijskog i teorijskog skupa [Slika 24]. Vidljivo je da točke prate pravac gotovo cijelom dužinom, manja odstupanja su vidljiva tek na repovima pravca.

Usporedbu empirijske i teorijske kumulativne funkcije pokreće naredba `cdfcomp`.

```
cdfcomp(trok_mle,
        plotstyle = "ggplot",
        xlab = "Vrijednost podataka",
        ylab = "Kumulativna funkcija raspodjele",
        main = "Usporedba empirijske i teorijske kumulativne raspodjele
funkcije")
```



Slika 25. Usporedba empirijske i teorijske kumulativne funkcije

Udaljenost crvenih točkica od crnog pravca označava razliku empirijske i teorijske razdiobe za određenu vrijednost podatka [Slika 25]. Vidljivo je da nema većeg odstupanja točki od pravca cijelom njegovom dužinom.

Zaključno se može reći da je grafičkom analizom utvrđeno da empirijski podaci pripadaju pretpostavljenoj teorijskoj raspodjeli tj. trokutastoj raspodjeli.

5.2.2. Numerička analiza

Numeričko testiranje adekvatnosti odabrane raspodjele provodi se pomoću naredbe `gofstat`.

```
gofstat(trok_mle)

## Goodness-of-fit statistics
##                                1-mle-triang
## Kolmogorov-Smirnov statistic   0.01898692
## Cramer-von Mises statistic     0.03724142
## Anderson-Darling statistic     0.29413533
##
```

```
## Goodness-of-fit criteria
##                                1-mle-triang
## Akaike's Information Criterion     5614.366
## Bayesian Information Criterion     5629.089
```

Granične vrijednosti za ranije opisane testove očitane su ili po potrebi računate prema intervalu vjerodostojnosti:

$$1 - \alpha = 1 - 0{,}05 = 0{,}95 = 95\ \% \quad \textbf{(2)}$$

Gdje je:

α vjerojatnost pogreške

1) Kolmogorov-Smirnov test

H_0: Podaci prate trokutastu raspodjelu podataka.

H_1: Podaci ne prate trokutastu raspodjelu podataka.

Ako je izračunata vrijednost veća od granične, H_0 se odbija uz vjerojatnost pogreške od $\alpha = 5\ \%$.

Granična vrijednost prema tablici [Slika 19.] iznosi:

$$Granična\ vrijednost = \frac{1{,}3581}{\sqrt{1000}} = 0{,}042947 \quad \textbf{(3)}$$

$$Izračunata\ vrijednost = 0{,}0189692 \quad \textbf{(4)}$$

$$0{,}042947 > 0{,}01898692 \quad \textbf{(5)}$$

Zaključak: Ne odbaciti H_0.

2) Cramer-von Mises

H_0: Podaci prate trokutastu raspodjelu podataka.

H_1: Podaci ne prate trokutastu raspodjelu podataka.

Ako je izračunata vrijednost veća od granične, H_0 se odbija uz vjerojatnost pogreške od $\alpha = 5\ \%$.

Granična vrijednost prema tablici [Slika 20.] iznosi:

$$Granična\ vrijednost = 0{,}22 \quad \textbf{(6)}$$

$$Izračunata\ vrijednost = 0{,}03724142 \quad \textbf{(7)}$$

$$0{,}22 > 0{,}03724142 \quad \textbf{(8)}$$

Zaključak: Ne odbaciti H_0.

3) Anderson-Darling test

Anderson–Darling test nije pouzdan za trokutastu raspodjelu, te se u ovom slučaju neće uzeti u obzir.

Kao što je u uvodnom dijelu ovog poglavlja objašnjeno Akaikeov i Bayesov informacijski kriterij sam za sebe ne znači ništa dok se ne usporedi s nekim drugim modelima. Stoga su AIC i BIC u ovom slučaju su zanemareni jer ne postoji referentni model.

Nakon što je provedena prilagodba trokutaste raspodjele empirijskim podacima te zatim provedena grafička i numerička analiza može se zaključiti kako se set empirijskih podataka ponaša po trokutastoj raspodjeli.

5.3. Prilagodba pravokutne teorijske raspodjele empirijskim podacima

U nastavku je prikazana prilagodba pravokutne raspodjele na set empirijskih podataka te je prema ranije navedenim smjernicama ispitano opisuje li odabrana teorijska raspodjela empirijske podatke.

Kako su se podaci za oba primjera nalazila u istom dokumentu, prva faza učitavanja podataka je već obavljena. Također, jednom kada se aktiviraju programski paketi nakon pokretanja R-a, nije ih potrebno ponovno aktivirati, stoga će i to izostati.

Funkcijom `fitdist` napravljena je prilagodba pravokutne teorijske raspodjele empirijskim podacima prema metodi momenta.

```
prav_mme <- fitdist(Podaci$pravokutna,
                    "unif",
                    method="mme")
```

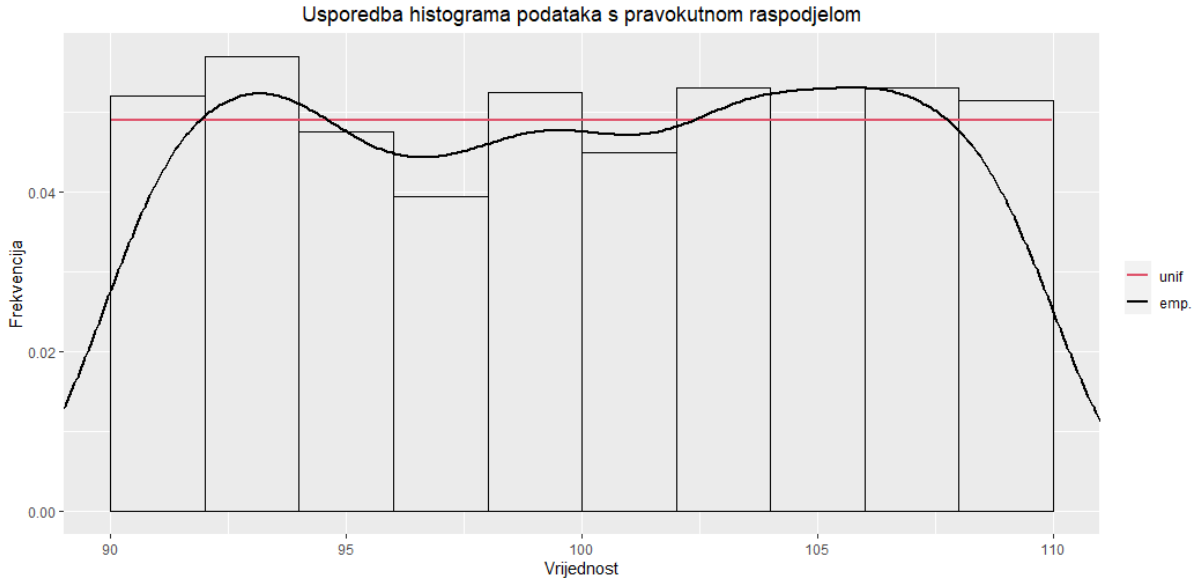
Kao izlazna vrijednost naredbe `print` dobiju se parametri pravokutne raspodjele za empirijske podatke koji su prikazani u nastavku.

```
print(prav_mme)

## Fitting of the distribution ' unif ' by matching moments
## Parameters:
##    estimate
## 1  89.79512
## 2 110.20220
```

5.3.1. Grafička analiza

U nastavku će biti napravljena grafička analiza. Usporedba teorijske funkcije gustoće vjerojatnosti s histogramom i empirijskom funkcijom gustoće vjerojatnosti biti će pokrenuta naredbom `denscomp`.

```
denscomp(prav_mme, demp = TRUE,
         plotstyle = "ggplot",
         xlab = "Vrijednost",
         ylab = "Frekvencija",
         main = "Usporedba histograma podataka s pravokutnom raspodjelom")
```

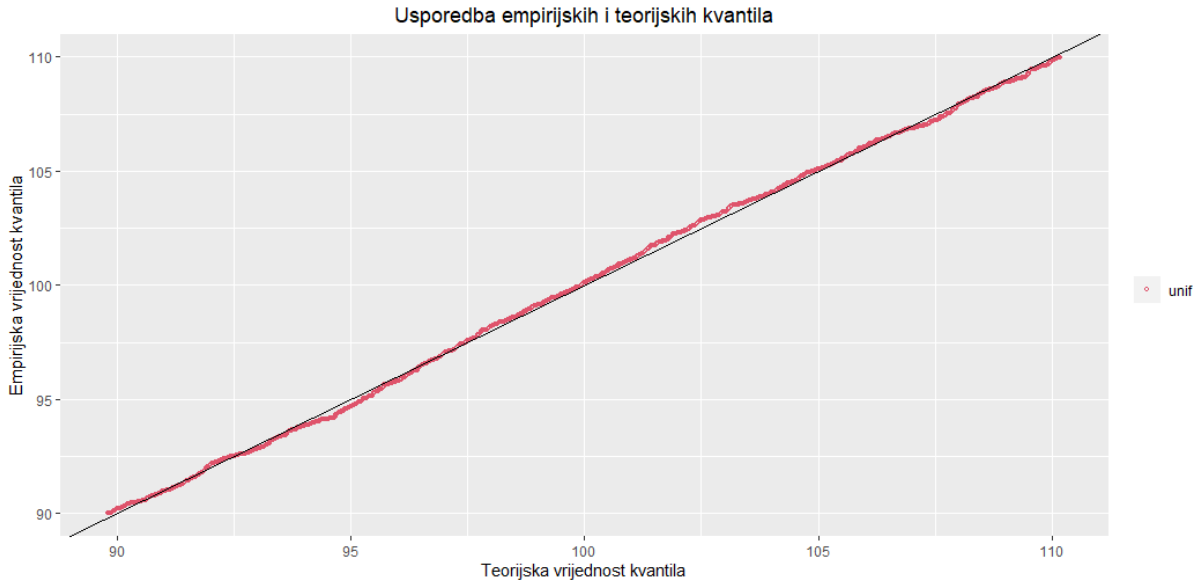


Slika 26. Usporedba histograma podataka s pravokutnom raspodjelom

Crvena linija označava teorijsku funkciju gustoće vjerojatnosti pravokutne raspodjele dok crna označava empirijsku funkciju gustoće [Slika 26.]. Vrhovi stupaca histograma se nalaze blizu linije funkcije gustoće pravokutne raspodjele dok krivulja empirijske funkcije gustoće na repovima dosta pada i na taj način se udaljava od teorijske linije. Grešku na repovima možemo pripisati R programskom jeziku jer je vidljivo da su visine prvog i zadnjeg stupca iznad linije teorijske raspodjele.

Usporedbu teorijskih i empirijskih kvantila pokreće naredba `qqcomp`.

```
qqcomp(prav_mme,
       plotstyle = "ggplot",
       xlab = "Teorijska vrijednost kvantila",
       ylab = "Empirijska vrijednost kvantila",
       main = "Usporedba empirijskih i teorijskih kvantila")
```

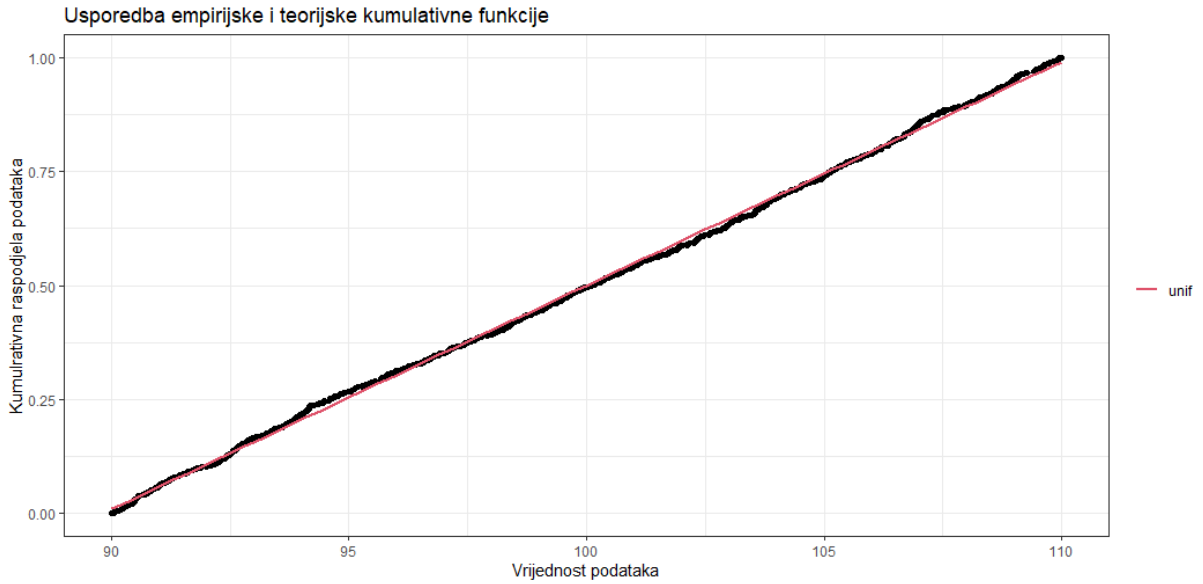


Slika 27. Usporedba empirijskih i teorijskih kvantila

Udaljenost crvenih točkica od crnog pravca označava razliku između vrijednosti kvantila empirijskog i teorijskog skupa [Slika 27.]. Vidljivo je da točke prate pravac gotovo cijelom dužinom, uz manja odstupanja za vrijednosti oko 95 i 102,5.

Usporedbu empirijske i teorijske kumulativne funkcije pokreće naredba `cdfcomp`.

```
cdfcomp(prav_mme,
        plotstyle = "ggplot",
        xlab = "Vrijednost podataka",
        ylab = "Kumulrativna raspodjela podataka",
        main = "Usporedba empirijske i teorijske kumulativne funkcije")
```



Slika 28. Usporedba empirijske i teorijske kumulativne funkcije

Udaljenost crvenih točkica od crnog pravca označava razliku empirijske i teorijske razdiobe za određenu vrijednost podatka [Slika 28.]. Vidljivo je da nema većeg odstupanja točki od pravca cijelom njegovom dužinom.

Zaključno se može reći da je grafičkom analizom utvrđeno da empirijski podaci pripadaju pretpostavljenoj teorijskoj raspodjeli tj. pravokutnoj raspodjeli.

5.3.2. Numerička analiza

Numeričko testiranje adekvatnosti odabrane raspodjele provodi se pomoću naredbe `gofstat`.

```
gofstat(prav_mme)
```

```
## Goodness-of-fit statistics
##                                1-mme-unif
## Kolmogorov-Smirnov statistic 0.02219213
## Cramer-von Mises statistic   0.06921527
## Anderson-Darling statistic   0.74441124
##
## Goodness-of-fit criteria
##                                  1-mme-unif
## Akaike's Information Criterion   6035.764
## Bayesian Information Criterion   6045.579
```

Granične vrijednosti za ranije opisane testove očitane su ili po potrebi računate prema intervalu vjerodostojnosti:

$$1 - \alpha = 1 - 0{,}05 = 0{,}95 = 95\ \% \quad \textbf{(9)}$$

1) Kolmogorov-Smirnov test

H_0: Podaci prate trokutastu raspodjelu podataka.

H_1: Podaci ne prate trokutastu raspodjelu podataka.

Ako je izračunata vrijednost veća od granične, H_0 se odbija uz vjerojatnost pogreške od $\alpha = 5\ \%$.

Granična vrijednost prema tablici [Slika 19.] iznosi:

$$Granična\ vrijednost = \frac{1{,}3581}{\sqrt{1000}} = 0{,}042947 \quad \textbf{(10)}$$

$$Izračunata\ vrijednost = 0{,}02219213 \quad \textbf{(11)}$$

$$0{,}042947 > 0{,}02219213 \quad \textbf{(12)}$$

Zaključak: Ne odbaciti H_0.

2) Cramer-von Mises test

H_0: Podaci prate trokutastu raspodjelu podataka.

H_1: Podaci ne prate trokutastu raspodjelu podataka.

Ako je izračunata vrijednost veća od granične, H_0 se odbija uz vjerojatnost pogreške od $\alpha = 5\ \%$.

Granična vrijednost prema tablici [Slika 20.] iznosi:

$$Granična\ vrijednost = 0{,}22 \quad \textbf{(13)}$$

$$Izračunata\ vrijednost = 0{,}06921527 \quad \textbf{(14)}$$

$$0{,}22 > 0{,}06921527 \quad \textbf{(15)}$$

Zaključak: Ne odbaciti H_0.

3) Anderson-Darling test

H_0: Podaci prate trokutastu raspodjelu podataka.

H_1: Podaci ne prate trokutastu raspodjelu podataka.

Ako je izračunata vrijednost veća od granične, H_0 se odbija uz vjerojatnost pogreške od $\alpha = 5\ \%$.

Granična vrijednost prema tablici [Slika 21.] iznosi:

$$Granična\ vrijednost = 2{,}49 \quad \textbf{(16)}$$

$$Izračunata\ vrijednost = 0{,}74441124 \quad \textbf{(17)}$$

$$2{,}49 > 0{,}74441124 \quad \textbf{(18)}$$

Zaključak: Ne odbaciti H_0.

Kao što je u uvodnom dijelu ovog poglavlja objašnjeno Akaikeov i Bayesov informacijski kriterij sam za sebe ne znači ništa dok se ne usporedi s nekim drugim modelima. Stoga su AIC i BIC u ovom slučaju su zanemareni jer ne postoji referentni model.

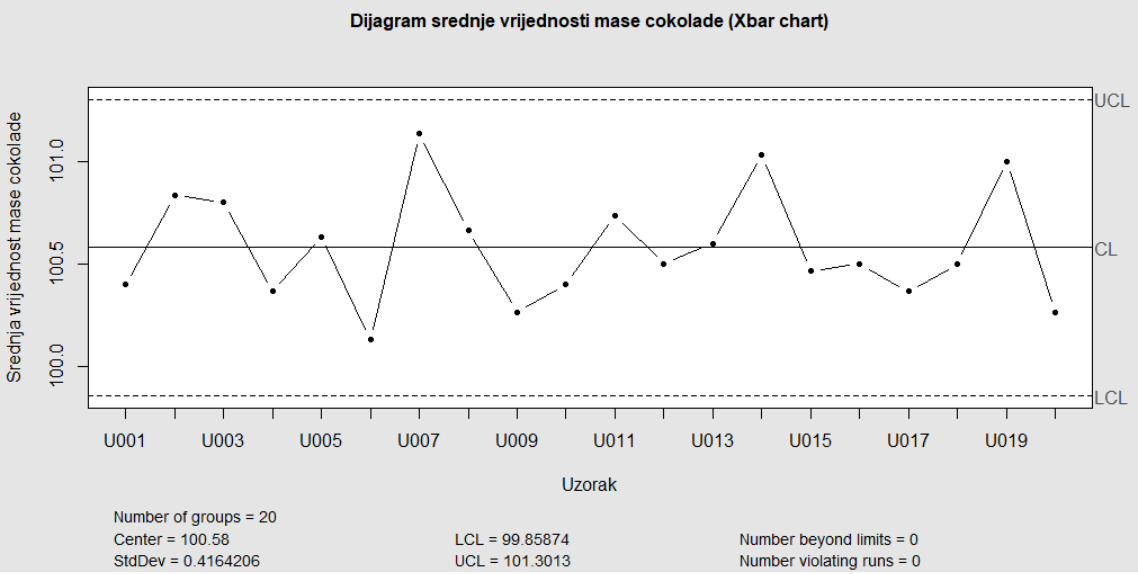
Nakon što je provedena prilagodba pravokutne raspodjele empirijskim podacima te zatim provedena grafička i numerička analiza može se zaključiti kako se set empirijskih podataka ponaša po pravokutnoj raspodjeli.

6. SPOSOBNOST PROCESA

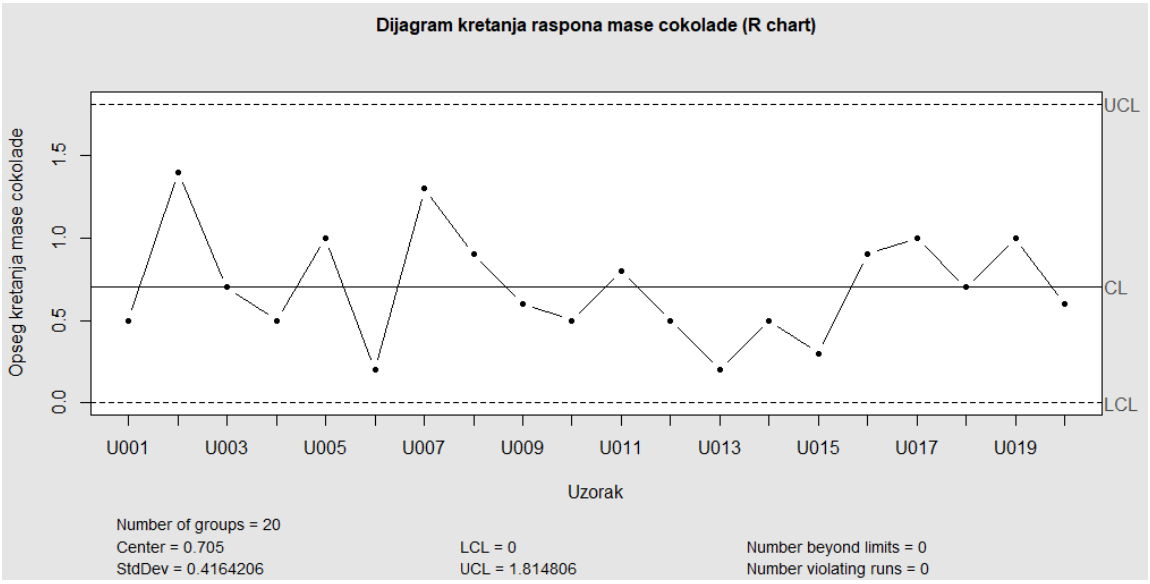
U proizvodnom procesu proizvodnje čokolade provodi se kontrola kvalitete krajnjeg proizvoda. Komponenta kvalitete koja se mjeri je masa čokolade koja ima donju granicu specifikacije od 100 g. Podlogu za analizu čine mjerenja 20 slučajnih uzoraka, veličina uzorka iznosi $n_{uzorka} = 3$. Zadatak je pratiti proces $\bar{x}$-R kontrolnom kartom te procijeniti sposobnost procesa. Podatke je bilo potrebno s kontrolne liste unijeti u R programski jezik.

```
uzorak_001 <- c(100.3, 100.2, 100.7)
uzorak_002 <- c(101.7, 100.5, 100.3)
uzorak_003 <- c(100.9, 101.1, 100.4)
uzorak_004 <- c(100.4, 100.6, 100.1)
uzorak_005 <- c(100.9, 101.0, 100.0)
uzorak_006 <- c(100.2, 100.2, 100.0)
uzorak_007 <- c(101.9, 100.6, 100.9)
uzorak_008 <- c(100.5, 100.3, 101.2)
uzorak_009 <- c(100.2, 100.0, 100.6)
uzorak_010 <- c(100.6, 100.1, 100.5)
uzorak_011 <- c(101.0, 101.0, 100.2)
uzorak_012 <- c(100.4, 100.8, 100.3)
uzorak_013 <- c(100.6, 100.5, 100.7)
uzorak_014 <- c(100.8, 101.3, 101.0)
uzorak_015 <- c(100.6, 100.3, 100.5)
uzorak_016 <- c(100.6, 100.9, 100.0)
uzorak_017 <- c(101.0, 100.0, 100.1)
uzorak_018 <- c(100.6, 100.1, 100.8)
uzorak_019 <- c(100.5, 101.0, 101.5)
uzorak_020 <- c(100.0, 100.6, 100.2)

cokolada <- data.frame(masa = c(uzorak_001, uzorak_002, uzorak_003,
uzorak_004, uzorak_005, uzorak_006, uzorak_007,
                               uzorak_008, uzorak_009, uzorak_010,
uzorak_011, uzorak_012, uzorak_013, uzorak_014,
                               uzorak_015, uzorak_016, uzorak_017,
uzorak_018, uzorak_019, uzorak_020),
                       uzorak =
rep(c("U001","U002","U003","U004","U005","U006","U007",
"U008","U009","U010","U011","U012","U013","U014",
"U015","U016","U017","U018","U019","U020"), each =3 ))
library(qcc)

masa_uzorak <- qcc.groups(
  data = cokolada$masa,
  sample = cokolada$uzorak)
X_graf <- qcc(data = masa_uzorak,
              type = "xbar",
              xlab = "Uzorak",
              ylab = "Srednja vrijednost mase cokolade",
```

```
              title = "Dijagram srednje vrijednosti mase cokolade (Xbar
chart)")

R_graf <- qcc(data = masa_uzorak,
              type = "R",
              xlab = "Uzorak",
              ylab = "Opseg kretanja mase cokolade",
              title = "Dijagram kretanja raspona mase cokolade (R chart)")
```



Slika 29. $\bar{X}$ kontrolna karta

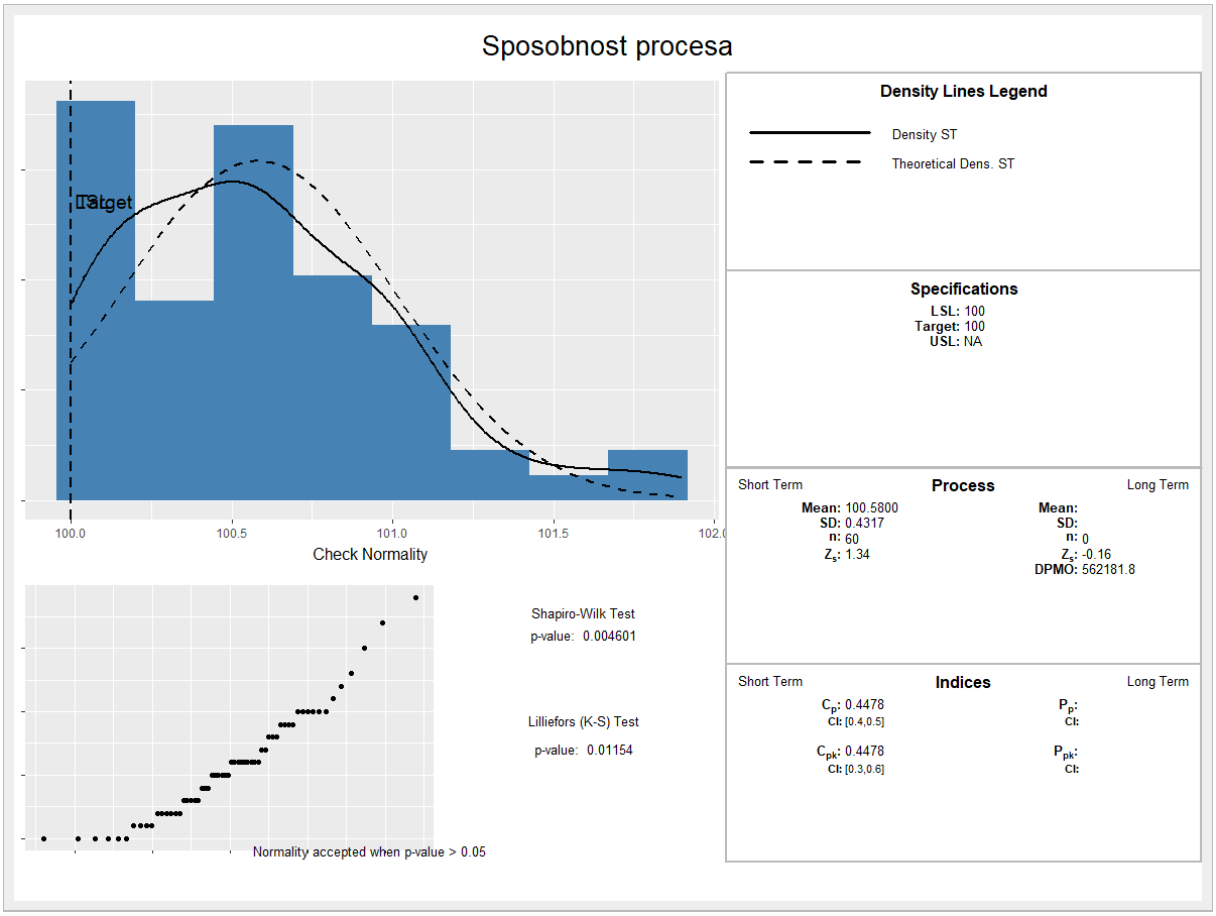


Slika 30. *R* kontrolna karta

Nakon unosa podataka i formiranja tablice aktiviran je programski paket `qcc` te su izrađene kontrolne karte `xbar` i `R` tipa kao što je navedeno u zadatku. Iz $\bar{x}$-*R* [Slika 29., Slika 30.] kontrolne karte vidljivo da ni jedan uzorak ne izlazi izvan kontrolnih granica što znači da ne postoje značajne varijacije u procesu te se za proces može reći da je pod kontrolom.

```
library(SixSigma)

ss.study.ca(cokolada$masa,
            LSL = 100,
            Target = 100,
            alpha = 0.05,
            f.main = "Sposobnost procesa")
```

```
## Warning in ss.study.ca(cokolada$masa, LSL = 100, Target = 100, alpha =
0.05, :
## Normality test/s failed
```



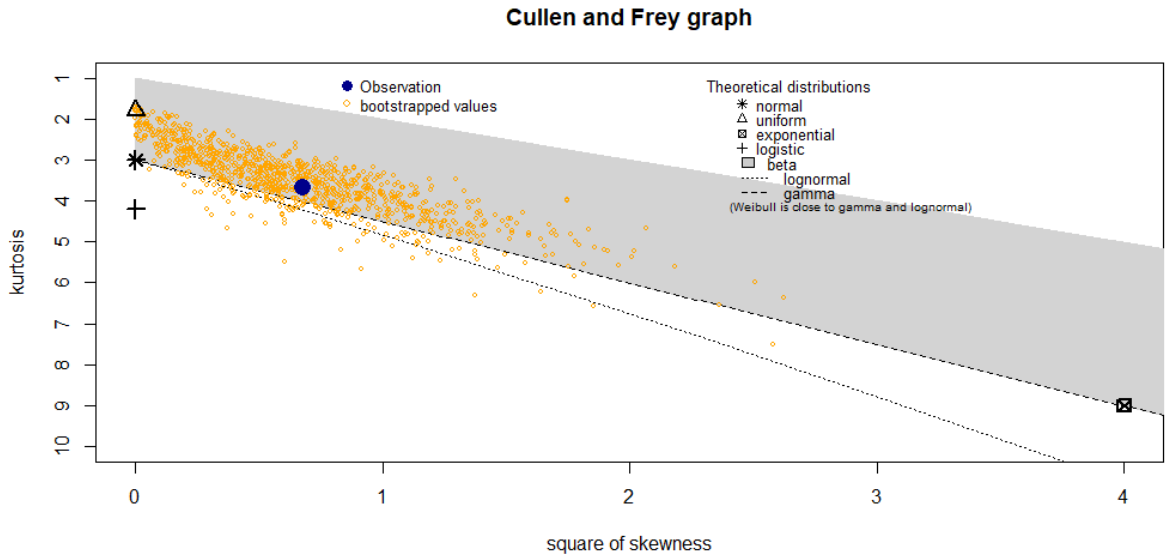
Slika 31. Sposobnost procesa

Za izračun sposobnosti procesa korišten je programski paket SixSigma. Osim uvjeta da je proces stabilan, odnosno „pod kontrolom“ drugi uvjet za pravilnu interpretaciju sposobnosti procesa je da se promatrani podaci mogu opisati normalnom raspodjelom. U izvještaju iznad slike je vidljivo da program upozorava korisnika porukom kako njegov skup podataka nije prošao test normalnosti. Prema histogramu [Slika 31.] koji je prikazan u izvještaju vidljivo je da se radi o pozitivnoj asimetriji koja je česta u mjerenjima koja imaju određenu donju granicu specifikacije. Za situacije kod nenormalne raspodjele podataka moguće je pokušati odrediti prate li podaci neku drugu poznatu raspodjelu podataka ili pomoću transformacije

podataka generirati model transformiranih vrijednosti i granica tako da odgovaraju normalnoj raspodjeli.

Cullen-Frey graf (engl. *Cullen and Frey graph*) osim podataka deskriptivne statistike koje je moguće vidjeti ispod u izvještaju, procjenjuje i kojom bi se teorijskom raspodjelom mogli opisati podaci. Procjenu provodi preko kvadrata koeficijenta asimetrije (engl. *skewness*) koji označava apscisu grafa te koeficijenta spljoštenosti (engl. *kurtosis*) koji označava ordinatu grafa. Koeficijent asimetrije predstavlja mjeru nagnutosti raspodjele podataka na lijevu ili desnu stranu dok koeficijent spljoštenosti predstavlja mjeru zaobljenosti (spljoštenosti) raspodjele.

```
library(fitdistrplus)

descdist(cokolada$masa, boot = 1000)

## summary statistics
## ------
## min:  100   max:  101.9
## median:  100.55
## mean:  100.58
## estimated sd:  0.4317093
## estimated skewness:  0.8224825
## estimated kurtosis:  3.677828
```

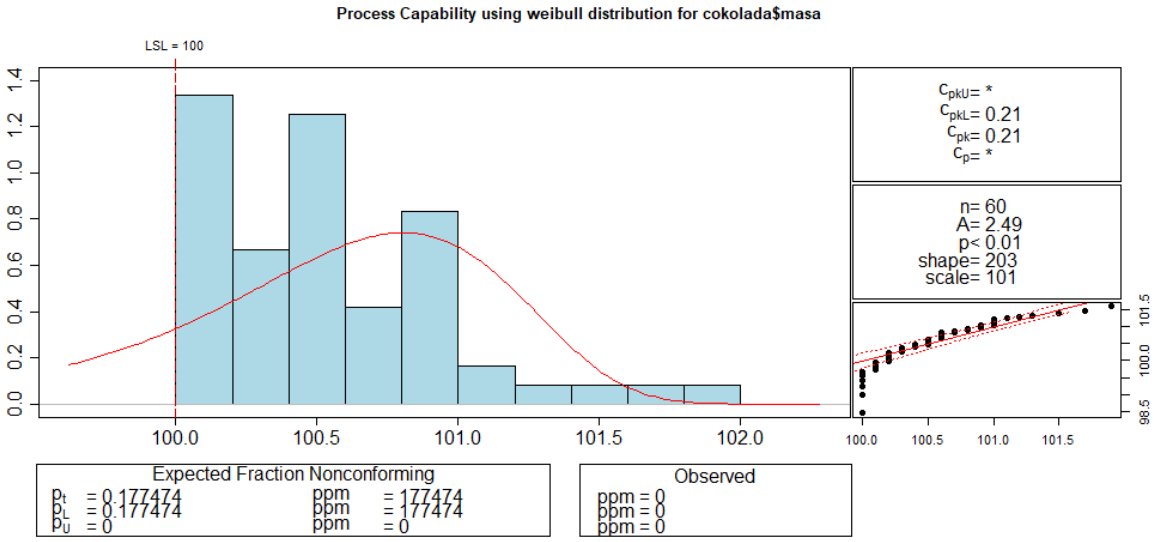


Slika 32. Cullen-Frey graf

Prema grafu [Slika 32.] je vidljivo da su promatrani podaci najbliže log-normalnoj, Weibullovoj i gamma raspodjeli, stoga će u nastavku biti procijenjena sposobnost podataka prema tim raspodjelama.

```
library(qualityTools)
```

```
cp(cokolada$masa,"weibull", lsl = 100)

##  Anderson Darling Test for weibull distribution
##
## data:  cokolada$masa
## A = 2.4942, shape = 203.043, scale = 100.807, p-value <= 0.01
## alternative hypothesis: true distribution is not equal to weibull
```



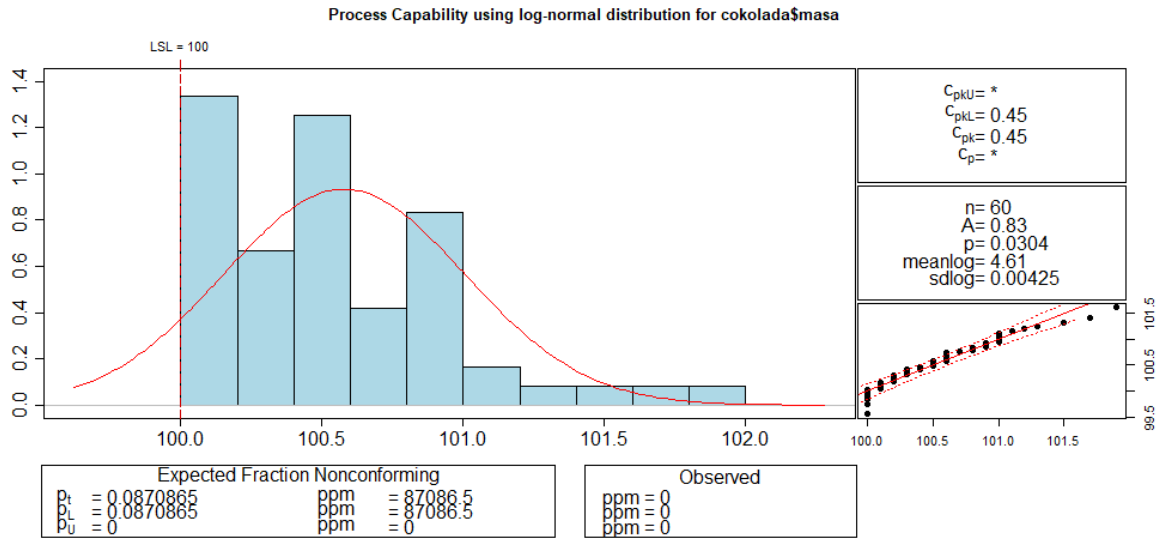
Slika 33. Sposobnost procesa primjenom Weibullove raspodjele

Kako bi se moglo zaključiti da raspodjela empirijskih podataka prati neku od teorijskih raspodjela, *p* vrijednost Anderson-Darling testa mora biti veća od razine značajnosti α.

$$p > \alpha = 0{,}05$$

Iz histograma [Slika 33.] i dobivenih podataka vidljivo je da je *p* vrijednost Anderson Darling testa manja od razine značajnosti što označava da podaci ne prate Weibullovu raspodjelu.

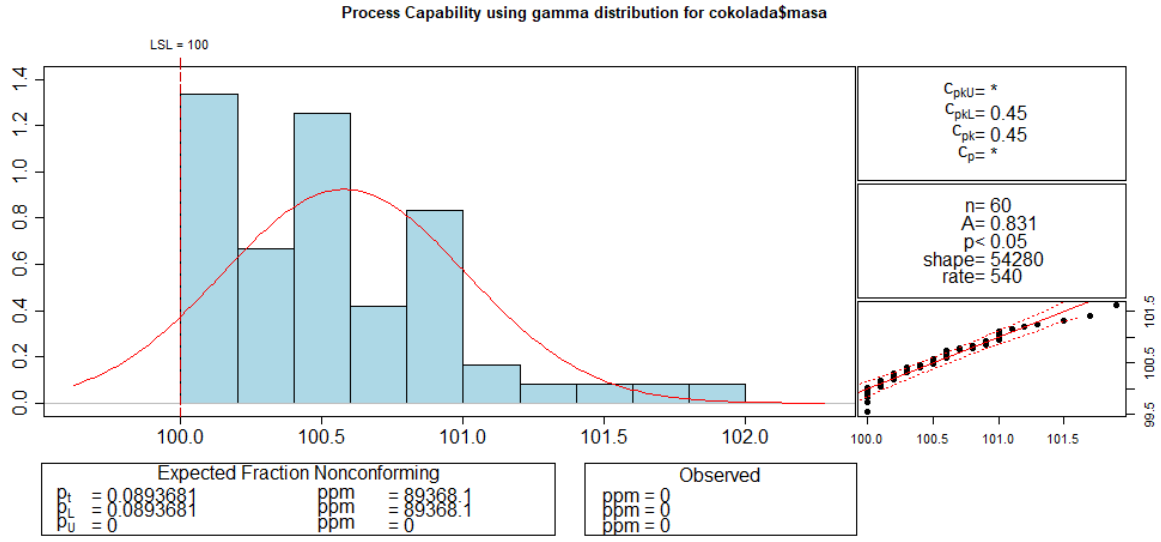
```
cp(cokolada$masa,"log-normal", lsl = 100)

##  Anderson Darling Test for log-normal distribution
##
## data:  cokolada$masa
## A = 0.8301, meanlog = 4.611, sdlog = 0.004, p-value = 0.03037
## alternative hypothesis: true distribution is not equal to log-normal
```



Slika 34. Sposobnost procesa primjenom log-normalne raspodjele

Iz histograma [Slika 34.] i dobivenih podataka vidljivo je da je *p* vrijednost Anderson Darling testa manja od razine značajnosti što označava da podaci ne prate lognormalnu raspodjelu.

```
cp(cokolada$masa, "gamma", lsl = 100)

##  Anderson Darling Test for gamma distribution
##
## data:  cokolada$masa
## A = 0.8305, shape = 54280.088, rate = 539.671, p-value <= 0.05
## alternative hypothesis: true distribution is not equal to gamma
```



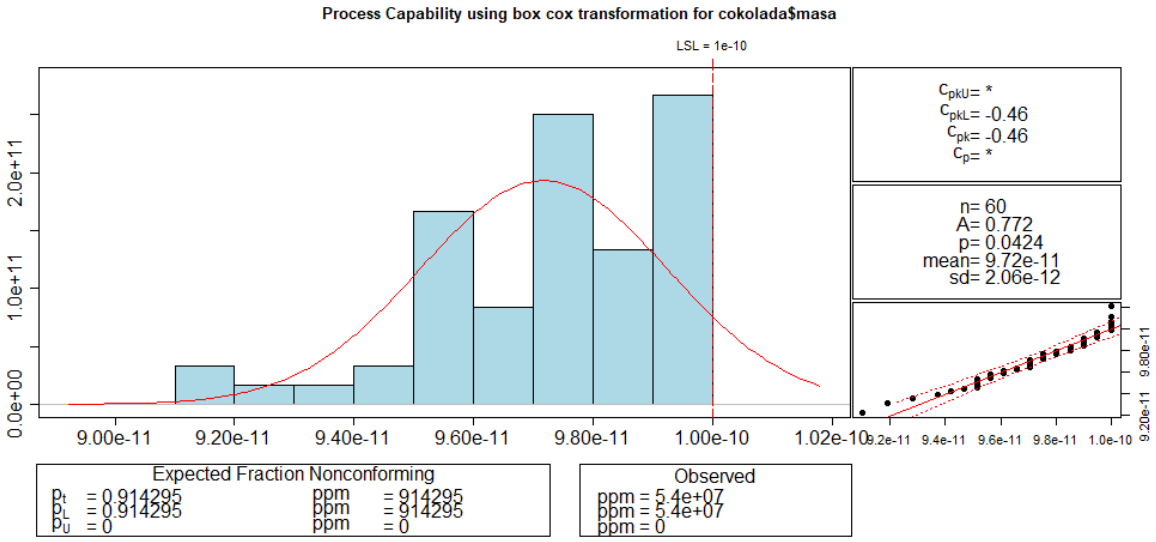
Slika 35. Sposobnost procesa primjenom gamma raspodjele

Iz histograma [Slika 35.] i dobivenih podataka vidljivo je da je *p* vrijednost Anderson Darling testa manja od razine značajnosti što označava da podaci ne prate gamma raspodjelu.

Zaključno se može reći da se podaci iz zadatka ne mogu opisati ni jednom od tri predložene teorijske raspodjele podataka. Ako se pogledaju sve izračunate demonstrirane izvrsnosti procesa C_{pk}, vidljivo je da gama, log-normalna i normalna raspodjela imaju približno jednake vrijednosti, a Weibullova raspodjela manju od njih.

Kao što je već rečeno, ako podaci ne prate normalnu ni jednu drugu teorijsku raspodjelu, moguće je pomoću transformacije podataka generirati model transformiranih vrijednosti i granica tako da odgovaraju normalnoj raspodjeli, što će biti učinjeno u nastavku pomoću Box-Cox transformacije.

```
cp(cokolada$masa, boxcox = TRUE, lsl = 100)
```



Slika 36. Sposobnost procesa primjenom Box-Cox transformacije

```
##  Anderson Darling Test for normal distribution
##
## data:  cokolada$masa
## A = 0.772, mean = 0, sd = 0, p-value = 0.04243
## alternative hypothesis: true distribution is not equal to normal
```

Vidljivo je da Box-Cox transformacija nije dobro provedena te da ni pomoću nje nije moguće odrediti sposobnost procesa [Slika 36.]. Na osnovu svega dosad navedenog može se zaključiti da proces nije sposoban te da je potrebno napraviti analizu proizvoda, proizvodnog procesa i načina kontrole konačnog proizvoda kako bi se utvrdili problemi i ispravile greške u sustavu.

7. UTJECAJ OCJENE SUKLADNOSTI PROIZVODA NA PROCJENU RIZIKA PROIZVOĐAČA I POTROŠAČA

Rizik kao efekt neizvjesnosti u postizanju ciljeva postaje glavni čimbenik u planiranju i organizaciji poslovanja, a time i u upravljanju kvalitetom. Ocjena sukladnosti se definira kao svaka aktivnost koja je poduzeta kako bi se izravno ili neizravno utvrdilo zadovoljava li proizvod, proces, sustav, osoba ili tijelo relevantne standarde i ispunjava li određene zahtjeve. Prilikom donošenja odluka o ocjeni sukladnosti mora se procijeniti rizik od pogrešno odbijenog proizvoda, kao i rizik da potrošač koristi prihvaćeni nesukladan proizvod. Donošenje pouzdane odluke u ocjenjivanju sukladnosti moguće je samo nakon izračuna vjerojatnosti sukladnosti, rizika potrošača (ako je artikl prihvaćen), rizika proizvođača (ako je artikl odbijen) i intervala prihvaćanja [25].

Kako bi se osiguralo da rizici proizvođača i potrošača ne prelaze dopuštene vrijednosti, treba uzeti u obzir specifikacijske granice (interval tolerancije), interval prihvaćanja i standardnu nesigurnost rezultata mjerenja. Ovisno o kvaliteti rezultata mjerenja i pouzdanosti procjene rizika mogu se donijeti različite odluke, npr. odluka o prihvaćanju proizvoda, ponovnom mjerenju proizvoda preciznijim mjernim sustavom, rekonstruiranju proizvoda, odbijanju proizvoda itd. [26].

Na primjeru proizvodnje aparata za zavarivanje biti će izračunat rizik proizvođača i potrošača. U primjeru je prema normi JCGM 106:2012 (Procjena mjerenih podataka – Uloga mjerne nesigurnosti u ocjeni sukladnosti) korišten postupak binarnog pravila za izračunavanje ocjene sukladnosti pomoću raspodjele vjerojatnosti. Također, utvrđen je i utjecaj mjerne nesigurnosti na rizik proizvođača i potrošača.

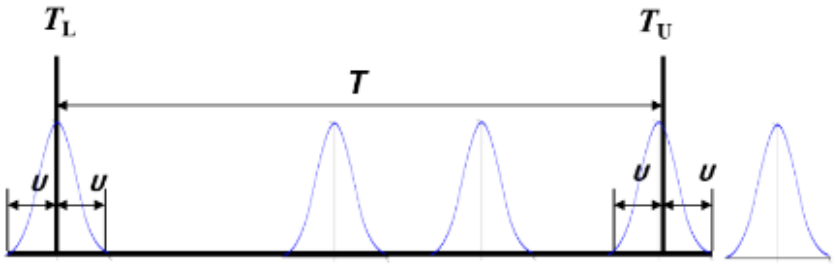
7.1. Pravila za donošenje odluka odluke

Zbog mjerne nesigurnosti uvijek postoji rizik pogrešnog donošenja odluke zadovoljava li predmet ili ne zadane zahtjeve bazirano na izmjerenim veličinama predmeta. Postoje dvije vrste pogrešaka: predmet prihvaćen kao sukladan, ali je u stvari nesukladan i predmet odbijen kao nesukladan, a zapravo je sukladan [25].

7.1.1. Pravilo jednostavnog prihvaćanja

Kada se za pravilo odluke koristi pravilo jednostavnog prihvaćanja [Slika 37.], proizvođač i potrošač prihvaćaju rezultat mjerenja unutar tolerancijskog intervala i dijele posljedice

netočnih odluka. U tom slučaju, mjerna nesigurnost mora biti prikladna i prihvatljiva za obje strane. Pravilo jednostavnog prihvaćanja i normalna raspodjela rezultata mjerenja može dovesti do vjerojatnosti prihvaćanja nesukladnog ili odbijanja sukladnog predmeta u iznosu od 50 % [25].



Slika 37. Pravilo jednostavnog prihvaćanja s gornjom i donjom granicom tolerancije [27]

Gdje je:

T_L – donja granica specifikacija

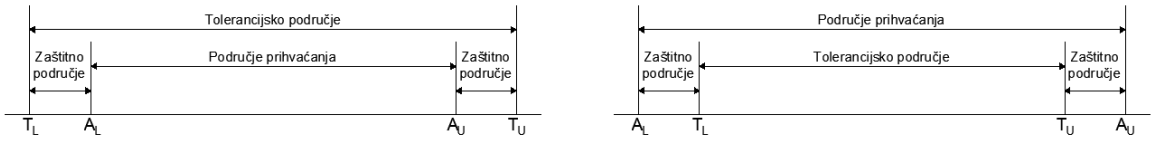
T_U – gornja granica specifikacija

T – tolerancijsko područje

U – mjerna nesigurnost

7.1.2. Binarno pravilo

Kada se ocjenjivanje sukladnosti utvrđuje pomoću binarnog pravila odlučivanja, mjeri se svojstvo predmeta i prihvaća se kao sukladno ako se izmjerena vrijednost svojstva nalazi unutar definiranog intervala prihvaćanja [Slika 38]. Na lijevom dijelu slike binarno pravilo je korišteno kako bi se smanjio rizik potrošača, dok je na desnom dijelu prikaz područja prihvaćanja kojim se smanjuje rizik proizvođača.



Slika 38. Binarno pravilo ocjene sukladnosti [25]

Gdje je:

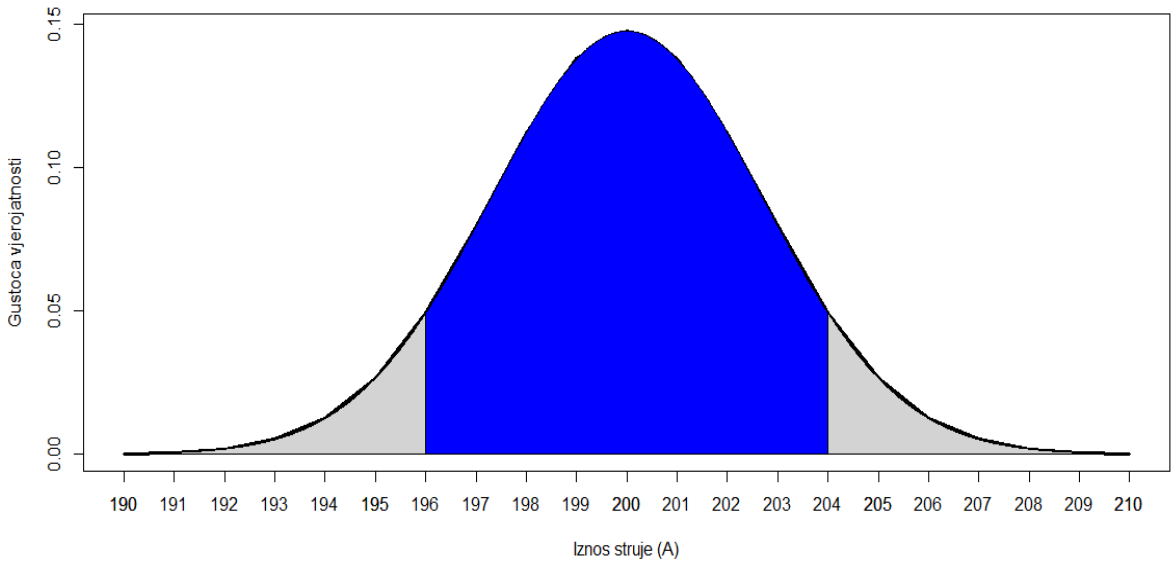
A_L donja granica prihvaćanja

A_U gornja granica prihvaćanja

Uvođenjem intervala prihvaćanja, rizik pogrešnog donošenja odluke koji je izazvan utjecajem mjerne nesigurnosti može se balansirati i na taj način minimizirati troškove povezane s mogućim pogrešnim odlukama [26].

7.2. Primjer

Proizvođač sustava i opreme za zavarivanje proizvodi aparate za zavarivanje. Aparat za zavarivanje koristi konstantan iznos struje, ali tijekom postupka zavarivanja struja oscilira u iznosu od nekoliko ampera. Oscilacije nisu dobra stvar, a ako pređu određenu granicu stroj će prestati pravilno raditi te zavar neće biti pouzdan. Nazivna struja za ovu vrstu postupka je 200 A. Stroj radi pravilno ako je iznos struje unutar tolerancijskog intervala definiranog granicama tolerancije s donjom granicom specifikacije $\mathcal{T}_L$ = 196 A i gornjom granicom specifikacije $\mathcal{T}_U$ = 204 A. Nakon provođenja mjerenja u svrhu kontrole, histogram [Slika 39.] dobivenih vrijednosti izgleda kao normalna krivulja raspodjele sa srednjom vrijednošću y_0 = 200 A te standardnim odstupanjem $\sigma = u_0$ = 2,7 A [27].



Slika 39. Vjerojatnost sukladnosti

U nastavku je prikazan programski kod u kojem su u prvom dijelu napisani ulazni podaci koji su dani u tekstu primjera, te nakon toga izračunate vjerojatnosti za sukladne i nesukladne proizvode.

```
mean = 200
sigma = 2.7
TL = 196
TU = 204
```

```
AL =  196.6
AU = 203.4
um = 1.2

pnorm(TL, mean, sigma) + pnorm(TU, mean, sigma, lower.tail = FALSE)

## [1] 0.1384783

1-(pnorm(TL, mean, sigma) + pnorm(TU, mean, sigma, lower.tail = FALSE))

## [1] 0.8615217
```

$$p_s = 0{,}8615 \quad \textbf{(19)}$$

$$\overline{p_s} = 0{,}1384 \quad \textbf{(20)}$$

Gdje je:

p_s vjerojatnost sukladnih strojeva

$\overline{p_s}$ vjerojatnost nesukladnih strojeva

Prema izračunu vjerojatnost nesukladnih strojeva koje proizvodi proizvođač iznosi 13,85 %, dok vjerojatnost sukladnih strojeva iznosi 86,15 %. To znači da približno 14 % isporučenih strojeva neće biti u skladu sa specifikacijama što nije prihvatljiva razina sukladnosti.

Iako kupnja novog i skupljeg stroja za proizvodnju može smanjiti varijabilnost strojeva, proizvođač je odlučio zadržati trenutni stroj za proizvodnju aparata te umjesto toga uvesti postupak kontrole kvalitete s ciljem otkrivanja i otklanjanja nesukladnih aparata. Postupak se sastoji od kontrole aparata specijalnom opremom čija mjerna nesigurnost iznosi u_m = 1,2 A. S ciljem smanjenja vjerojatnosti isporuke aparata za zavarivanje koji ne udovoljavaju potrebnim zahtjevima, granice prihvaćanja (A_L = 196,6 A i A_U = 203,4 A) postavljaju se unutar granica specifikacija na jednakoj udaljenosti dužine sigurnosnog područja s obje strane. Duljina sigurnosnog područja ω izračunava se kao ω = 204 A – 203,4 A = 0,6 A [27].

Rizik proizvođača R_P i rizik potrošača R_C izračunavaju se prema sljedećim formulama:

$$R_P = \int_{\frac{T_U - y_0}{u_0}}^{\frac{T_L - y_0}{u_0}} (1 - F(z))\varphi_0(z)dz \quad \textbf{(21)}$$

$$R_C = \int_{-\infty}^{\frac{T_L - y_0}{u_0}} F(z)\varphi_0(z)dz + \int_{\frac{T_U - y_0}{u_0}}^{\infty} F(z)\varphi_0(z)dz \quad \textbf{(22)}$$

Gdje je:

$$\varphi_0(z) = (1/\sqrt{2\pi})\exp(-z^2/2) \quad \textbf{(23)}$$

$$F(z) = \Phi\left(\frac{A_{\mathrm{U}} - y_0 - u_0 z}{u_{\mathrm{m}}}\right) - \Phi\left(\frac{A_{\mathrm{L}} - y_0 - u_0 z}{u_{\mathrm{m}}}\right) \quad \textbf{(24)}$$

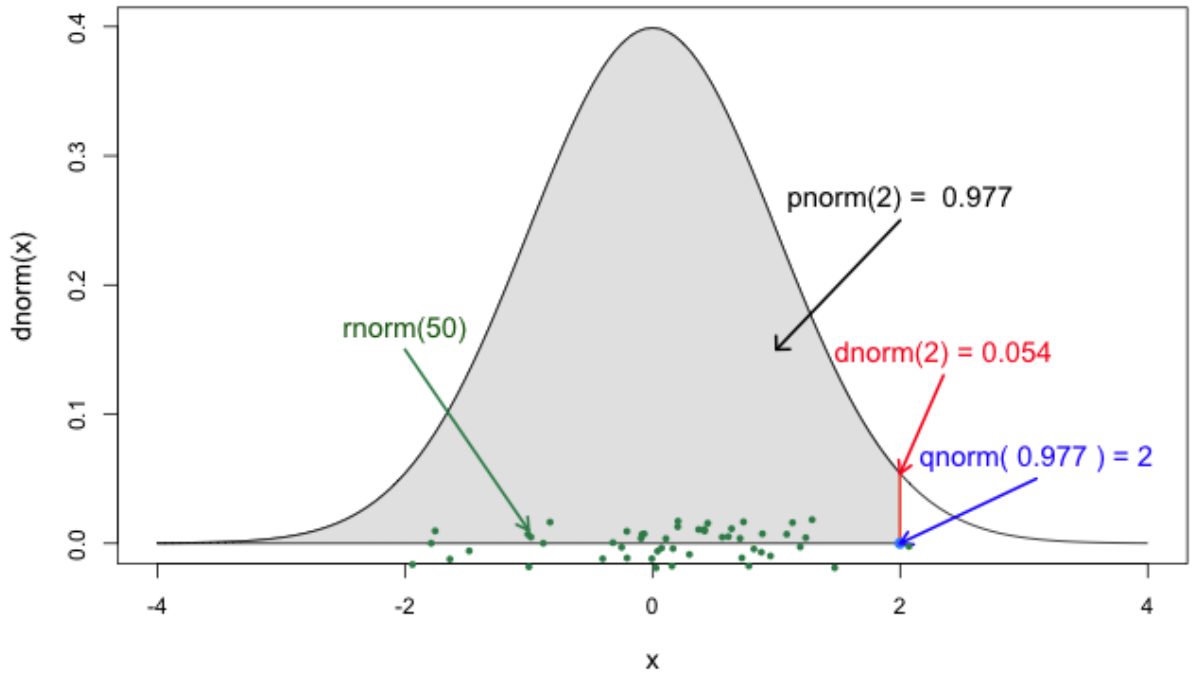
Gdje je:

$$z = \frac{\bar{x} - \mu}{\sigma} \quad \textbf{(25)}$$

Notacija navedenih jednadžbi:

z	varijabla standardizirane normalne razdiobe
$F(z)$	Standardna normalna kumulativna funkcija raspodjele s varijablom *z*
$\varphi_0(z)$	Standardna normalna funkcija gustoće vjerojatnosti s varijablom *z*

Radi jednostavnijeg razumijevanja koda napisanog u R programskom jeziku u nastavku će biti objašnjene četiri osnovne funkcije povezane s normalnom raspodjelom podataka. Kumulativna funkcija raspodjele opisana je funkcijom `pnorm(x)` te je označena površinom ispod krivulje koja daje vjerojatnost *p* za dani *x* [Slika 40.]. Njoj inverzna funkcija je `qnorm(x)` koja za zadanu vjerojatnost *p* izračunava vrijednost *x*. Funkcija `dnorm(x)` označava funkciju gustoće vjerojatnosti. Ona predstavlja vjerojatnost da varijabla poprimi određenu vrijednost te je na slici prikazana visinom funkcije u određenoj točki. Posljednja funkcija `rnorm(n)` generira *n* broj varijabli koje se ponašaju prema normalnoj razdiobi.



Slika 40. Funkcije povezane s normalnom raspodjelom podataka unutar R programskog jezika [28]

U nastavku je prikazan programski kod pomoću kojeg su izračunate vrijednosti rizika proizvođača R_P te rizika potrošača R_C za određenu mjernu nesigurnost u_m.

```
RC <- function(z) {((pnorm((AU-mean-sigma*z)/um) - pnorm((AL-mean-
sigma*z)/um)))*dnorm(z)}
RC1 <- integrate(RC, lower = -Inf, upper = (TL-mean)/sigma)
RC2 <- integrate(RC, lower = (TU-mean)/sigma, upper = Inf)
RC1$value + RC2$value

## [1] 0.01668411

RC1$abs.error + RC2$abs.error

## [1] 6.180198e-07

RP <- function(z) {(1 - (pnorm((AU-mean-sigma*z)/um) - pnorm((AL-mean-
sigma*z)/um)))*dnorm(z)}
integrate(RP, lower = ((TL-mean)/sigma), upper = ((TU-mean)/sigma))

## 0.1280509 with absolute error < 5.6e-08
```

$$Rizik\ potrošača\ R_C = 0{,}0167 = 1{,}67\ \% \quad \textbf{(26)}$$

$$Rizik\ proizvođača\ R_P = 0{,}1280 = 12{,}81\ \% \quad \textbf{(27)}$$

Ako se prikažu dobiveni rezultati ovog zadatka na uzorku od 100 proizvedenih aparata za zavarivanje, 86 strojeva zadovoljava zahtjeve potrošača dok 14 strojeva ne zadovoljava definirane zahtjeve. Prema izračunatome riziku proizvođača od 86 ispravnih strojeva prihvaćeno ih je 75, dok je njih 11 (12,8 %) proglašeno nesukladnim iako su zadovoljavali uvjete. Od 14 strojeva koji nisu zadovoljavali propisane uvjete svi su odbijeni kao nesukladni. Dakle, broj prihvaćenih nesukladnih strojeva se smanjio s 14 % na 1,67 % što znači da niti jedan kupac nije primio nevažeći stroj te je cilj ove kontrole kvalitete ispunjen. Zaključno, cijena koju proizvođač mora platiti kako bi se smanjio rizik od prihvaćanja nesukladnih strojeva očitava se u broju ispravnih strojeva koji su proglašeni nesukladnim, njih je 11 od 25 odbijenih što čini 44 %.

Ako proizvođač želi smanjiti svoj rizik i samim time vjerojatnost da se ispravan stroj proglasi nesukladnim, potrebno je uložiti u opremu za kontrolu kvalitete koja ima manju mjernu nesigurnost. U nastavku je napisan programski kod koji računa utjecaj mjerne nesigurnosti na rizik proizvođača i kupca.

```
t <- data.frame(x1 = c(200),
                x2 = c(2.7),
                x3 = c(196),
                x4 = c(204),
                x5 = c(196.000005, 196.00005, 196.0005,196.0025, 196.005,
1:500*0.05+196),
```

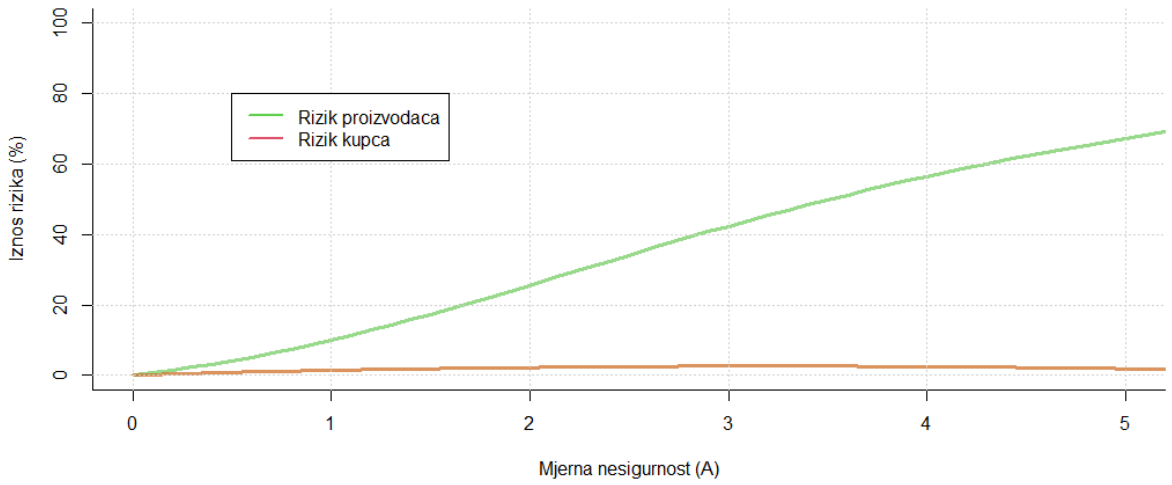
```
                x6 = c(203.999995, 203.99995, 203.9995,203.9975, 203.995,
-(1:500*0.05)+204),
                x7 = c(0.00001, 0.0001, 0.001, 0.005, 0.01, 1:500*0.1))

Func <- function(z, X) {
  x1 <- X[[1]]
  x2 <- X[[2]]
  x3 <- X[[3]]
  x4 <- X[[4]]
  x5 <- X[[5]]
  x6 <- X[[6]]
  x7 <- X[[7]]
  (1 - (pnorm((x6-x1-x2*z)/x7) - pnorm((x5-x1-x2*z)/x7)))*dnorm(z)}

t$RP <- apply(t, 1, function(x){
  integrate(Func, lower = ((TL-mean)/sigma) , upper = ((TU-mean)/sigma), X
= x)$value})

Func2 <- function(z, X) {
  x1 <- X[[1]]
  x2 <- X[[2]]
  x3 <- X[[3]]
  x4 <- X[[4]]
  x5 <- X[[5]]
  x6 <- X[[6]]
  x7 <- X[[7]]
  (pnorm((x6-x1-x2*z)/x7) - pnorm((x5-x1-x2*z)/x7))*dnorm(z)}

t$RC1 <- apply(t, 1, function(x){
  integrate(Func2, lower = -Inf , upper = ((TL-mean)/sigma), X =
x)$value})

t$RC2 <- apply(t, 1, function(x){
  integrate(Func2, lower = ((TU-mean)/sigma) , upper = Inf, X = x)$value})

t$RC = t$RC1 + t$RC2
t$RC_p = t$RC*100
t$RP_p = t$RP*100

names(t)[1] <- "srednja vrijednost"
names(t)[2] <- "sigma"
names(t)[3] <- "TL"
names(t)[4] <- "TU"
names(t)[5] <- "AL"
names(t)[6] <- "AU"
names(t)[7] <- "um"

t[10:25,c(7, 12, 13)]

##     um      RC_p      RP_p
## 10 0.5 0.8438893  4.200331
## 11 0.6 0.9841781  5.234538
## 12 0.7 1.1161677  6.335717
```

```
## 13 0.8 1.2403551  7.503455
## 14 0.9 1.3572216  8.736571
## 15 1.0 1.4672284 10.033121
## 16 1.1 1.5708185 11.390420
## 17 1.2 1.6684111 12.805091
## 18 1.3 1.7604031 14.273137
## 19 1.4 1.8471677 15.790027
## 20 1.5 1.9290571 17.350797
## 21 1.6 2.0063928 18.950155
## 22 1.7 2.0794665 20.582584
## 23 1.8 2.1485238 22.242433
## 24 1.9 2.2137484 23.924006
## 25 2.0 2.2752435 25.621627
```

Na kraju izraza generirana je tablica u kojoj prvi stupac označava redni broj, drugi mjernu nesigurnost, a treći i četvrti rizik proizvođača i kupca u postocima. Iz tablice se može vidjeti da ako proizvođač investira u bolju opremu za kontrolu kvalitete s manjom mjernom nesigurnosti dolazi i do smanjenja rizika i za proizvođača i za potrošača. Primjerice, ulaganje u opremu mjerne nesigurnosti 0,7 A smanjilo bi rizik proizvođača s 12,81 % na 6,33 % dok bi se rizik kupca smanjio s 1,67 % na 1,11 %. Drugi slučaj koji se može promatrati je da je početna investicija bila u opremu za kontrolu kvalitete s mjernom nesigurnosti 2 A. Korištenje takve opreme povećalo bi rizik potrošača na 2,27 %, dok bi se rizik proizvođača povećao čak na 25,62 %.

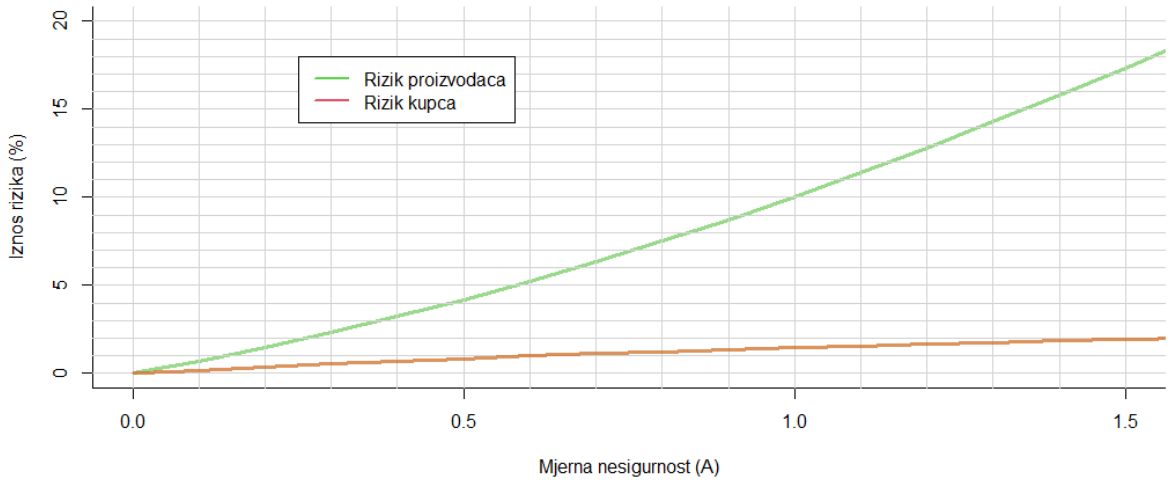
Sljedeći programski kod ranije spomenutu tablicu prikazuje grafički.

```
plot(t$RP_p ~ t$um ,
     type="l", bty="l" ,
     xlab="Mjerna nesigurnost (A)", ylab="Iznos rizika (%)" ,
     col=rgb(0.2,0.7,0.1,0.5),
     lwd=3 , pch=17 ,
     ylim = c(0,100), xlim = c(0,5))
lines(t$RC_p ~ t$um ,
      col=rgb(0.8,0.4,0.1,0.7),
      lwd=3 , pch=19 , type="l" )
grid()
legend(x = 0.5, y = 80,
       legend = c("Rizik proizvodaca", "Rizik kupca"),
       lty = c(1, 1),
       col = c(3, 2),
       lwd = 2)
```



Slika 41. Graf ovisnosti mjerne nesigurnosti o iznosu rizika proizvođača i kupca

```
plot(t$RP_p ~ t$um , type="l" , bty="l" ,
     xlab="Mjerna nesigurnost (A)" , ylab="Iznos rizika (%)" ,
     col=rgb(0.2,0.7,0.1,0.5),
     lwd=3 , pch=17 ,
     ylim=c(0,20), xlim=c(0,1.5))
abline(h = seq(0, 20, 1), col = "light grey")
abline(v = seq(0, 1.5, 0.1), col ="light grey")
lines(t$RC_p ~ t$um , col=rgb(0.8,0.4,0.1,0.7) ,
      lwd=3 , pch=19 , type="l" )

legend(x = 0.25, y = 18,
       legend = c("Rizik proizvodaca", "Rizik kupca"),
       lty = c(1, 1),
       col = c(3, 2),
       lwd = 2)
```

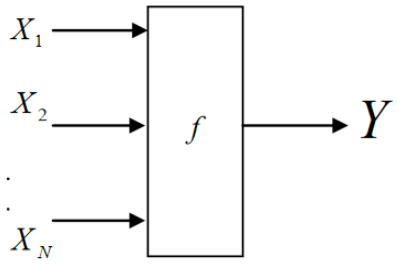


Slika 42. Graf ovisnosti mjerne nesigurnosti o iznosu rizika proizvođača i kupca - uvećano

Prema načinjenim linijskim grafovima [Slika 41., Slika 42.] vidljivo je da mjerna nesigurnost ima značajan utjecaj na rizik proizvođača za razliku od rizika kupca koji ima izrazito blagu krivulju rasta s porastom mjerne nesigurnosti. U idealnom slučaju kada bi mjerni sustav bio u potpunosti točan i precizan, tj. s vrlo malom mjernom nesigurnosti rizik proizvođača i kupca bio bi zanemariv. Kako idealan slučaj nije moguć, potrebno je pronaći balans između investiranja u mjerni sustav kojim se smanjuje broj krivo procijenjenih, tj. sukladnih proizvoda koji su proglašeni nesukladnim ili nesukladnih koji su proglašeni sukladnim te troškova koji proizvođač ima za svaki krivo procijenjeni proizvod.

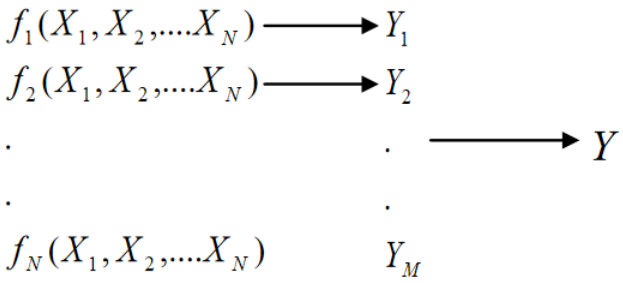
8. PROCJENA MJERNE NESIGURNOSTI

Mjerna nesigurnost opisuje rasipanje vrijednosti koja bi se razumno mogla pripisati mjerenoj veličini. U većini slučajeva mjerena veličina Y ne mjeri se izravno nego se određuje iz N drugih veličina $X_1, X_2, \dots, X_n$ na temelju funkcijskog odnosa koji predstavlja osnovni matematički model za potpuno određenje mjerene veličine [29].



Slika 43. Skalarni odnos između ulaznih veličina i mjerene veličine [29]

Ulazne veličine $X_1, X_2, \dots, X_n$ o kojima ovisi izlazna veličina Y mogu se promatrati kao mjerene veličine i mogu ovisiti o drugim veličinama dovodeći tako do složenog funkcijskog odnosa koji se ne mora uvijek moći eksplicitno napisati [29].



Slika 44. Vektorski odnos između ulaznih veličina i mjerene veličine [29]

Monte Carlo je metoda statističkog modeliranja koja se koristi u postupku procjenjivanja mjerne nesigurnosti rezultata mjerenja. Temelji se na generiranju određenog broja slučajnih varijabli koji su opisani funkcijom raspodjele za svaku ulaznu veličinu X_i i stvaranju odgovarajuće vrijednosti izlazne veličine Y, kombinirajući različite razdiobe kojima su definirane ulazne veličine. Postupak se ponavlja M puta te se na taj način tvori eksperimentalna funkcija gustoće vjerojatnosti podataka izlazne veličine koja se temelji na $M \cdot Y$ vrijednosti. Iz eksperimentalne funkcije gustoće vjerojatnosti slijedi procjena izlazne

veličine y, procijenjeno standardno odstupanje, te procjena intervala $y_{\left(\left(\frac{1-P}{2}\right)\cdot M\right)}, y_{\left(\left(\frac{1+P}{2}\right)\cdot M\right)}$ za zadanu vjerojatnost *P* [29].

Prema publikaciji EA-4/02 naslova „Izražavanje mjerne nesigurnosti pri umjeravanju“ u nastavku je prikazana usporedba GUM metode i Monte Carlo metode procjene mjerne nesigurnosti na primjeru umjeravanja pomičnog mjerila. Monte Carlo simulacije provedene su u R programskom jeziku.

Procjena mjerne nesigurnosti GUM metodom

Provedena je procjena mjerne nesigurnosti rezultata mjerenja umjeravanja pomičnog mjerila (0 – 150) mm / 0,05 mm.

Pogreška pokazivanja E_X pomičnog mjerila na referentnoj temperaturi t_0 = 20 °C dobiva se iz izraza:

$$E_{\mathrm{x}} = l_{\mathrm{iX}} - l_{\mathrm{S}} + L_{\mathrm{S}} \cdot \bar{\alpha} \cdot \Delta t + \delta L_{\mathrm{iX}} + \delta L_{\mathrm{M}} \qquad \textbf{(28)}$$

Gdje su:

l_{iX}	mm	pokazivanje pomičnog mjerila
l_{S}	mm	duljina granične mjerke
L_{S}	mm	nazivna duljina granične mjerke
$\bar{\alpha}$	$°\mathrm{C}^{-1}$	prosječni koeficijent toplinskog širenja pomičnog mjerila i granične mjerke
Δt	°C	razlika temperatura između pomičnog mjerila i granične mjerke
δL_{iX}	mm	ispravak zbog rezolucije pomičnog mjerila
δL_{M}	mm	ispravak zbog mehaničkih utjecaja i Abbeovog principa

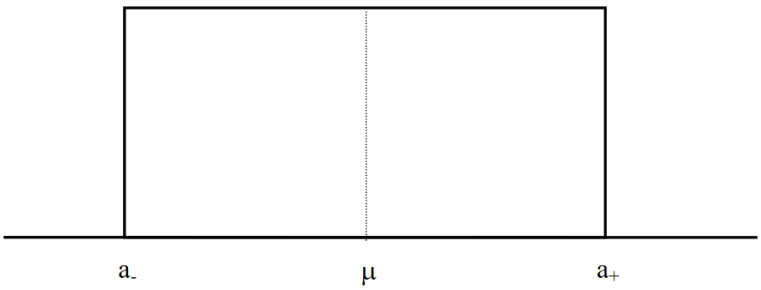
Procjena ulaznih veličina temelji na procjeni granica intervala pojavljivanja (od $-a_{\mathrm{p}}$ do $+a_{\mathrm{p}}$), uz jednaku vjerojatnost pojavljivanja unutar cijelog procijenjenog intervala, što znači da ne postoji vjerojatnost da ulazna veličina leži izvan tog intervala. U tom slučaju standardna se nesigurnost dobiva iz pravokutne razdiobe mogućih vjerojatnosti [Slika 45.] [29]. Svođenje procijenjenog standardnog odstupanja na razinu standardne nesigurnosti izvodi se prema izrazu:

$$u(x_{\mathrm{i}}) = a_{\mathrm{p}}/\sqrt{3} \qquad \textbf{(29)}$$

Gdje je:

$u(x_\mathrm{i})$ standardna nesigurnost

a_p granica intervala pravokutne raspodjele



Slika 45. Pravokutna razdioba [24]

Tablica 4. prikazuje doprinose nesigurnosti pojedine sastavnice mjerne nesigurnosti.

Tablica 4. Doprinosi mjerne nesigurnosti

Veličina X_i	Procjena x_i	Standardno odstupanje	Razdioba vjerojatnosti	Standardna nesigurnost $u(x_\mathrm{i})$	Koeficijent osjetljivosti c_i	Doprinos nesigurnosti $u_\mathrm{i}(y)$
$\boldsymbol{l_{\mathrm{iX}}}$	150,10 mm					
$\boldsymbol{\delta L_\mathrm{S}}$	150,00 mm	± 0,8 μm	pravokutna	0,46 μm	-1,0	-0,46 μm
$\boldsymbol{\delta t}$	0	± 2 °C	pravokutna	1,15 K	1,7	2,0 μm
$\boldsymbol{\delta l_{\mathrm{iX}}}$	0	± 25 μm	pravokutna	15 μm	1,0	15 μm
$\boldsymbol{\delta L_\mathrm{M}}$	0	± 50 μm	pravokutna	29 μm	1,0	29 μm
$\boldsymbol{E_\mathrm{x}}$	0,10 mm					33 μm

Prosječni koeficijent toplinskog širenja pomičnog mjerila i granične mjerke korišten za izračun vrijednosti δ*t* iznosi $\bar{\alpha} = 11{,}5 \cdot 10^{-6}\ °\mathrm{C}^{-1}$.

Procjenu (x_i) i standardno odstupanje prikazuje Tablica 4. Standardna nesigurnost za pravokutnu razdiobu vjerojatnosti izračunata je prema formuli (29), dok je doprinos mjernoj nesigurnosti izračunat prema sljedećoj formuli:

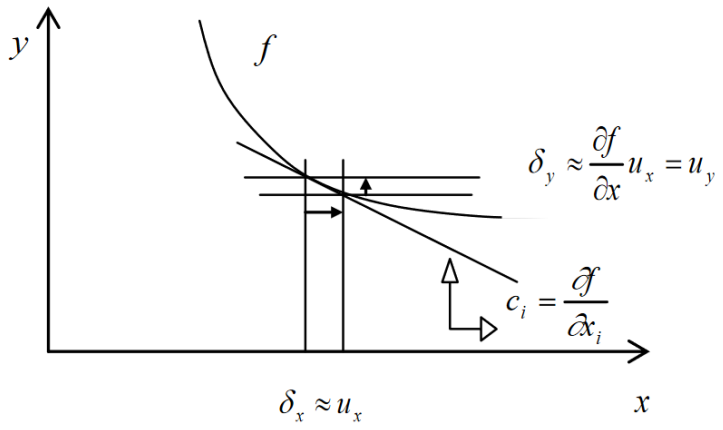
$$u_\mathrm{i}(y) = u(x_\mathrm{i}) \cdot c_\mathrm{i} \quad \textbf{(30)}$$

Gdje je:

c_i koeficijent osjetljivosti

$$c_\mathrm{i} = \frac{\partial f}{\partial x_\mathrm{i}} \quad \textbf{(31)}$$

Koeficijent osjetljivosti je derivacija koje opisuje kako se procjena vrijednosti izlazne veličine y mijenja s promjenama vrijednosti procjena ulaznih veličina $x_1, x_2, \ldots, x_n$ [29].



Slika 46. Koeficijent osjetljivosti [26]

Iz posljednjeg stupca tablice, tj. doprinosa mjerne nesigurnost vidljivo je da su ispravci zbog rezolucije pomičnog mjerila te zbog mehaničkih utjecaja i Abbeovog principa dva dominantna člana. Kada se javi takva situacija, može se primijeniti metoda kojom se ta dva člana kombiniraju u jedan dominantan član te se u tome slučaju standardna nesigurnost koju treba pridružiti mjernom članu može napisati kao [30]:

$$u(y) = \sqrt{{u_0}^2(y) + {u_\mathrm{R}}^2(y)} \quad \textbf{(32)}$$

pri čemu:

$$u_0(y) = \sqrt{{u_1}^2(y) + {u_2}^2(y)} \quad \textbf{(33)}$$

$$u_0(y) = \sqrt{15^2 + 29^2} \quad \textbf{(34)}$$

$$u_0(y) = 32{,}65\ \mu\mathrm{m} \quad \textbf{(35)}$$

označuje doprinos mjernoj nesigurnosti dominantnih članova, dok

$$u_\mathrm{R}(y) = \sqrt{\sum_{\mathrm{i}=3}^{N} {u_\mathrm{i}}^2(y)} \quad \textbf{(36)}$$

$$u_\mathrm{R}(y) = \sqrt{(-0.46)^2 + 2^2} \quad \textbf{(37)}$$

$$u_\mathrm{R}(y) = 2{,}05\ \mu\mathrm{m} \quad \textbf{(38)}$$

doprinos mjernoj nesigurnosti preostalih nedominantnih članova.

Zaključno, mjerna nesigurnost koju treba pridružiti mjernom članu iznosi:

$$u_0(y) = \sqrt{32{,}65^2 + 2{,}05^2}$$

$$u_0(y) \approx 33\ \mu\text{m}$$

Ako dva dominantna doprinosa potječu iz pravokutnih razdioba vrijednosti s poluširinama a_1 i a_2, razdioba koja se dobije kao rezultat njihove konvolucije simetrična je trapezna razdioba s poluširinama [30]:

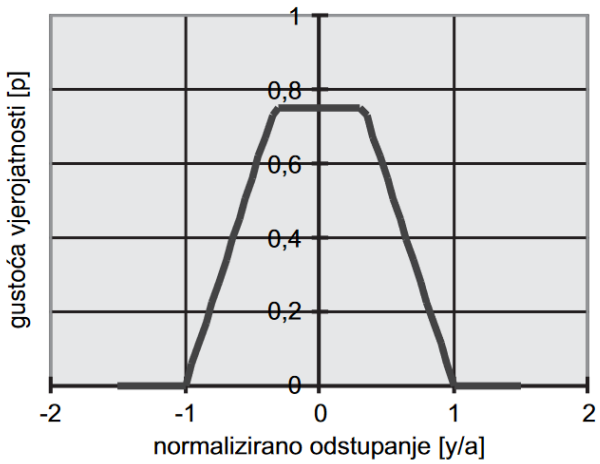
$$a = a_1 + a_2 \; i \; b = |a_1 - a_2| \tag{39}$$

Gdje je:

a	donja osnovica trapeza
b	gornja osnovica trapeza

$$a = 25 + 50 = 75\ \mu m \tag{40}$$

$$b = |25 - 50| = 25 \mu m \tag{41}$$



Slika 47. Simetrična trapezna razdioba [27]

Kako se radi o trapeznoj raspodjeli, da bi se izračunala proširena mjerna nesigurnost *U* potrebno je odrediti faktor pokrivanja i vrijednost parametra brida prema idućim formulama:

$$\beta = \frac{b}{a} = \frac{|a_1 - a_2|}{a_1 + a_2}$$

Gdje je:

β	parametar brida

$$\beta = \frac{|25 - 50|}{25 + 50} = 0{,}33$$

Faktor pokrivanja k_p za vjerojatnost pokrivanja od 95 % koja odgovara trapeznoj razdiobi s parametrom brida $\beta < 0{,}95$ izračunava se iz izraza:

$$k_\mathrm{p} = \frac{1 - \sqrt{(1-p)(1-\beta^2)}}{\sqrt{\frac{1+\beta^2}{6}}}$$

$$k_\mathrm{p} = \frac{1 - \sqrt{(1-0.95)(1-0.33^2)}}{\sqrt{\frac{1+0.33^2}{6}}}$$

$$k_p = 1{,}83$$

Zaključno, proširena mjerna nesigurnosti iznosi:

$$U = k_\mathrm{p} \cdot u(E_x) = 1{,}83 \cdot 0{,}033 \text{ mm} = 0{,}06 \text{ mm}.$$

Procjena mjerne nesigurnosti Monte Carlo metodom

Pomoću Monte Carlo simulacije čije su glavne karakteristike opisane u uvodnom dijelu ovog poglavlja provedena je procjena mjerne nesigurnosti na primjeru umjeravanja pomičnog mjerila.

Prema tablici ulazne veličine (x_i) definirane su funkcijama gustoće vjerojatnosti $g(x_i)$. Ulazne veličine se koriste za računanje pogreške pokazivanja E_X pomičnog mjerila na referentnoj temperaturi t_0 = 20 °C prema već ranije spomenutoj formuli:

$$E_\mathrm{x} = l_\mathrm{iX} - l_\mathrm{S} + L_\mathrm{S} \cdot \bar{\alpha} \cdot \Delta t + \delta L_\mathrm{iX} + \delta L_\mathrm{M} \qquad \textbf{(42)}$$

Za izvedbu Monte Carlo simulacije korišten je R programski jezik.

Tablica 5. Ulazne veličine (x_i) definirane su funkcijama gustoće vjerojatnosti $g(x_i)$ za izračun pogreške pokazivanja E_x

Naziv	**Veličina x_i**	**Oznaka veličine u R kodu x_i**	**Funkcija gustoće vjerojatnosti $g(x_i)$**
pokazivanje pomičnog mjerila	l_iX	x_1	150,10 mm
duljina granične mjerke	l_S	x_2	Pravokutna razdioba (M; 150 mm – 0,8 μm; 150 + 0,8 μm)
nazivna duljina granične mjerke	L_S	x_3	150 mm

prosječni koeficijent toplinskog širenja pomičnog mjerila i granične mjerke	$\bar{\alpha}$	x_4	$11{,}5 \cdot 10^{-6}\ °C^{-1}$
razlika temperatura između pomičnog mjerila i granične mjerke	Δt	x_5	Pravokutna razdioba (M; -2 °C; 2 °C)
ispravak zbog rezolucije pomičnog mjerila	δl_{iX}	x_6	Pravokutna razdioba (M; -25 µm; 25 µm)
ispravak zbog mehaničkih utjecaja i Abbeovog principa	δL_M	x_7	Pravokutna razdioba (M; -50 µm; 50 µm)

U nastavku je prikazan programski kod pomoću kojeg su ulazne vrijednosti (x_i) i pripadne vrijednosti funkcija gustoće vjerojatnosti $g(x_i)$ pohranjene u R programski jezik kao vrijednosti veličina.

```
M=1000000
set.seed(1234)
x1 <- 150.1
x2 <- runif(M, min = 150-0.8*10^(-3), max= 150+0.8*10^(-3))
x3 <- 0.15
x4 <- 11.5 *10^(-6)
x5 <- runif(M, min = -2, max = 2)
x6 <- runif(M, min = -25*10^(-3), max = 25*10^(-3))
x7 <- runif(M, min = -50*10^(-3), max = 50*10^(-3))
```

Funkcija `set.seed` postavlja početni broj koji se koristi za generiranje niza slučajnih brojeva, time se osigurava da će se u simulaciji dobiti isti rezultat svaki put kada pokrenete isti postupak koji koristi isti skup generiranih varijabli. Oznaka M predstavlja broj simulacija.

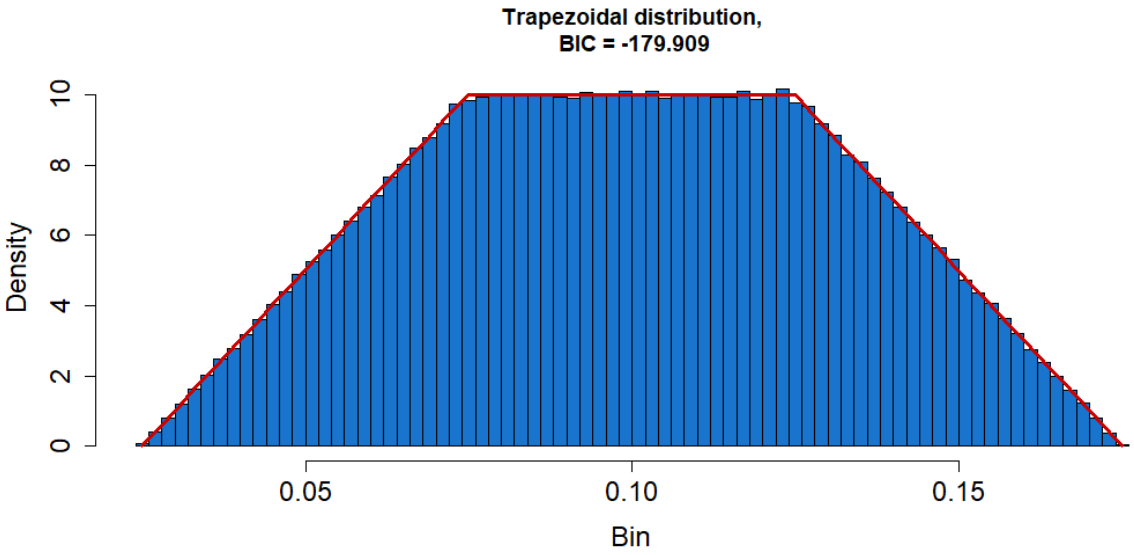
Nakon unosa podataka učitan je ranije instalirani programski paket nadogradnje `propagate`. Označene su veličine koje će biti korištene u daljnjem postupku i zapisana funkcija po kojoj se računa procjena izlazne veličine E_x. Funkcija `propagate` istoimenog paketa provodi prijenos razdiobi na temelju razvoja prvog i drugog Taylorog reda i Monte Carlo simulacije.

Prijenos razdiobi je metoda koja se upotrebljava za određivanje razdiobe vjerojatnosti izlazne veličine iz razdioba vjerojatnosti pridruženih ulaznim veličinama o kojima ovisi izlazna veličina [31].

Funkcijom `fitDistr` ranije određena funkcija razdiobe vjerojatnosti izlazne veličine E_x se uklapa u 32 različite funkcije razdiobe te se određuje najbolja.

```
library(propagate)
```

```
Ex <- cbind(x1, x2, x3, x4, x5, x6, x7)
Exx <- expression(x1 - x2 + (x3 * x4 * x5) + x6 +x7)
Exxx <- propagate(Exx, data = Ex, do.sim = TRUE)
fitDistr(Exxx)
```



Slika 48. Funkcija gustoće vjerojatnosti izlazne veličine (Trapezna razdioba)

```
## Best fit is Trapezoidal Distribution.

## Parameters:

##          a          b          c          d
## 0.02495840 0.07479625 0.12498074 0.17508338

## Standard errors:

##             a             b             c             d
## 0.0001234340 0.0001502054 0.0001426365 0.0001209881

## Goodness of fit:

## BIC = -179.9088
```

Pokretanjem ranije napisanog koda dobije se histogram podataka i pripadna funkcija gustoće vjerojatnosti koja je dodijeljena izlaznoj veličini. U ovom slučaju je to trapezna razdioba sa parametrima a_t, b_t, c_t *i* d_t koji označavaju vrijednosti *x* koordinata vrhova trapeza. Vrijednosti parametara trapeza, pripadne standardne pogreške i vrijednost BIC parametra prikazani su u gornjem izvještaju.

Koordinate vrhova trapeza na x-osi iznose:

$$a_t = 0{,}02495840 \text{ mm} \quad \textbf{(43)}$$

$$b_t = 0{,}07479625 \text{ mm} \quad \textbf{(44)}$$

$$c_t = 0{,}12498074 \text{ mm} \quad \textbf{(45)}$$

$$d_t = 0{,}17508338 \text{ mm} \quad \textbf{(46)}$$

prema tome donja osnovica trapeza iznosi:

$$d_t - a_t = 0{,}17508338 - 0{,}02495840 = 0{,}15 \text{ mm} = 150 \text{ µm} \quad \textbf{(47)}$$

dok gornja iznosi:

$$c_t - b_t = 0{,}12498074 - 0{,}07479625 = 0{,}05 \text{ mm} = 50 \text{ µm} \quad \textbf{(48)}$$

Prema grafu te prema numeričkim podacima vidljivo je da su koordinate točki gornje i donje osnovice trapeza približno simetrične te se može utvrditi da se radi simetričnom trapezu s poluširinama donje osnovice (a) i gornje osnovice (b) u iznosima:

$$a \approx 75 \text{ µm} \quad \textbf{(49)}$$

$$b \approx 25 \text{ µm} \quad \textbf{(50)}$$

čime je potvrđena konstatacija da ako dva dominantna doprinosa nesigurnosti potječu iz pravokutnih razdioba vrijednosti s poluširinama a_1 i a_2, razdioba koja se dobije kao rezultat njihove konvolucije simetrična je trapezna razdioba s poluširinama $a = a_1 + a_2$ i $b = |a_1 - a_2|$.

Nadalje potrebno je odrediti interval povjerenja $(1 - \alpha)$ od 95 %, odnosno odrediti u kojem intervalu oko osi simetrije trapeza ploština trapeza iznosi 95 %.

Nakon učitavanja programskog paketa `trapezoid` u kojem se nalaze razne funkcije za trapeznu raspodjelu koristi se funkcija `qtrapezoid` koja za određeni interval povjerenja p određuje vrijednost x.

```
library(trapezoid)

qtrapezoid(0.975,
           min = 0.02495840,
           mode1 = 0.07479625,
           mode2 = 0.12498074,
           max = 0.17508338,
           lower.tail = TRUE)

## [1] 0.1592435

qtrapezoid(0.975,
           min = 0.02495840,
           mode1 = 0.07479625,
           mode2 = 0.12498074,
```

```
            max = 0.17508338,
            lower.tail = FALSE)

## [1] 0.04075634
```

Kako se radi o simetričnoj trapeznoj raspodjeli, određena je vrijednost *x* za interval povjerenja 97,5 % s desne strane osi te za isti takav interval povjerenja s lijeve strane osi.

Nakon pokretanja programskog koda u izvještaju su prikazane tražene granice intervala koje iznose:

$$x_1 = 0{,}04 \text{ mm} \quad \textbf{(51)}$$

$$x_2 = 0{,}16 \text{ mm} \quad \textbf{(52)}$$

Kada se uzme u obzir da se os trapeza nalazi na 0,1 mm, vidljivo je da je su dobivene vrijednosti x_1 i x_2 simetrične što potvrđuje dobivenu proširenu mjernu nesigurnost od 0,06 mm.

Zaključno se može konstatirati da su dobivene iste vrijednosti proširene mjerne nesigurnosti pomoću GUM metode i pomoću Monte Carlo metode.

STATISTIČKA ANALIZA TOLERANCIJA

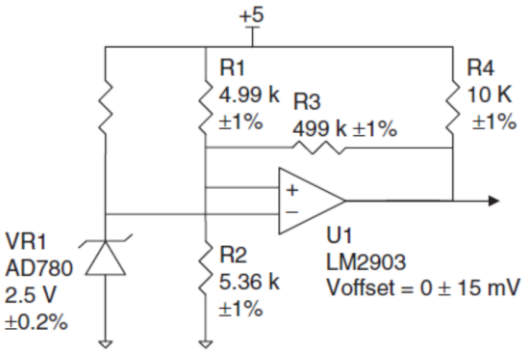
Tolerancija je dopušteno odstupanje od ciljane vrijednosti. Statistička analiza tolerancija se provodi kada se želi utvrditi izlazna vrijednost određene karakteristike kvalitete koja ovisi o nekoliko ulaznih faktora, te je tu ovisnost moguće opisati određenom funkcijom. Na osnovu vrijednosti ulaznih varijabli i pripadnih dopuštenih odstupanja računa se vrijednost izlazne veličine. U nastavku je primjenom već spomenutog postupka Monte Carlo simulacije prikazan primjer utjecaja statističkih tolerancija na izlaznu vrijednost napona strujnog kruga.

Primjer: Strujni krug

Potrebno je konstruirati napajanje od 5 V, koje uključuje krug za reguliranje ispravnosti funkcije sustava. Ako iznos izlaza napajanja od 5 V padne prenisko, detektor prisiljava mikroprocesor da prijeđe u sigurno stanje (engl. *safe shutdown state*). Kada se uključi napajanje, iznos izlaza polako raste. Detektor mora držati mikroprocesor u sigurnom stanju dok iznos napona bude ponovno prihvatljiv. Sustav, tj. krug ima dva ključna parametra, $V_{\text{Trip-Up}}$ te $V_{\text{Trip-Down}}$. $V_{\text{Trip-Up}}$ je napon kada detektor mijenja stanje dok napon raste. Slično, $V_{\text{Trip-Down}}$ je napon kada detektor mijenja stanje kada napon pada. Napajanje bi trebalo regulirati svoj izlaz na 5 V +/- 2 %, što bi iznosilo 4,90 V – 5,10 V. Mikroprocesor zahtjeva 5 V +/- 5 % što

bi iznosilo 4,75 V – 5,25 V. Da bi krug normalno funkcionirao, mora vrijediti: $V_{\text{Trip-Up}} > V_{\text{Trip-Down}}$. S time se izbjegavaju nestabilni signali. Gornja granica tolerancija $V_{\text{Trip-Up}}$ je na 4,90, a donja granica tolerancija $V_{\text{Trip-Down}}$ je na 4,75 [32].

Krug uključuje V_{R1}, referentu diodu, te U_1 komparator. Otpornici R_1 i R_2 namještaju napon promjene stanja, dok otpornik R_3 daje pozitivnu povratnu informaciju da bi se utvrdilo vrijedi li [32]: $V_{\text{Trip-Up}} > V_{\text{Trip-Down}}$



Slika 49. Shema strujnog kruga

Naponi promjene stanja strujnog kruga definirani su jednadžbama:

$$V_{\text{Trip-Up}} = (V_{\text{R1}} + V_{\text{Offset}})\left[1 + \frac{R_1(R_2 + R_3)}{R_2 \cdot R_3}\right] \quad \textbf{(53)}$$

$$V_{\text{Trip-Down}} = (V_{\text{R1}} + V_{\text{Offset}})\left[1 + \frac{R_1(R_3 + R_4)}{R_2(R_1 + R_3 + R_4)}\right] \quad \textbf{(54)}$$

Pomoću Monte Carlo simulacije čije su glavne karakteristike opisane u uvodnom dijelu ovog poglavlja provedena je analiza tolerancija te procjena sposobnosti procesa na primjeru strujnog kruga.

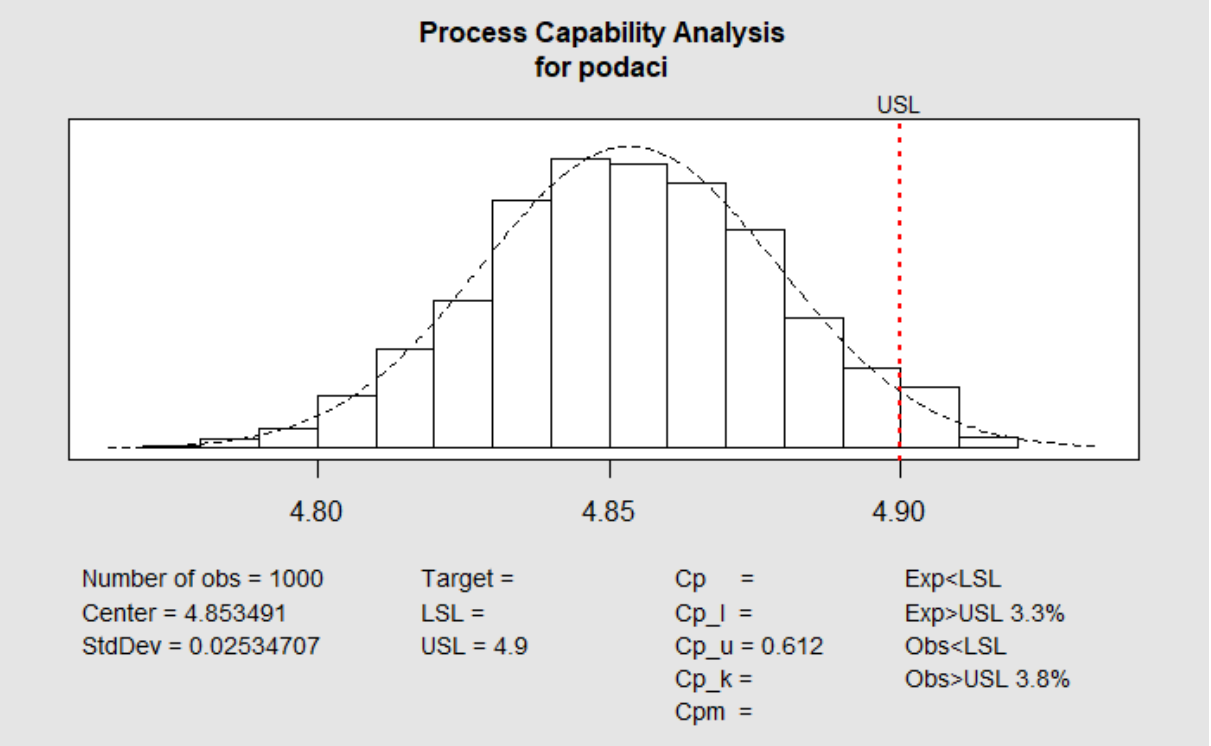
Ulazne veličine (x_i) definirane su funkcijama gustoće vjerojatnosti $g(x_i)$ gdje je $M = 1000$ [Tablica 6.]. Ulazne veličine se koriste za računanje faktora kvalitete koji se testiraju, a to su vrijednosti napona $V_{Trip\text{-}Up}$ i $V_{Trip\text{-}Down}$.

Tablica 6. Ulazne veličine (x_i) definirane su funkcijama gustoće vjerojatnosti $g(x_i)$ za izračun karakteristika kvalitete strujnog kruga

Naziv	**Veličina** x_i	**Funkcija gustoće vjerojatnosti $g(x_i)$**
Otpor na R_1	R_1	Pravokutna raspodjela (M; 4940,1; 5039,9) Ω
Otpor na R_2	R_2	Pravokutna raspodjela (M; 5306,4; 5413,6) Ω
Otpor na R_3	R_3	Pravokutna raspodjela (M; 494010; 503990) Ω

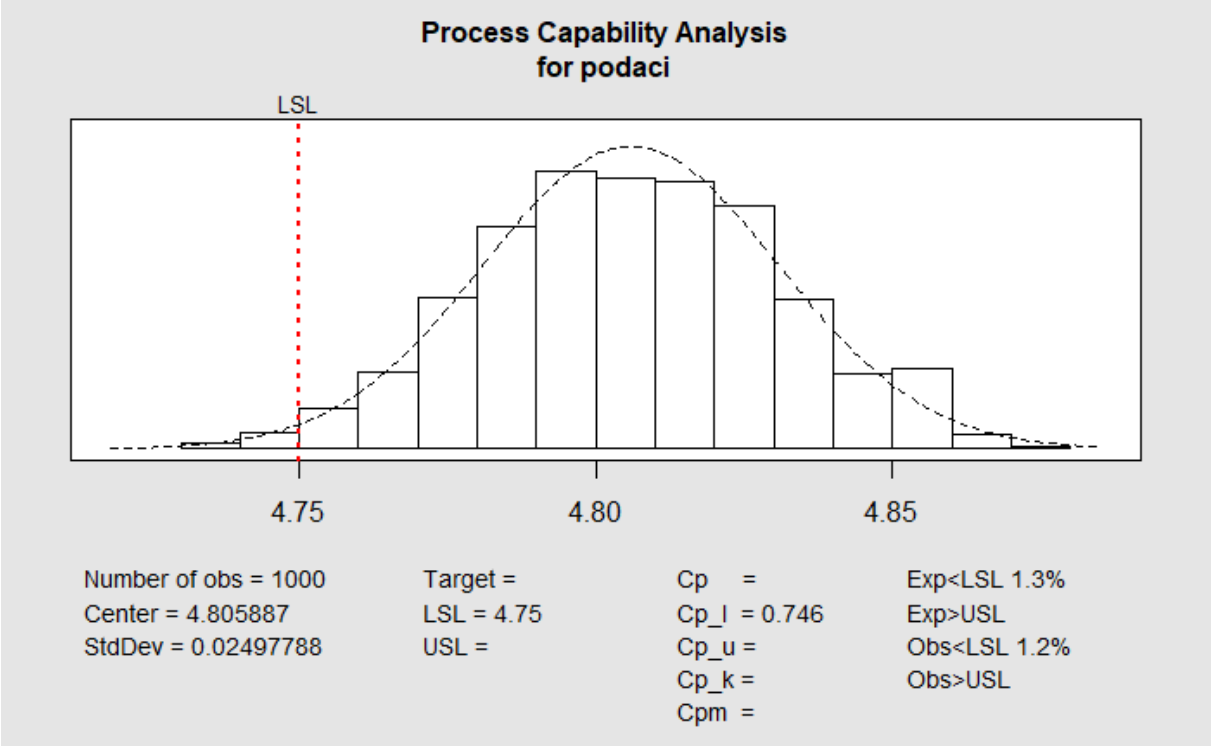
Otpor na R4	R_4	Pravokutna raspodjela (*M*; 9900; 10100) Ω
Napon na diodi	V_{R1}	Pravokutna razdioba (*M*; 2,495; 2,505) V
Odstupanje napona	V_{Offset}	Pravokutna razdioba (*M*; -0,015 µm; 0,015 µm)

```
library(mc2d)

set.seed(1234)
R1 <- runif(1000, min = 4940.1, max = 5039.9)
R2 <- runif(1000, min = 5306.4, max = 5413.6)
R3 <- runif(1000, min = 494010, max = 503990)
R4 <- runif(1000, min = 9900, max = 10100)
VR1 <- runif(1000, min = 2.495, max = 2.505)
Voff <- runif(1000, min = -0.015, max = 0.015)

Vup = (VR1 + Voff)*(1+((R1 *(R2 + R3))/(R2 * R3)))
Vdown = (VR1 + Voff)*(1+((R1 *(R3 + R4))/(R2*(R1 + R3 + R4))))
zadatak <- data.frame(R1, R2, R3, R4, VR1, Voff, Vup, Vdown)

library(qcc)

Mjerenja <- Vup
velicina_uzorka <- 1
Broj_uzoraka <- length(Mjerenja)/velicina_uzorka
uzorak <- rep(1:Broj_uzoraka, each = velicina_uzorka)
df <- data.frame(Mjerenja, uzorak)
podaci <- with(df, qcc.groups(Mjerenja, uzorak))
xbar <- qcc(podaci, type = "xbar", plot = FALSE)
process.capability(xbar, spec.limits = c(NA, 4.9))
```




```
library(qcc)
```

```
Mjerenja <- Vdown
velicina_uzorka <- 1
Broj_uzoraka <- length(Mjerenja)/velicina_uzorka
uzorak <- rep(1:Broj_uzoraka, each = velicina_uzorka)
df <- data.frame(Mjerenja, uzorak)
podaci <- with(df, qcc.groups(Mjerenja, uzorak))
xbar <- qcc(podaci, type = "xbar", plot = FALSE)
process.capability(xbar, spec.limits = c(4.75, NA))
```



Kako bi proizvod postigao uspjeh na tržištu, potrebno je provoditi kontrolu kvalitete pri konstruiranju proizvoda. U zadatku postoje dva faktora kontrole kvalitete koji se testiraju i na osnovu čijih rezultata se odlučuje hoće li se ići u proizvodnju ili su potrebne dodatne konstrukcijske dorade. Prvi faktor je $V_{\text{Trip-Up}}$ koji ima određenu gornju granicu tolerancije kako bi sustav ispravno funkcionirao, a ona iznosi koja iznosi 4,9 V. Prema podacima koji su generirani u 3,8 % iznos napona prelazio je gornju granicu specifikacije, a prema procijenjenim očekivanjima taj iznos je nešto manji (3,3 %). Gornja potencijalna sposobnost ovog faktora kvalitete iznosi 0,612. Drugi faktor je $V_{\text{Trip-Down}}$ koji ima određenu donju granicu tolerancije kako bi sustav ispravno funkcionirao, a ona iznosi 4,75 V. Prema podacima koji su vidljivi na grafovima očekivanje da će napon biti manji od zadane granice iznosi 1,3 %, a prema simuliranim podacima 1,2 %. Donja potencijalna sposobnost ovog faktora iznosi 0,746.

Prema dobivenim podacima može se zaključiti kako proizvod nije dobro konstruiran, proizvodanja ovakvog proizvoda donijela bi samo troškove organizaciji jer ne bi bio konkurentan na tržištu. Potrebno je provesti daljnju analizu i rekonstruirati proizvod.

9. ZAKLJUČAK

Inženjerstvo kvalitete nameće se kao efikasan odgovor na osiguranje povoljnog omjera cijene i kvalitete proizvoda što osigurava njegovu konkurentnost na tržištu. Dostupnost informacija razvojem digitalizacije rapidno raste, strategija prikupljanja i obrade istih važan je dio u inženjerstvu kvalitete jer se njene tehnike i alati temelje na podacima o određenim procesima. Kada su podaci prikupljeni, potrebno je pronaći odgovarajuću računalnu podršku koja podržava tehnike i alate inženjerstva kvalitete, jedan od takvih je R programski jezik. R programski jezik je besplatan softver otvorenog koda čijom su primjenom prikazani neki od alata i tehnika te riješeni određeni primjeri iz inženjerstva kvalitete.

Na početku je prikazano 7 jednostavnih alata kontrole kvalitete. Kao što pokazuje krivulja učenja prikazana u radu [Slika 8] potrebno je dosta vremena da se nauče osnove rada u R-u. Vrlo je vjerojatno da bi se ovih 7 alata brže napravilo u nekom od konvencionalnih softvera koji korisniku omogućuju da slijednom klikova postigne željeni cilj.

Nakon prikupljanja podataka, kako bi se mogli koristiti u daljnjoj analizi potrebno ih je opisati nekom od teorijski raspodjela. Prilagodba pravokutne i trokutaste teorijske raspodjele na empirijske podatke prikazana je u radu. Ove raspodjele su odabrane jer neki drugi softveri koji imaju ovu funkciju nemaju mogućnost prilagodbe na spomenute teorijske razdiobe.

Važno je naglasiti da se R programski jezik sastoji od velikog broja programskih paketa koji funkcioniraju kao mali softveri unutar njega, stoga postoje različiti oblici istog alata. Primjerice, u radu su korištena tri različita paketa kojima se računa sposobnost procesa. Svaki od njih ima neku prednost. Prvi upozorava korisnika ukoliko normalnost podataka nije zadovoljena, drugi uz sposobnost računa i postotak podataka koji se ne nalazi unutar tolerancijskog intervala, a treći omogućuje računanje sposobnosti procesa za različite teorijske raspodjele.

Kvaliteta je definirana zahtjevima kupaca. Stoga kada kupac postavi određene zahtjeve proizvođač mora znati je li ih u mogućnosti ispuniti i koliki rizik preuzima. U radu je prikazan način na koji mjerna nesigurnost utječe na ocjenu sukladnosti proizvoda, a time i na podjelu rizika između kupca i proizvođača. Zaključeno je da proizvođač mora pronaći balans između investiranja u mjerni sustav kojim se smanjuje broj krivo procijenjenih tj. sukladnih proizvoda

koji su proglašeni nesukladnim ili nesukladnih koji su proglašeni sukladnim te troškova koji proizvođač ima za svaki krivo procijenjeni proizvod.

Nadalje, napravljena je usporedba GUM metode i metode Monte Carlo procjene mjerne nesigurnosti na primjeru umjeravanja pomičnog mjerila. Vrijednosti ulaznih parametara simulirane su pomoću Monte Carlo metode te su korištenjem paketa `propagate` dobiveni podaci o proširenoj mjernoj nesigurnoati u postuku umjeravanja pomičnog mjerila koji se podudaraju s rezultatima GUM metode.

Nakon umjeravanja pomičnog mjerila primjenom Monte Carlo simulacija napravljena je analiza statističkih tolerancija dviju komponenti kvalitete konstruiranog napajanja od 5 V. Zaključeno je da proizvod nije dobro konstruiran te da je potrebna analiza i dorada proizvoda prije same proizvodnje.

Može se zaključiti da je primjena R programskog jezika vrlo široka. Nakon što se nauče osnove jezika i prihvati drugačiji način rada (pisanje koda, a ne kliktanje ili „drag and drop“), mogućnosti korisnika rastu te prerastaju mogućnosti ostalih tradicionalnih programa koji se koriste u postupcima ostvarivanja kvalitete. Strojno učenje bio bi idući korak u primjeni R programskog jezika u inženjerstvu kvalitete.

LITERATURA

[1] Oakland, J. *Statistical process control*. Jordan Hill, Oxford, UK, 2008.

[2] ISO 9000:2015: Quality management systems — Fundamentals and vocabulary https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en Pristupljeno: 28. travnja 2021.

[3] Klaić, B. Rječnik stranih riječi: Tuđenice i posuđenice. Zagreb: Nakladni zavod MH, 1990.

[4] Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, 2021.. http://www.enciklopedija.hr/Natuknica.aspx?ID=34865 Pristupljeno: 3. svibnja 2021.

[5] Runje B. Predavanja iz kolegija Osnove osiguravanja kvalitete. Zagreb: Fakultet strojarstva i brodogradnje, 2018.

[6] Chao-Ton Su. Quality Engineering: Off-Line Methods and Applications. Boca Raton: CRC Press; 2013.

[7] Emilio L. Cano, Javier M. Moguerza, Mariano Prieto Corcoba: Quality Control with R. 2nd ed, Springer; 2009.

[8] Rostamkhani R, Karbasian M. Quality Engineering Techniques. 1st ed. Boca Raton: CRC Press; 2020.

[9] Runje B, Horvatić Novak A. Predavanja iz kolegija Upravljanje rizicima, Metode Procjene rizika. Zagreb: Fakultet strojarstva i brodogradnje, 2019.

[10] Williams A.R.T., Van der Wiele A., B.G. Dale B.G. Quality costing: a management review. International Journal of Management Reviews; 1999.

[11] Roger D. Peng. R programming for Dana Science: https://bookdown.org/rdpeng/rprogdatascience/ Pristupljeno: 17. ožujka 2021.

[12] R Advantages and Disadvantages: https://www.javatpoint.com/r-advantages-and-disadvantages, Pristupljeno: 25. ožujka 2021.

[13] Lisjak D, Kolar D. Predavanja iz kolegija Informatički menadžment, Prediktivna analitika. Zagreb: Fakultet strojarstva i brodogradnje; 2020.

[14] Službena web stranica R programskog jezika: https://www.r-project.org/ Pristupljeno: 1. ožujka 2021.

[15] St. Olaf College, Minnesota. R vs. RStudio: https://wp.stolaf.edu/it/r-vs-rstudio/, Pristupljeno: 25. ožujka 2021.

[16] Razumić A. Osnovni statistički alati za analizu podataka [diplomski rad]. Zagreb: Fakultet strojarstva i brodogradnje; 2018.

[17] Erdeg M. Metode procjene parametara [diplomski rad]. Osijek: Sveučilište J. J. Strossmayera u Osijeku, Odjel za matematiku; 2016.

[18] Fit of univariate distributions to non-censored data: https://rdrr.io/cran/fitdistrplus/man/fitdist.html Pristupljeno: 11. travnja 2021.

[19] Engineering statistics handbook. Kolmogorov-Smirnov Goodness-of-Fit Test: 1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test (nist.gov) Pristupljeno: 12. travnja 2021.

[20] Kolmogorov-Smirnov tablica: https://www.real-statistics.com/statistics-tables/kolmogorov-smirnov-table/ Pristupljeno: 12. travnja 2021.

[21] Goodness of Fit Test for a Growth Analysis: https://www.ge.com/digital/documentation/meridium/V36000/WebHelp/mergedProjects/WebHelp_Master/mergedProjects/Reliability/Goodness_of_Fit_Test_for_a_Growth_Analysis.htm Pristupljeno 13. travnja 2021.

[22] Engineering statistics handbook. Anderson-Darling test: 1.3.5.14. Anderson-Darling Test (nist.gov) Pristupljeno 13. travnja 2021.

[23] Jäntschi L, Bolboacă S.D. Computation of Probability Associated with Anderson–Darling Statistic. Cluj: Technical University of Cluj-Napoca; 2018.

[24] Malić I. Generalizirani linearni model na panel podatcima [diplomski rad]. Osijek: Sveučilište J. J. Strossmayera u Osijeku, Odjel za matematiku; 2014.

[25] JCGM 106:2012 Evaluation of measurement dana – The role of measurement uncertainty in conformity assessment. JCGM, 2012.

[26] Forbes A. B. Measurment uncertainty and otpimized conformance assessment. Measurement Vol 39, 2006, str.808-814.

[27] Runje B; Horvatic Novak A; Razumic A; Piljek P; Strbac B & Orosnjak M (2019). Evaluation of Consumer and Producer Risk in Conformity Assessment Decisions, Proceedings of the 30th DAAAM International Symposium, pp.0054-0058, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-22-8, ISSN 1726-9679, Vienna, Austria DOI: 10.2507/30th.daaam.proceedings.007

[28] Lecture 13, Prof. Jack Weiss, University of North Carolina https://sakai.unc.edu/access/content/group/3d1eb92e-7848-4f55-90c3-7c72a54e7e43/public/docs/lectures/lecture13.htm Pristupljeno: 15. travnja 2021.

[29] Runje, B. Predavanja iz kolegija Teorija i tehnika mjerenja. Zagreb: Fakultet strojarstva i brodogradnje; 2017.

[30] EA 4-02 – Izračunavanje mjerne nesigurnosti pri umjeravanju. Zagreb: Državni zavod za mjeriteljstvo, 2008.

[31] JCGM 101:2008 – Vodič za procjenu mjerne nesigurnosti – Dodatak 1: Monte Carlo metoda

[32] Mažar M. Statistička analiza tolerancija [završni rad]. Zagreb: Fakultet strojarstva i brodogradnje; 2011.