Primjena statističkih metoda analize i optimizacije proizvodnog procesa uporabom dostupnog programskog paketa

Lukač, Sandro

Undergraduate thesis / Završni rad

2015

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:235:220118

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-04-25**

Repository / Repozitorij:

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

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

ZAVRŠNI RAD

Sandro Lukač

Zagreb, 2015. godina.

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

ZAVRŠNI RAD

Mentori:

Prof. dr. sc. Nedeljko Štefanić, dipl. ing.
Dr. sc. Hrvoje Cajner, dipl. ing.

Student:

Sandro Lukač

Zagreb, 2015. godina.

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

Zahvaljujem se svome mentoru, **dr. sc. Hrvoju Cajneru** i **prof. dr. sc. Nedeljku Štefaniću** na stručnoj pomoći i savjetima kojima su mi omogućili da uspješno završim ovaj rad.

Zahvaljujem se obitelji na pruženom strpljenu i povjerenju koje su mi pružili tijekom studija i tijekom pisanja ovog rada.

Sandro Lukač

SVEUČILIŠTE U ZAGREBU

FAKULTET STROJARSTVA I BRODOGRADNJE

Središnje povjerenstvo za završne i diplomske ispite

Povjerenstvo za završne ispite studija strojarstva za smjerove:
proizvodno inženjerstvo, računalno inženjerstvo, industrijsko inženjerstvo i menadžment, inženjerstvo materijala i mehatronika i robotika

Sveučilište u Zagrebu Fakultet strojarstva i brodogradnje	
Datum 26 -02- 2015	Prilog
Klasa: 602-04/15- 6/3	
Ur.broj: 15- 1703- 15- 128	

ZAVRŠNI ZADATAK

Student: **SANDRO LUKAČ** Mat. br.: 0035186580

Naslov rada na hrvatskom jeziku: **Primjena statističkih metoda analize i optimizacije proizvodnog procesa uporabom dostupnog programskog paketa**

Naslov rada na engleskom jeziku: **Use of statistical methods for analyzing and optimization of production process using available computer software**

Opis zadatka:

Potreba za brzom analizom, kao i formiranje prijedloga optimizacije proizvodnih procesa, zahtijeva određeni stupanj teorijskog znanja, ali i poznavanja određenih računalnih i programskih rješenja. Budući da je dostupnost programskih rješenja nerijetko ograničavajući faktor, uporaba programskog paketa MS Excel, kao široko dostupnog alata, daje mogućnost efikasne statističke analize krajnjim korisnicima. Zadaća ovog rada jest dati presjek takvih mogućnosti navedenog programskog paketa kroz niz primjera iz prakse, te općenito pokazati kako pristupiti problemima analize i optimizacije proizvodnog sustava kroz ovakvo programsko rješenje.

Rad mora posebice sadržavati:

1. Presjek mogućnosti programskog paketa MS Excel po pitanju statističke analize i optimizacije.
2. Detaljan opis teorijskih podloga te programskih procedura statističkih metoda.
3. Provedenu statističku analizu na podacima iz stvarnih proizvodnih sustava.
4. Optimizaciju proizvodnog sustava korištenjem numeričke optimizacijske metode dostupne u navedenom programskom paketu.

Zadatak zadan:
25. studenog 2014.

Zadatak zadao:
Prof. dr. sc. Nedeljko Štefanić

Rok predaje rada:
1. rok: 26. veljače 2015.
2. rok: 17. rujna 2015.

Predviđeni datumi obrane:
1. rok: 2., 3. i 4. ožujka 2015.
2. rok: 21., 22. i 23. rujna 2015.

Predsjednik Povjerenstva:
Prof. dr. sc. Zoran Kunica

SADRŽAJ

SADRŽAJ I
POPIS SLIKA II
POPIS TABLICA IV
POPIS OZNAKA V
SAŽETAK VII
1. UVOD 1
2. TEORIJSKA PODLOGA 2
3. PRIMJENA RAČUNALA U STATISTIČKOJ ANALIZI 16
4. STATISTIČKA ANALIZA 41
5. OPTIMIZACIJA PROIZVODNOG SUSTAVA 46
6. ZAKLJUČAK 51
LITERATURA 52
PRILOG 53

POPIS SLIKA

Slika 1. Prikaz sučelja STATISTICA-e 17
Slika 2. Prikaz grafikona u programskom paketu Minitab 18
Slika 3. Površinski 2D grafikon 20
Slika 4. Histogram diskretnih vrijednosti 20
Slika 5. Histogram kontinuiranih vrijednosti 21
Slika 6. Paretov dijagram 21
Slika 7. „Box-whisker“ grafikon 22
Slika 8. Linijski grafikon 23
Slika 9. „Scatter plot“ 23
Slika 10. Funkcija AVEDEV 24
Slika 11. Funkcija CORREL 25
Slika 12. Funkcija CONFIDENCE.NORM 26
Slika 13. Funkcija STANDARDIZE 27
Slika 14. Funkcija NORM.DIST 28
Slika 15. Funkcija VAR.P 29
Slika 16. Funkcija STDEV.P 30
Slika 17. Funkcija AVERAGE 30
Slika 18. Funkcija DEVSQ 31
Slika 19. Funkcija HARMEAN 32
Slika 20. Funkcija MEDIAN 32
Slika 21. Funkcija MODE.SNGL 33
Slika 22. Funkcija QUARTILE.INC 34
Slika 23. Tablica rezultata alata deskriptivne statistike 37
Slika 24. T-test uparena dva uzorka 42
Slika 25. T-test dva uzorka jednakih varijanci 43
Slika 26. F-test dva uzorka 44
Slika 27. Z-test dva uzorka 45
Slika 28. Strukturni model 48
Slika 29. Parametri 48

Slika 30. Rezultati..... 49
Slika 31. Stanje ograničenja 50

POPIS TABLICA

Tablica 1. ANOVA s jednim promjenjivim faktorom 7
Tablica 2. ANOVA s dva promjenjiva faktora..... 9
Tablica 3. Odabir hipoteze kod t-testa..... 10
Tablica 4. Veličine t-testa uparena dva uzorka 11
Tablica 5. Odabir hipoteze za t-test dva uparena uzorka 12
Tablica 6. Veličine za t-test jednakih varijanci 12
Tablica 7. Odabir hipoteze za t-test jednakih varijanci 13
Tablica 8. Odabir hipoteze za z-test 14
Tablica 9. Veličine z-testa dva uzorka 14
Tablica 10. Z-test dva uzorka 15
Tablica 11. Setovi podataka 22
Tablica 12. Izlazne vrijednosti „quarta“ 34
Tablica 13. Podaci vrijednosti vlačnih čvrstoća [17] 41
Tablica 14. Dimenzije proizvoda dva stroja [18] 43
Tablica 15. Proizvodni podaci [20] 46

POPIS OZNAKA

Oznaka	Jedinica	Opis
$\bar{x}$		Aritmetička sredina
H		Harmonijska sredina
G		Geometrijska sredina
R		Raspon
α		Pogreška
M_o		Mod
L_1		Donja granica razreda
M_e		Medijan
σ^2		Varijanca
σ		Standardna devijacija
V		Koeficijent varijacije
Z		Standardizirano obilježje
n		Broj elemenata
m		Broj elemenata
μ		Očekivanje
M_r		Centralni moment
m_r		Pomoćni moment
S^2		Nepristrana procjena varijance
ε_{ij}		Slučajna odstupanja unutar uzorka
SKO		Suma kvadrata odstupanja
F		Računska vrijednost F-testa
k_b		Stupanj slobode brojnika
k_n		Stupanj slobode nazivnika
χ^2		Hi kvadrat varijabla

H_0	Nulta hipoteza
H_1	Alternativna hipoteza
P	Vrijednost vjerojatnosti
D	Pretpostavljeno očekivanje razlike među parovima
t	Računska vrijednost t-testa
z	Računska vrijednost z-testa
l	Veličina razreda
b	Frekvencija
a	Frekvencija razreda koji prethodi modalnom
c	Frekvencija razreda koji slijedi iza modalnog
f_{med}	Apsolutna frekvencija medijalnog razreda
$F(L_1)$	Kumulativna frekvencija
x_{max}	Najveća vrijednost niza podataka
x_{min}	Najmanja vrijednost niza podataka
x_{ij}	Vrijednost i-tog mjerenja u j-tom stupcu
$\Delta\bar{x_j}$	Doprinos aritmetičke sredine j-tog uzorka
$\Delta\bar{x_\iota}$	Doprinos aritmetičke sredine i-tog uzorka
s	Procjena standardne devijacije
$\bar{d}$	Očekivanje razlike među parovima
S_d	Procjena standardne devijacije uzorka razlike među parovima
μ_d	Stvarno očekivanje populacije razlika među parovima

SAŽETAK

Tema ovog rada je primjena statističkih metoda analize i optimizacije proizvodnog procesa uporabom dostupnog programskog paketa. Svaki proizvodni proces mora se kontrolirati, a kontrola podrazumijeva bilježenje značajnih podataka i vrijednosti, da bi te podatke znali analizirati i protumačiti. Uporabom računalne analize i optimizacije postupak se višestruko ubrzava i smanjuje se mogućnost pogreške.

U uvodu je ukratko opisana statistika te su navedeni programski paketi koji će se koristiti. U poglavlju računalne analize navedeni su neki programi koji se mogu koristiti kao zamjena za MS Excel. Također, opisane su neke od važnijih statističkih funkcija kao i programski paketi i alati koji dolaze uključeni u programski paket Excel, kao što su „Analysis ToolPak“ za statističku analizu i testove i „Solver“ koji se koristi za optimizaciju i linearno programiranje. Tema trećeg poglavlja je teorijska podloga u kojoj su navedene formule i postupci koji stoje iza računalnih funkcija i alata. Kao glavni primjer uzeti su F-test, t-test i z-test. U sljedećem poglavlju napravljeni su primjeri sa stvarnim proizvodnim podacima te tumačenje dobivenih rezultata. Na kraju rada napravljena je optimizacija proizvodnog sustava metodom linearnog programiranja alatom „Solver“.

Ključne riječi: Računalna analiza, statistička analiza, „Analysis ToolPak“, optimizacija, „Solver“, statistički testovi.

1. UVOD

Statistika je znanstvena metoda kojom se prikupljaju, uređuju, analiziraju i tumače raznovrsni brojčani podaci o različitim pojavama iz svakodnevnog života i procesi iz proizvodnih pogona. Zadaća statistike je uočiti zakonitosti u masovnim i slučajnim pojavama te donijeti zaključke koji bi te pojave i razne procese razjasnili te unaprijedili. Statistika kao znanstvena metoda proučavanja procesa i pojava dijeli se na deskriptivnu i inferencijalnu statistiku.

Deskriptivna statistika bavi se uređivanjem prikupljenih, empirijskih podataka te njihovim opisivanjem pomoću osnovnih statističkih veličina poput aritmetičke sredine, varijance i devijacije. Također, nastoji grafički prikazati podatke.

Inferencijalna statistika donosi zaključke uz pomoć metoda koje se provode na dijelu podataka, uzorku i prihvaća ih za cijelu populaciju.

U ovom radu dan je presjek mogućnosti širokodostupnog programskog paketa MS Excel u inženjerskoj statistici: prikaz često korištenih statističkih funkcija u analizi podataka, korištenje paketa „Analysis ToolPak“ s više alata za primjenu testova u teoriji uzoraka te korištenje alata „Solver“ za optimizaciju. Također, u radu je dana i detaljna teorijska podloga korištenih funkcija i alata.

2. TEORIJSKA PODLOGA

Poznavanje rada programa MS Excela ne znači nužno i poznavanje teorije iza svih funkcija. Svaka od funkcija unaprijed je isprogramirana i sadrži sva znanja iz područja statistike kako bi se korisniku olakšao rad. Neke od funkcija vrlo su jednostavne i lako shvatljive čak i onima koji nemaju znanja iz statističke analize dok su neke funkcije poprilično složene te zahtijevaju poduži račun na papiru. U ovom poglavlju razradit će se neke od statističkih funkcija i metoda često korištenih u analizi i optimizaciji. Također, te funkcije koristit će se i na primjeru gdje će se optimizirati proizvodni sustav.

2.1. Osnovni pojmovi u statistici

Da bi razumjeli većinu statističkih funkcija moramo razjasniti neke osnovne pojmove vezane za statističku analizu. Najčešće podaci dolaze u velikim setovima iz kojih je teško izvaditi korisne informacije. Zbog toga se takvi setovi podataka grupiraju te se iz njih računanjem dobivaju osnovne statističke veličine kao što su srednje vrijednosti oko koje se raspoređuju elementi seta ili niza.

Srednja vrijednost je centralna vrijednost obilježja nekog niza prema kojoj će se grupirati ostali članovi niza. Mogu se podijeliti na[1]:

- potpune srednje vrijednosti
- položajne srednje vrijednosti.

Potpune srednje vrijednosti računaju se korištenjem svih podataka u zadanom statističkom nizu. Tu spadaju aritmetička sredina, geometrijska sredina, harmonijska sredina i moment. Položajne srednje vrijednosti dobivaju se s obzirom na položaj podataka u nizu.

Aritmetička sredina je omjer zbroja svih članova niza i broja članova tog istog niza.

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n} \tag{2.1}$$

Geometrijska sredina je n-ti korijen umnoška svih članova skupa. Koristi se kao mjera prosječne brzine nekih promjena.

$$G = \sqrt[n]{x_1 \times \cdots \times x_n} \tag{2.2}$$

Harmonijska sredina je recipročna vrijednost aritmetičke sredine recipročnih vrijednosti za koje se sredina izračunava.

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{2.3}$$

Moment je aritmetička sredina niza odstupanja vrijednosti numeričke varijable od njezine aritmetičke sredine (centralni moment) ili neke druge vrijednosti (pomoćni moment) podignuta na neku potenciju $r \in N_0$ [2].

Centralni moment r-tog reda je moment oko centra (aritmetička sredina) [1]:

$$M_r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^r}{n} \tag{2.4}$$

Centralni moment r-te potencije 2 je zapravo varijanca dok potencije 3 i 4 predstavljaju koeficijent asimetrije odnosno spljoštenosti.

Pomoćni moment r-tog reda je moment oko točke 0 [1]:

$$m_r = \frac{\sum_{i=1}^{n} {x_i}^r}{n} \tag{2.5}$$

Pomoćni moment r-te potencije 1 predstavlja aritmetičku sredinu.

Mod s oznakom M_o, je obilježje statističke varijable koje ime najveću frekvenciju [2], tj. onaj podatak koji se najviše puta ponavlja. Ako su elementi raspoređeni u razrede, modalni je razred onaj s najvećom frekvencijom b. L_1 je donja granica toga razreda, l je veličina razreda, a je frekvencija razreda koji prethodi modalnom i c je frekvencija razreda koja slijedi iza modalnog.

$$M_o = L_1 + \frac{(b-a)}{(b-a)+(b-c)} \times l \tag{2.6}$$

Kvantili su numeričke vrijednosti koje niz uređen po veličini dijele na n jednakih dijelova.

Medijan je vrijednost središnjeg člana niza u nizu sa neparnim brojem elemenata. U slučaju da je broj elemenata u nizu paran, definira se kao aritmetička sredina središnjih dvaju članova, tj.

$$M_e = \frac{n_{r-1} + n_r}{2} \tag{2.7}$$

gdje je n_r vrijednost središnjeg člana niza. Medijan ima svojstvo da je zbroj apsolutnih odstupanja svih vrijednosti varijable od nekog fiksnog broja minimalan upravo za medijan. Također, ako su elementi grupirani u razrede onda je medijalni razred prvi razred po redu čija je kumulativna frekvencija veća ili jednaka $\frac{n}{2}$. Prema jednadžbi ispod, f_{med} je apsolutna frekvencija medijalnog razreda, l njegova veličina i $F(L_1)$ kumulativna frekvencija (zbroj svih frekvencija) do medijalnog razreda.

$$M_e = L_1 + \frac{\frac{n}{2} - F(L_1)}{f_{med}} \times l \tag{2.7}$$

Uz medijan, u kvantile spadaju kvartili koji dijele niz na 4 jednaka dijela, decili na 10 jednakih dijelova i percentili na 100 jednakih dijelova.

Uz navedene srednje vrijednosti često spominjani pojmovi su varijanca i standardna devijacija.

Varijanca je zbroj svih umnožaka između kvadrata razlike vrijednosti elementa x_i i aritmetičke sredine $\bar{x}$ podijeljen s brojem elemenata u nizu. To je minimum svih srednjih vrijednosti kvadratnih odstupanja od aritmetičke sredine. Uzima u obzir sve elemente numeričkog niza.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \tag{2.8}$$

Standardna devijacija je drugi korijen varijance i izražena je u originalnim jedinicama mjere. Zbog toga se vrlo često upotrebljava. Definira se kao prosječno odstupanje vrijednosti elementa od aritmetičke sredine. Pomoću standardne devijacije u originalnim mjernim jedinicama obilježja može se uspoređivati raspršenost oko aritmetičke sredine nizova koji su grupirani po jednakom obilježju [3].

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \tag{2.9}$$

Nepristrana procjena varijance je zbroj svih umnožaka između kvadrata razlike vrijednosti elementa x_i i aritmetičke sredine $\bar{x}$ podijeljen s korigiranim brojem uzoraka uzetih iz populacije kako prikazuje jednadžba (2.10):

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \tag{2.10}$$

Nepristrana procjena devijacije je drugi korijen nepristrane procjene varijance.

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{2.11}$$

Koeficijent varijacije prikazuje koliki odnos vrijednosti (u%) aritmetičke sredine iznosi vrijednost standardne devijacije. Naziva se još i relativna mjera rasipanja[1].

$$V = \frac{\sigma}{\bar{x}} \times [100\%] \tag{2.12}$$

Standardizirano obilježje oznake Z_i je linearna transformacija originalnih vrijednosti elementa x_i i prikazuje njihovo odstupanje od aritmetičke sredine u standardnim devijacijama.

$$Z_i = \frac{x_i - \bar{x}}{\sigma}, \qquad i \in N \tag{2.13}$$

Raspon je razlika najveće i najmanje vrijednosti u nekom nizu podataka.

$$R_x = x_{max} - x_{min} \tag{2.14}$$

2.2. Statističke metode i testovi

U ovom poglavlju bit će opisane statističke metode s pripadajućim testovima koji se pojavljuju u paketu „Analysis ToolPak“. Neke od metoda koje će se ukratko objasniti su analiza varijance (u nastavku ANOVA) s jednim i dva promjenjiva faktora, F-test, t- test, Z-test.

2.2.1. ANOVA s jednim promjenjivim faktorom

ANOVA je tehnika upotrebe razlika između prosjeka uzoraka u zaključivanju o postojanju ili ne postojanju razlika između prosjeka populacija. Postupak kojim je moguće raščlaniti i procijeniti varijabilnosti uvjetovane različitim čimbenicima tj., izvorima. U proizvodnim uvjetima analiza varijance je postupak koji se koristi u provjeri djelovanja promjene stanja nekog proizvodnog faktora na mjerenu vrijednost – rezultat [4]. Analizom varijance provjeravaju se promjene aritmetičkih sredina uzoraka. Primjena analize varijance moguća je ako su varijance svih promatranih uzoraka jednake i ako je mjerena varijabla normalno distribuirana.

Izgled modela prikazan je sljedećom jednadžbom:

$$x_{ij} = \mu + \Delta\bar{x}_j \pm \varepsilon_{ij} \tag{2.15}$$

gdje je x_{ij} vrijednost i-tog mjerenja u j-tom stupcu, μ aritmetička sredina svih podataka, $\Delta\bar{x}_j$ doprinos ar. sredine j-tog uzorka i ε_{ij} slučajno odstupanje unutar uzorka. Uzeta je u obzir pretpostavka da se slučajna odstupanja unutar uzorka ponašaju također po normalnoj razdiobi.

$$\varepsilon_{ij} \rightarrow N\left\{E(\varepsilon_{ij}) = 0; \sigma^2_{(\varepsilon_{ij})}\right\} \tag{2.16}$$

2.2.1.1. Dekompozicija Sume Kvadrata Odstupanja (SKO)

Ideja ANOVA-e je razdvajanje varijabilnosti na dva dijela: varijabilnost među uzorcima i varijabilnost unutar uzoraka (rezidualna varijabilnost).

Prema jednadžbi, ukupna varijabilnost prikazana je ukupnom sumom kvadrata odstupanja[5]:

$$\sum_{i,j}(x_{ij} - \mu)^2 = \sum_{i,j}(x_{ij} - \bar{x}_j)^2 + \sum_{i,j}(\bar{x}_j - \mu)^2 \tag{2.17}$$

Jednadžba (2.17) može se pisati:

$$SKO_{ukupno} = SKO_{unutar\ uzoraka} + SKO_{između\ uzoraka} \tag{2.18}$$
$$S^2_{ukupno} = S^2_{unutar\ uzoraka} + S^2_{između\ uzoraka}$$

Budući da postoji k razina uzoraka $SKO_{unutar\ uzoraka}$ ima k-1 stupnjeva slobode, te postoji N broja podataka, $SKO_{između\ uzoraka}$ ima n-k stupnjeva slobode. Omjer sume kvadrata odstupanja i broj stupnjeva slobode naziva se srednji kvadrat odstupanja između uzorka i označava se [5]:

$$S^2_{između\ uzoraka} = \frac{SKO_{između\ uzoraka}}{k-1} \tag{2.19}$$

te srednji kvadrat odstupanja unutar uzorka

$$S^2_{unutar\ uzorka} = \frac{SKO_{unutar\ uzoraka}}{N-k} \tag{2.20}$$

U modelu analize varijance s jednim promjenjivim faktorom pretpostavljaju se sljedeće hipoteze[4]:

$$H_0: \Delta x_1 = \Delta x_2 = \cdots = \Delta\bar{x}_k = 0 \tag{2.21}$$
$$H_1: bar\ jedan\ \Delta\bar{x}_j \neq 0$$

Za provjeru gornje hipoteze H_0 nužno je odrediti dvije nezavisne varijance, u ovom slučaju to će biti $S^2_{između\ uzoraka}$ i $S^2_{unutar\ uzoraka}$ kako bi se formirao F-test:

$$F = \frac{S^2_{između\ uzoraka}}{S^2_{unutar\ uzoraka}} \tag{2.22}$$

koji se ponaša po F razdiobi s $k_b = n_1 - 1$ i $k_n = n_2 - 1$ stupnjeva. Zaključak se donosi na temelju usporedbe izračunate veličine F s teorijskom vrijednošću F_0. Nulta hipoteza H_0 prihvaća se ukoliko je $F \leq F_0$, a odbacuje se ako je $F > F_0$ uz pogrešku α.

Tablica 1. ANOVA s jednim promjenjivim faktorom

Izvor varijacije	SKO	Stupnjevi slobode	srednji kvadrat odstupanja	F
Između uzoraka (faktor)	$SKO_{između}$	k-1	$S^2_{između}$	$\frac{S^2_{između}}{S^2_{ostatak}}$
Unutar uzoraka (ostatak)	$SKO_{ostatak}$	N-k	$S^2_{ostatak}$	-
Ukupno:	SKO_{ukupno}	N-1	-	-

2.2.2. F test i F razdioba

Ponekad je bitno znati da li su varijance različitih populacija jednake jer ta informacija pokazuje njihov međusobni odnos stupnja disperzije. Da bi se testirala činjenica o jednakosti varijanci pretpostavljaju se hipoteze. Pretpostavljamo da su na raspolaganju dva slučajna uzorka, veličine n_1 iz populacije 1 i veličine n_2 iz populacije 2, gdje su S_1^2 i S_2^2 nepristrane procjene varijance. Testirat će se hipoteze:

$$H_0: \sigma_1^2 = \sigma_2^2 \tag{2.23}$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Razvoj procedure za ovaj test i navedene hipoteze zahtjevaju novu razdiobu vjerojatnosti, F razdiobu. Slučajna varijabla F definirana je kao omjer dvije nezavisne hi-kvadrat (χ^2) varijable, svaka podijeljena sa svojim stupnjem slobode [2]. Neka su σ_1^2 i σ_2^2 varijance, a S_1^2 i S_2^2 nepristrane procjene varijance nezavisnih populacija 1 i 2.

Tada je omjer:

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \tag{2.24}$$

za bilo koja dva takva uzorka, distribuiran po F razdiobi s $n_1 - 1$ stupnjeva slobode brojnika i $n_2 - 1$ stupnjeva slobode nazivnika. Dakle testiranje hipoteza:

$$\frac{\sigma_1^2}{\sigma_2^2} = 1, \qquad \frac{\sigma_1^2}{\sigma_2^2} \geq 1, \qquad \frac{\sigma_1^2}{\sigma_2^2} \leq 1 \tag{2.25}$$

odnosno njihovih negacija, provodi se usporedbom vrijednosti dobivene F testom, $F = \frac{S_1^2}{S_2^2}$ i tabličnih vrijednosti F_α ili $F_{\alpha/2}$ gdje je α pogreška prve vrste.

2.2.3. ANOVA s dva promjenjiva faktora

U analizi varijance s jednim promjenjivim faktorom razmatrao se utjecaj jednog faktora na vrijednosti slučajne varijable. U praksi, međutim, nije rijetkost da na neku varijablu utječu istovremeno dva ili više faktora. Dakle, kada se želi provjeriti utjecaj djelovanja promjene stanja dva proizvodna faktora na mjerenu vrijednost, rezultat, koristi se analiza varijance s dva promjenjiva faktora.

Model analize varijance s dva promjenjiva faktora je sljedeći [4]:

$$x_{ij} = \mu + \Delta\bar{x}_j + \Delta\bar{x}_i \pm \varepsilon_{ij} \tag{2.26}$$

$$x_{ij} \rightarrow vrijednost\ i-tog\ retka\ i\ j-tog\ stupca$$

$$\mu \rightarrow aritmetička\ sredina\ svih\ podataka$$

$$\Delta\bar{x}_j \rightarrow doprinos\ ar.sredine\ j-tog\ stupca$$

$$\Delta\bar{x}_i \rightarrow doprinos\ ar.sredine\ i-tog\ stupca$$

$$\varepsilon_{ij} \rightarrow slučajno\ odstupanje\ (ostatak)$$

$$Pretpostavka:$$

$$\varepsilon_{ij} \rightarrow N\left\{E(\varepsilon_{ij}) = 0; \sigma^2_{(\varepsilon_{ij})}\right\}$$

U ovom situaciji formiraju se dva F testa[5]:

- Po redovima

$$H_0: \Delta\bar{x}_1 = \Delta\bar{x}_i = \cdots = \Delta\bar{x}_k = 0 \tag{2.27}$$

$$H_1: bar\ jedan\ \Delta\bar{x}_i \neq 0$$

- Po stupcima

$$H_0: \Delta\bar{x}_1 = \Delta\bar{x}_j = \cdots = \Delta\bar{x}_k = 0 \tag{2.28}$$
$$H_1: bar\ jedan\ \Delta\bar{x}_j \neq 0$$

U postupku analize varijance navedene hipoteze se testiraju pomoću dekompozicije sume kvadrata odstupanja. Ukupna varijabilnost podataka prikazana je ukupnom sumom kvadrata odstupanja:

$$\sum_{ij}(x_{ij}-\mu)^2 = \sum_{i,j}(x_j-\mu)^2 + \sum_{i,j}(\bar{x}_i-\mu)^2 + SKO_{ostatka} \tag{2.29}$$

odnosno:

$$SKO_{ukupno} = SKO_{stupci} + SKO_{redovi} + SKO_{ostatka} \tag{2.30}$$
$$S^2_{ukupno} = S^2_{stupci} + S^2_{redovi} + S^2_{ostatka}$$

Tablica 2. ANOVA s dva promjenjiva faktora

Izvor varijacije	Suma kvadrata odstupanja	Stupnjevi slobode	Srednji kvadrat odstupanja	F
Faktor 1 (redovi)	SKO_{redovi}	l-1	S^2_{redovi}	$S^2_{redovi}/S^2_{ostatak}$
Faktor 2 (stupci)	SKO_{stupci}	k-1	S^2_{stupci}	$S^2_{stupci}/S^2_{ostatak}$
Ostatak	$SKO_{ostatak}$	(k-1)(l-1)	$S^2_{ostatak}$	-
Ukupno	SKO_{ukupno}	N-1	-	-

Zaključak se donosi na temelju usporedbe izračunate veličine F s teorijskom vrijednošću F_0. Nulte hipoteze se prihvaćaju ukoliko je $F \leq F_0$, a odbacuje se ako je $F > F_0$ uz pogrešku α.

2.2.4. Z-test i t-test

Testiranje hipoteze o pretpostavljenoj vrijednosti očekivanja osnovnog skupa se provodi tako da se iz osnovnog skupa izuzme slučajan uzorak veličine n članova. Ako imamo veliki uzorak (n > 30), riječ je o testu velikim uzorkom (z-test), gdje je vrijednost varijable z iz standardizirane normalne razdiobe, a ako je u uzorku trideset i manje članova (n ≤ 30), tada se radi o malenom uzorku, koristi se studentova t – razdioba i provodi se t-test.

2.2.4.1. T-test

Ako postoji osnovni skup koji ima normalnu razdiobu, ali se ne zna njegova varijanca, testiranje hipoteze o pretpostavljenoj vrijednosti očekivanja osnovnog skupa temeljit će se na Studentovoj razdiobi[6]. Pretpostavljaju se sljedeće hipoteze:

$$H_0: \mu = \mu_0 \tag{2.31}$$
$$H_1: \mu \neq \mu_0 \; ili \; \mu > \mu_0 \; ili \;\; \mu < \mu_0$$

Izračuna se test veličina:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \tag{2.32}$$

gdje je μ_0 pretpostavljena vrijednost očekivanja osnovnog skupa, $\bar{x}$ aritmetička sredina uzorka i $\frac{s}{\sqrt{n}}$ standardna pogreška aritmetičke sredine[7]. Stupanj slobode je n-1.

Sljedeći korak je pronalazak P vrijednosti. P vrijednost ovisi o tome koja se alternativna hipoteza promatra. P vrijednost je vjerojatnost tj. površina ispod repova t distribucije s n -1 stupnjeva slobode. Tri su moguće varijante[6]:

- ako je $H_1: \mu > \mu_0$, P vrijednost = P (T ≥ t)
- ako je $H_1: \mu < \mu_0$, P vrijednost = P (T ≤ t)
- ako je $H_1: \mu \neq \mu_0$, P vrijednost = P (T ≤ − |t| ili T ≥ |t|) ili 2P(T ≥ |t|)

Kada je P vrijednost poznata uspoređuje se s pogreškom α, te ako je:

Tablica 3. Odabir hipoteze kod t-testa

P vrijednost < α	odbacuje se H_0
P vrijednost > α	prihvaća se H_0

2.2.4.2. T-test uparena dva uzorka za pretpostavljenu vrijednost očekivanja

Ovaj test se koristi kada imamo dva uzorka iste veličine, uzorci su zavisni (2 opažanja na istim subjektima, prije i poslije pokusa) i oba uzorka su dovoljno velika da su njihova očekivanja normalno distribuirana.

Tablica 4. Veličine t-testa uparena dva uzorka

x_i	podaci iz uzorka 1
y_i	podaci iz uzorka 2
$d_i = x_i - y_i$	razlika među parovima
$\bar{d}$	očekivanje razlike među parovima
S_d	procjena standardne devijacije uzorka razlike među parovima
n	veličina uzorka
μ_d	stvarno očekivanje populacije razlika među parovima
D	pretpostavljeno očekivanje razlike među parovima

gdje je [6]:

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} \tag{2.33}$$

$$S_d = \frac{\sqrt{\sum_{i=1}^{n}(d_i - \bar{d})^2}}{n-1} \tag{2.34}$$

Prvo će se pretpostaviti hipoteze:

$$H_0: \mu_d = D \tag{2.35}$$

$$H_1: \mu_1 - \mu_2 \neq D \; ili \; \mu_1 - \mu_2 > D \; ili \; \mu_1 - \mu_2 < D$$

Zatim se odabire razina povjerenja uz pogrešku α, te se izračuna testna veličina:

$$t = \frac{\bar{d} - D}{\frac{S_d}{\sqrt{n}}} \tag{2.36}$$

Odredi se stupanj slobode koji iznosi n-1 i pronađe se P vrijednost koja ponovno ovisi o tome koja se hipoteza koristi[8]:

- ako je H_1: $\mu_d > D$, P vrijednost = P(T ≥ t)
- ako je H_1: $\mu_d < D$, P vrijednost = P(T ≤ t)

- ako je H_1: $\mu_d \neq D$, P vrijednost = P(T ≤ − |t| or T ≥ |t|) or 2P(T ≥ |t|)

Kada je P vrijednost poznata, usporedi se s pogreškom α te ako je:

Tablica 5. Odabir hipoteze za t-test dva uparena uzorka

P vrijednost < α	odbacuje se H_0
P vrijednost > α	prihvaća se H_0

2.2.4.3. T-test dva nezavisna uzorka jednakih varijanci za pretpostavljenu vrijednost očekivanja

Ovaj test se koristi kada postoje dva uzorka, koji ne moraju biti iste veličine, iz dvije populacije. Uzorci su nezavisni, veličine uzorka su dovoljno velike da je očekivanje normalno distribuirano ili su obje populacije normalno distribuirane[8]. Standardne devijacije obje populacije su nepoznate ali se pretpostavljaju jednake.

Tablica 6. Veličine za t-test jednakih varijanci

Populacija	Uzorak	Očekivanje	Standardna devijacija	Veličina uzorka	Ar. sredina uzorka	Procjena stan. dev.
1	x_i	μ_x	σ_x	n	$\bar{x}$	S_x
2	y_i	μ_y	σ_y	m	$\bar{y}$	S_y

Prvo će se pretpostaviti hipoteze:

$$H_0: \mu_x - \mu_y = D \tag{2.37}$$

$$H_1: \mu_x - \mu_y \neq D \; ili \; \mu_x - \mu_y > D \; ili \; \mu_x - \mu_y < D$$

gdje je *D* pretpostavljena razlika u očekivanju.

Zatim se računa testna veličina[6]:

$$t = \frac{(\bar{x} - \bar{y}) - D}{\sqrt{S_d^2 \left(\frac{1}{n} + \frac{1}{m}\right)}} \tag{2.38}$$

gdje je:

$$S_d^2 = \frac{(n-1) \times S_x^2 + (m-1) \times S_y^2}{n + m - 2} \tag{2.39}$$

Odredi se stupanj slobode:

$$SS = n + m - 2 \tag{2.40}$$

Zatim se pronađe P vrijednost koja ponovno ovisi o hipotezi koja se koristi:

- ako je H_1: $\mu_x - \mu_y > D$, P vrijednost = P(T ≥ t)
- ako je H_1: $\mu_x - \mu_y < D$, P vrijednost = P(T ≤ t)
- ako je H_1: $\mu_x - \mu_y \neq D$, P vrijednost = P(T ≤ − |t| or T ≥ |t|) ili 2P(T ≥ |t|)

Sada poznatu P vrijednost, uspoređuje se s pogreškom α, te ako je [10]:

Tablica 7. Odabir hipoteze za t-test jednakih varijanci

P vrijednost < α	odbacuje se H_0
P vrijednost > α	prihvaća se H_0

2.2.4.4. T-test dva uzorka nejednakih varijanci za pretpostavljenu vrijednost očekivanja

Ovaj test jednak je prethodnom testu uz razliku što se ovaj provodi uz pretpostavku nejednakih varijanci[8]. Uz iste hipoteze, odabire se razina povjerenja te se izračuna testna veličina:

$$t = \frac{(\bar{x} - \bar{y}) - D}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \tag{2.41}$$

Zatim se nađe stupanj slobode:

$$SS = \frac{\left(\frac{S_x^2}{n} - \frac{S_y^2}{m}\right)^2}{\frac{S_x^4}{n^2(n-1)} + \frac{S_y^4}{m^2(m-1)}} \tag{2.42}$$

Ovaj podatak rijetko će biti cijeli broj pa se može zaokružiti na najbližu cjelobrojnu vrijednost. Daljnji postupak identičan je postupku iz prethodnog T-testa.

2.2.4.5. Z-test

Kao što se već spomenulo z-test koristi se kada je populacija normalno distribuirana i kada postoji uzorak od 30 ili više elemenata. Standardna devijacija populacije je poznata. Pretpostavljaju se hipoteze:

$$H_0: \mu = \mu_0 \tag{2.43}$$

$$H_1: \mu \neq \mu_0 \; ili \; \mu > \mu_0 \; ili \;\; \mu < \mu_0$$

Odabire se razina povjerenja i izračuna se testna veličina[7]:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \tag{2.44}$$

Zatim se pronađe P vrijednost ovisno o hipotezi:

- ako je $H_1: \mu > \mu_0$, P vrijednost = P ($Z \geq z$)
- ako je $H_1: \mu < \mu_0$, P vrijednost = P ($Z \leq z$)
- ako je $H_1: \mu \neq \mu_0$, P vrijednost = P ($Z \leq -|z|$ ili $Z \geq |z|$) ili $2P(Z \geq |z|)$

Uz poznatu vrijednost P donosi se zaključak:

Tablica 8. Odabir hipoteze za z-test

P vrijednost < α	odbacuje se H_0
P vrijednost > α	prihvaća se H_0

2.2.4.6. Z-test s dva uzorka za srednje vrijednosti s poznatim varijancama

Ovaj test koristi se kada imamo dva uzorka iz dvije populacije koji mogu biti različitih veličina. Obje populacije su normalno distribuirane ili su oba uzorka, koji su nezavisni, dovoljno veliki da su očekivanja normalno distribuirana ($n \geq 30$)[8]. Standardne devijacije obje populacije su poznate.

Tablica 9. Veličine z-testa dva uzorka

Populacija	Uzorak	Očekivanje	Standardna devijacija	Veličina uzorka	Ar. sredina uzorka
1	x_i	μ_x	σ_x	n	$\bar{x}$
2	y_i	μ_y	σ_y	m	$\bar{y}$

Prvo će se pretpostaviti hipoteze:

$$H_0: \mu_x - \mu_y = D \tag{2.45}$$

$$H_1: \mu_x - \mu_y \neq D \; ili \; \mu_x - \mu_y > D \; ili \; \mu_x - \mu_y < D$$

Odabire se razina povjerenja, te se izračuna testna veličina:

$$z = \frac{(\bar{x} - \bar{y}) - D}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \tag{2.46}$$

Zatim se pronađe P vrijednost ovisno o hipotezi:

- ako je $H_1: \mu_x - \mu_y > D$, P vrijednost = $P(Z \geq z)$

- ako je H_1: $\mu_x - \mu_y < D$, P vrijednost = P(Z ≤ z)
- ako je H_1: $\mu_x - \mu_y \neq D$, P vrijednost = P(Z ≤ − |z| or Z ≥ |z|) ili 2P(Z ≥ |z|)

Kada je P vrijednost poznata, uspoređuje se s pogreškom α, te se donosi zaključak:

Tablica 10. Z-test dva uzorka

P vrijednost < α	odbacuje se H_0
P vrijednost > α	prihvaća se H_0

3. PRIMJENA RAČUNALA U STATISTIČKOJ ANALIZI

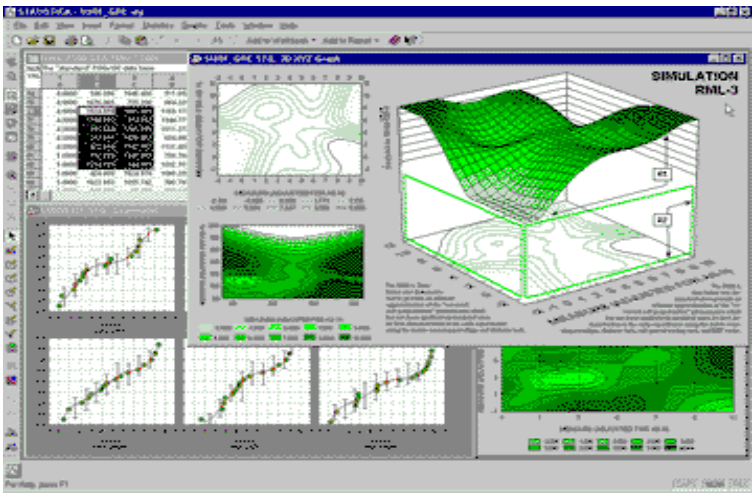
U ovom poglavlju prikazat će se mogućnosti računalne statističke analize podataka te tumačenje dobivenih rezultata.

3.1. Primjena računala i programski paketi koji se koriste za statističku obradu

U današnjem svijetu potreba za što bržom obradom podataka i analizom dovela je do velikog razvoja računalnih paketa koji te procese ne samo olakšavaju već i unaprjeđuju. Programiranje ali i korištenje tih paketa zahtijeva određeni stupanj teorijskog znanja ali je postupak cjelokupnog dobivanja korisnih podataka olakšan. Uz pomoć takvih paketa ne rijetko se može provesti i optimizacija procesa koja zbog svoje zamršenosti ne bi bila izvediva „ručnim postupkom“ u realnom vremenu. Neki od postupaka koje takvi paketi sadrže uz samu analizu su upravljanje podacima, rudarenje podataka, razni statistički testovi (F-test, t-test...) i vizualni prikaz podataka. Na žalost, mnogi od tih računalnih paketa dostupni su samo uz plaćanje, stoga će se u ovom radu samo predstaviti dva paketa dok će se rad bazirati na poznatom i dostupnom paketu Microsoft Excel.

3.1.1. Statistica

Statistica je analitički i statistički računalni paket razvijen od StatSoft kompanije koja je podružnica megakompanije za računalnu opremu i programe Dell. Sadrži cijeli niz mogućnosti za obradu, upravljanje, prikaz i rudarenje podataka kao i raznolikost funkcija prediktivnog modeliranja za predviđanje ishoda, klasifikacijske i Cluster metode te metode analiziranja setova podataka u svrhu pronalaska njihove glavne karakteristike[9]. Što se tiče grafičkih mogućnosti, dvodimenzionalni i trodimenzionalni grafovi dobivenih podataka lako se generiraju kako prikazuje slika [Slika 1] dolje:



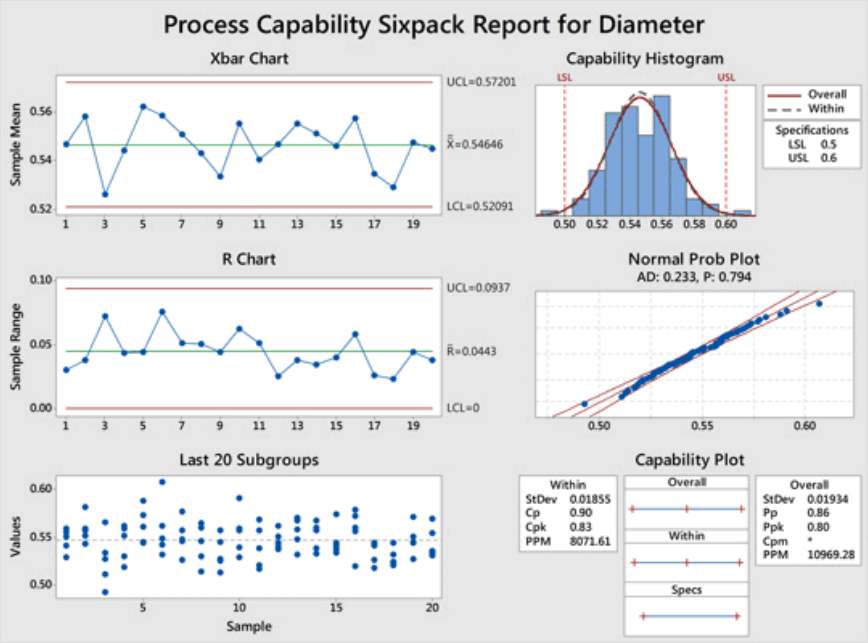
Slika 1. Prikaz sučelja STATISTICA-e

Također, označavanje i obilježavanje te uspoređivanje s glavnom karakteristikom seta podataka omogućuje lakši uvid ekstrema te njihovo primjereno izuzimanje iz seta.

3.1.2. Minitab

Sljedeći od programskih paketa koje ćemo spomenuti kao uvod u MS Excel je Minitab. Minitab je programski paket sa statističkim metodama razvijen na Državnom Sveučilištu Pensilvanije. Danas se koristi u svezi s implementacijom „Six sigma-e“, „CMMI“ (eng. Capability Maturity Model Integration, sposobnost i zrelost integracije modela) i drugih, statistikom baziranih modela poboljšanja procesa. Program je vrlo razvijen i raširen jer ima integrirane metode koje se koriste u Američkom Ministarstvu Obrane. Samo neke od mogućnosti su ANOVA, Monte Carlo metoda, MANOVA (višestruka analiza varijanci) i razni statistički testovi (z-test, t-test, korelacijski testovi)[10]. Također, trodimenzionalni grafovi, dijagrami i interaktivni grafikoni lako su dostupni za prikaz. Za kontrolu i optimizaciju procesa koriste se Pareto dijagrami, dijagrami uzroka i posljedice, $\bar{x}$ (aritmetička sredina), σ (standardna devijacija) i R (raspon) dijagrami.

Prednosti Minitaba su jednostavno intuitivno sučelje, grafičke ikone za lakše snalaženje, automatsko osvježavanje tablica u slučaju promjene podataka ili vrijednosti, jednostavna izrada tablica u podlošku sličnom MS Excelu, pregled aktivnosti u prozoru događaja i jednostavan pregled dobivenih grafova kako je prikazano na slici dolje [Slika 2].



Slika 2. Prikaz grafikona u programskom paketu Minitab

Iako su predstavljeni programski paketi zanimljivi i izrazito korisni u statističkoj analizi nisu besplatni i otvoreni za korištenje bez kupnje licence stoga će se ovaj rad temeljiti na programskom paketu MS Excel koji je pristupačnija varijanta.

3.2. MS Excel

Kako se i najavilo u prethodnom potpoglavlju ovaj rad temeljiti će se na programskom paketu MS Excel. Excel je jedan od najčešće korištenih računalnih programa za široku upotrebu i primjenu statističkih izračuna i analize. Gotovo svako računalo dolazi s instaliranim paketom Microsoft Office, stoga je često ako ne i uvijek prvi odabir upravo Excel. Moguća je jednostavna razmjena podataka i to na razini više različitih programa (grafičkih, tabličnih ili baza podataka)[11]. Što se tiče grafičkih prezentacija na raspolaganju se nalaze više tipova i podtipova grafikona. Vezani su podacima u tablici i ažuriraju se trenutno kako se promijene podaci u tablici. Moguće je vezati i preklapati više grafikona kako bi imali trenutni uvid u razliku između setova podataka i najvećih odstupanja. Trodimenzionalni grafovi mogu se orijentirati u prostoru u sve tri osi.

Još neke od dodatnih pogodnosti koje čine Excel tako raširenim su:

- jednostavne osnovne naredbe
- integriranost u Microsoft Office
- nije potrebno učiti nove metode manipuliranja podataka niti crtanja grafova jer se oni rade automatski

- već sadrži dodatke sa statističkim funkcijama i analizom podataka
- jednostavnost promjene podataka i parametara
- mogućnost izrade vlastitih funkcija.

To sve čini ga izvrsnim programom za učenje i primjenu statističkih alata što će biti pokazano u ovom radu. Program sadrži mnoštvo matematičkih, logičkih, financijskih i drugih funkcija ali će se u radu dati presjek statističkih funkcija. Također detaljno će se opisati mogućnosti paketa „Analysis ToolPak“ i alata „Solver“.

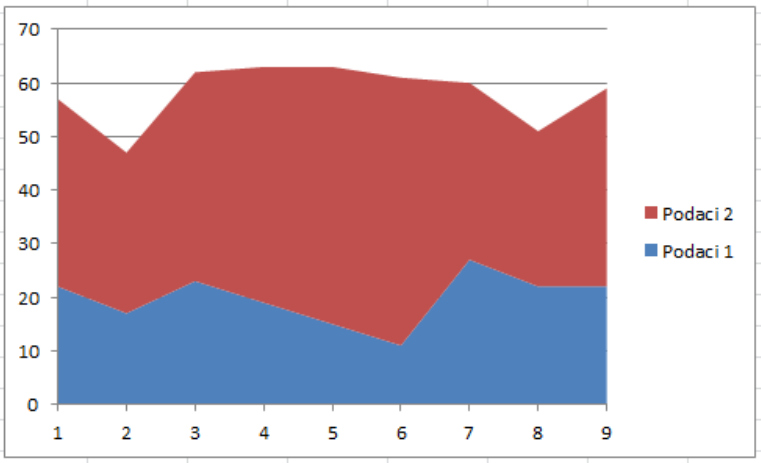
3.2.1. Izrada grafikona u Excelu

U ovom potpoglavlju pokazat će se mogućnosti i načini izrade grafikona vezanih uz statističku obradu podataka. Izrada grafikona vrlo je jednostavna te se jednostavnim selektiranjem niza podataka i odabirom vrste grafikona pojavljuje gotov grafički prikaz podataka. Grafikonima se na razumljiv i pregledan način, uz pomoć jednostavnih geometrijskih likova i ponešto brojčanih podataka prikazuju rezultati statističkih obrada i analiza. Nažalost, što je grafikon pregledniji, podaci u njemu su šturiji i manje precizni stoga treba biti oprezan pri njihovoj izradi.
Grafikon mora imati naslov, ishodišnu točku ili središte (i osi, ukoliko je njima određen), jedinice promatranih podataka (ako su zadane) te po potrebi njihov izvor i kazalo. On se zapravo sastoji od nekoliko komponenti, svaka međusobno podložna promjeni. U statističkoj obradi razlikuju se dvije namjene grafikona[12]:

- istraživački grafikoni
- prezentacijski grafikoni.

Površinski grafikoni bilo da su u dvodimenzionalnom [Slika 3] ili trodimenzionalnom obliku prikazuju trend vrijednosti tijekom vremena ili druge podatke iz nizova podataka. Na 3D površinskim grafikonima koriste se tri osi (vodoravna, okomita i os dubine) koje se mogu promijeniti.

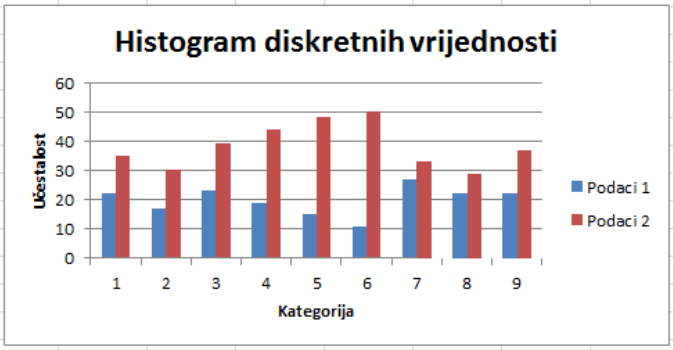
	Podaci1	Podaci2
2	22	35
3	17	30
4	23	39
5	19	44
6	15	48
7	11	50
8	27	33
9	22	29
10	22	37



Slika 3. Površinski 2D grafikon

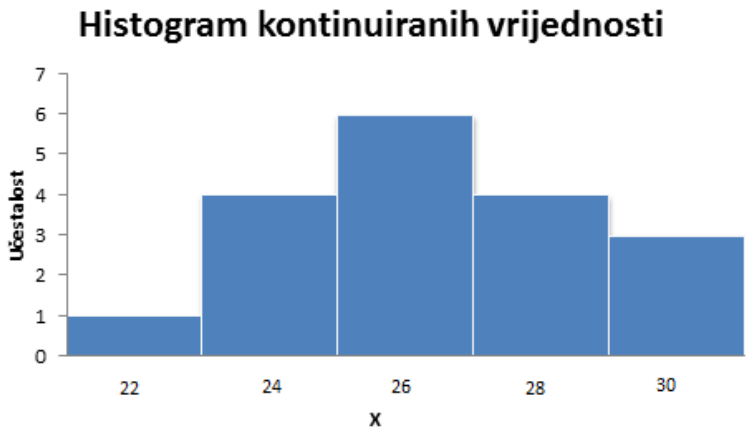
Prikazivanje podataka, koji su na radnom listu raspoređeni u stupce ili retke, i njihove učestalosti moguće je prikazati na histogramu. Na histogramu se kategorije najčešće prikazuju na vodoravnoj osi (osi kategorije), a vrijednosti na okomitoj osi (osi vrijednosti), kao što je prikazano na grafikonu ispod [Slika 4] za diskretne varijable:

	Podaci1	Podaci2
2	22	35
3	17	30
4	23	39
5	19	44
6	15	48
7	11	50
8	27	33
9	22	29
10	22	37



Slika 4. Histogram diskretnih vrijednosti

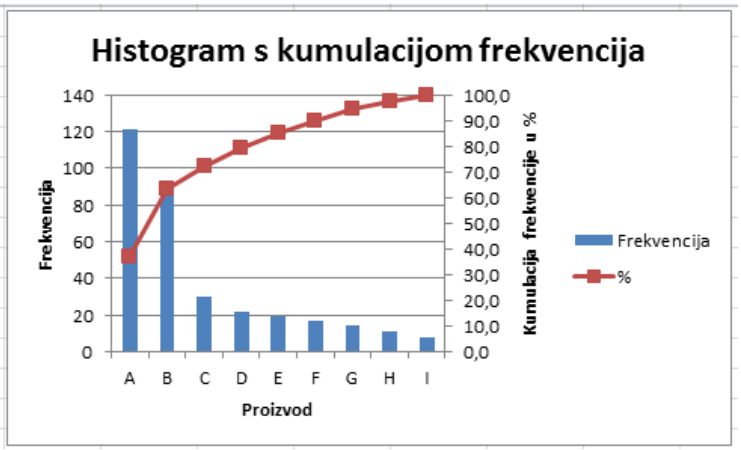
Iz ovakvog grafikona direktno se očitava vjerojatnost pojave pojedine vrijednosti varijable. Prikaz preko razreda podataka po kojima klasificiramo podatke je histogram za kontinuiranu varijablu, kako je prikazano na grafikonu ispod [Slika 5]. U tehnici se prikazuje s razredima (stupcima) jednake širine[1].



Slika 5. Histogram kontinuiranih vrijednosti

Također možemo prikazati histogram s frekvencijama koje se kumuliraju. Ovaj grafikon naziva se još i Paretov dijagram [Slika 6].

	Proizvod	Frekvencija	Kumulativ	%
2	A	121	121	36,7
3	B	88	209	63,3
4	C	30	239	72,4
5	D	22	261	79,1
6	E	19	280	84,8
7	F	17	297	90,0
8	G	14	311	94,2
9	H	11	322	97,6
10	I	8	330	100,0



Slika 6. Paretov dijagram

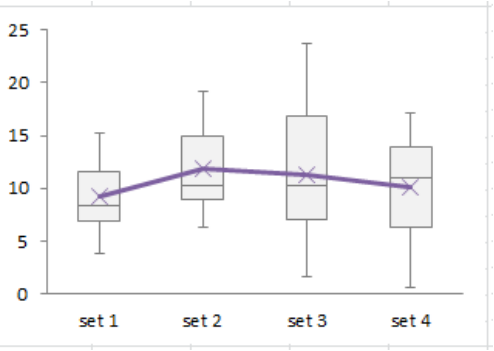
„Box-whisker“ grafikon [Slika 7] (eng. „kutija-brkovi“) jedno je od najčešćih prikaza podataka. Grafikon je moguće kreirati u različitim verzijama gdje je središnja točka medijan ili aritmetička sredina ili s raznim podjelama (npr. po intervalima promjene)[1]. Ovakvim prikazom jednostavno se uočavaju problematični podaci kao što su minimumi, maksimumi ili ekstremi u nekom drugom pogledu. Jednostavno se prikazuju dva ili više uzorka paralelno te se potom lako uviđa njihova karakteristika i međusobna relacija. Od navedenih podataka u tablici ispod [Tablica 11] napravljen je „box-whisker“ grafikon:

Tablica 11. Setovi podataka

set 1	set 2	set 3	set 4
14,67	17,40	10,14	6,42
10,06	17,15	1,62	17,18
8,26	19,17	20,25	10,91
6,95	9,83	5,92	6,80
5,05	8,26	9,68	6,22
11,47	8,18	17,46	5,53
11,95	12,07	7,43	4,88
3,88	17,28	17,78	4,86
8,31	9,05	10,22	14,21
6,54	6,36	23,72	6,62
4,48	7,40	2,65	15,73
8,79	14,20	8,29	12,70
10,97	9,86	3,35	14,94
8,15	18,08	17,99	13,43
7,13	11,75	10,21	8,63
15,14	10,29	5,45	13,92
7,67	9,79	10,94	0,57
15,24	10,36	13,73	12,69
14,08	8,86	12,42	15,36
6,07	12,06	16,71	11,00

Za izradu „box-whisker“ grafikona moramo se poslužiti funkcijama MIN, MAX, QUARTILE, MEDIAN i AVERAGE pomoću kojih se grafikon definira i koje će se kasnije u radu objasniti.

	set 1	set 2	set 3	set 4
Sr.vrijed.	9,24	11,87	11,30	10,13
St.dev.	3,56	3,95	6,16	4,65
Min	3,88	6,36	1,62	0,57
Q1	6,85	9,00	7,05	6,37
Median	8,29	10,33	10,22	10,96
Q3	11,59	14,94	16,90	13,99
Max	15,24	19,17	23,72	17,18



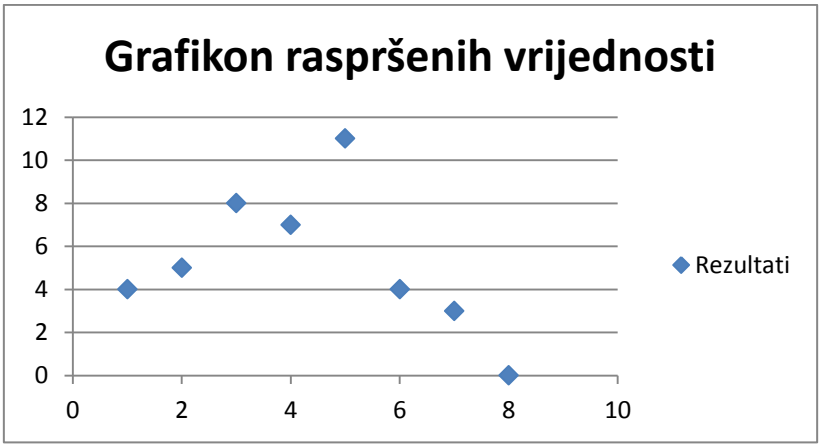
Slika 7. „Box-whisker“ grafikon

Na slici iznad srednja vrijednost tj. aritmetička sredina prikazana je znakom x u sredini „kutije“, minimum predstavlja najdonju točku dok maksimum najgornju točku „brka“, prvi kvartil (Q1) odnosno treći kvartil (Q3) su donja odnosno gornja stranica „kutije“ i medijan je horizontalna crta unutar „kutije“.

Linijski grafikoni [Slika 8], koji se prikazuju s oznakama za prikaz pojedinačnih vrijednosti podataka ili bez njih, mogu se koristiti za prikaz trendova tijekom vremena ili po ravnomjerno raspoređenim kategorijama, osobito kada postoje brojne točke podataka, a važan je i redoslijed kojim su predstavljene.

Slika 8. Linijski grafikon

Također postoje i raspršeni grafikoni (eng. Scatter plot) [Slika 9]. Vrijednosti za x unesu se u jedan stupac, a vrijednosti za y u drugi. Na raspršenom dijagramu vrijednosti x i y se objedinjuju u jedinstvene točke tih vrijednosti te se prikazuju u stvarnim često nepravilnim intervalima.



Slika 9. „Scatter plot“

3.2.2. Statističke funkcije MS Excela

Postoji mnogo statističkih funkcija u Excelu kojima se još mogu dodati i one funkcije koje nisu specificirane kao statističke ali mogu uvelike pomoći pri statističkoj obradi podataka. Statističke funkcije se pozivaju iz kartice „Formulas“ te se pod padajućim izbornikom odabiru statističke funkcije. Njih je otprilike 100, ovisno o verziji MS Excela te ćemo prema važnosti i preglednosti detaljno opisati samo neke [13]. Većina statističkih funkcija svakako se može izvesti i „klasičnim“ matematičkim funkcijama, ukoliko korisnik poznaje obrasce za

izračunavanje parametara. Efikasnost je važnija te će se tako analiza puno brže provesti uz pomoć gotovih funkcija.

Za većinu funkcija vrijedi sljedeće:

- podaci trebaju biti brojevi ili nazivi polja, odnosno reference koje sadrže brojeve
- ako podatak koji je polje ili referenca sadrži tekst, logičke vrijednosti ili prazne ćelije, te su vrijednosti zanemarene. Međutim, ćelije s vrijednošću nula su uključene.

AVEDEV

Vraća prosjek apsolutnih odstupanja točaka podataka od svojih srednjih vrijednosti tj., aritmetičkih sredina. Ova funkcija je mjera stupnja varijabilnosti u skupu podataka.

Sintaksa:

„Number1; number2;...“ su podaci za koje se izračunava prosjek apsolutnih odstupanja. Na njih utječe mjerna jedinica ulaznih podataka.

Napomene:

- jednadžba prosječnog odstupanja je:

$$f(x) = \frac{1}{n}\sum|x - \bar{x}| \tag{3.1}$$

Na slici ispod dan je primjer navedene funkcije:

Podaci				
1				
2	Formula			
3	AVEDEV(A2:A11)			
4				
5	Rezultat			
6	2,5			
7	Prosjek apsolutnih odstupanja podataka			
8	od njihove srednje vrijednosti			
9				
10				

Slika 10. Funkcija AVEDEV

CORREL

Ova funkcija prikazuje koeficijent korelacije dvaju niza podataka. Koristi se za određivanje odnosa između dva svojstva.

Sintaksa:

„Array1“ je prvi raspon ćelija s vrijednostima.

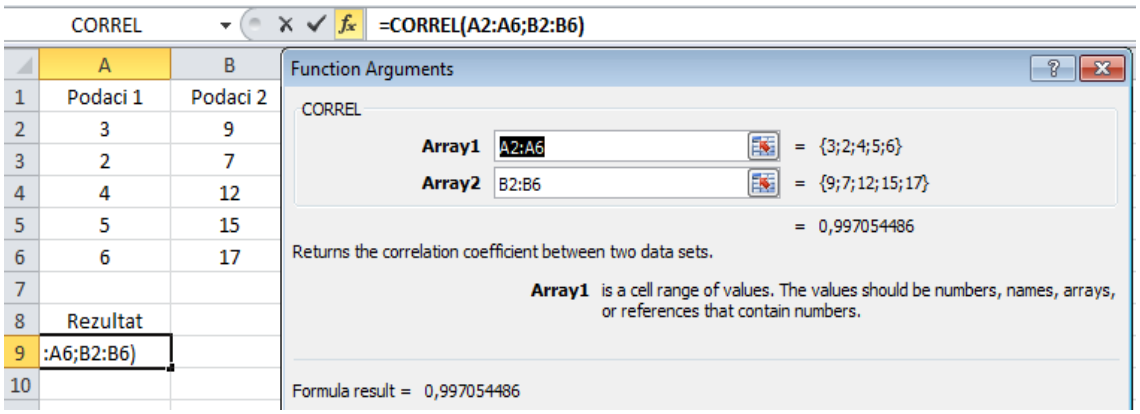
„Array2“ je drugi raspon ćelija s vrijednostima.

Napomene:

- ako prvi i drugi raspon ćelija imaju različit broj točaka, CORREL prikazuje vrijednost pogreške
- jednadžba:

$$Correl(X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \tag{3.2}$$

Primjer funkcije CORREL dan je na slici ispod:



Slika 11. Funkcija CORREL

CONFIDENCE.NORM

Funkcija se koristi da bi se izradio interval povjerenja aritmetičke sredine populacije. Interval povjerenja raspon je vrijednosti. Aritmetička sredina uzorka, $\bar{x}$, nalazi se u središtu tog raspona koji iznosi $\bar{x}$ ± CONFIDENCE.NORM. Na primjer, ako je $\bar{x}$ aritmetička sredina uzorka vremena isporuke za proizvode naručene poštom, $\bar{x}$ ± CONFIDENCE.NORM raspon je očekivanja populacije. Dakle, za populaciju μ_0, vjerojatnost dobivanja očekivanja populacije μ_0 u zadanom intervalu određena je razinom povjerenja, dok je za dobivanje vrijednosti izvan zadanog intervala vjerojatnost manja od pogreške α. Interval povjerenja ne dozvoljava zaključak da postoji vjerojatnost 1 – α da će se vrijeme isporuke sljedeće pošiljke nalaziti u intervalu povjerenja.

Sintaksa:

„Alpha“ je pogreška koja se koristi za izračun razine povjerenja. Razina povjerenja jednaka je 100*(1 – α)% ili drugim riječima, α s vrijednošću od 0,05 označava razinu povjerenja od 95%.

„Standard_dev“ je standardna devijacija populacije za raspon podataka i pretpostavlja se da je poznata.

„Size“ je veličina uzorka.

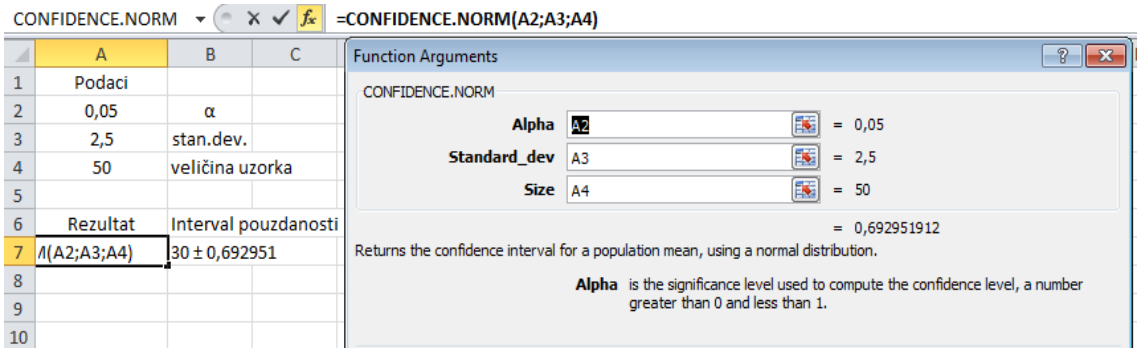
Napomene:

- Ako polje „size“ nije cijeli broj, decimalna mjesta se odbacuju
- Pretpostavi li se da je α jednak 0,05, potrebno je izračunati površinu ispod standardne normalne krivulje koja je jednaka 1 – α, tj. 95%. Ova vrijednost je ± 1,96. Stoga je interval povjerenja:

$$\bar{x} = \pm 1{,}96\left(\frac{\sigma}{\sqrt{n}}\right) \tag{3.3}$$

Primjer:

Pretpostavi li se da je u uzorku od 50 putnika prosječno trajanje putovanja do posla 30 minuta sa standardnom devijacijom populacije od 2,5. Ako je α = 0,05, CONFIDENCE.NORM vraća 0,69291. Odgovarajući interval pouzdanosti tada je 30 ± 0,69291 tj., približno od 29,3 do 30,7 kako prikazuje slika dolje [Slika 12]:



Slika 12. Funkcija CONFIDENCE.NORM

STANDARDIZE

Funkcija vraća normaliziranu vrijednost razdiobe određene s očekivanjem i standardnom devijacijom.

Sintaksa:

„X“ je vrijednost koja se želi normalizirati.

„Mean“ je očekivanje.

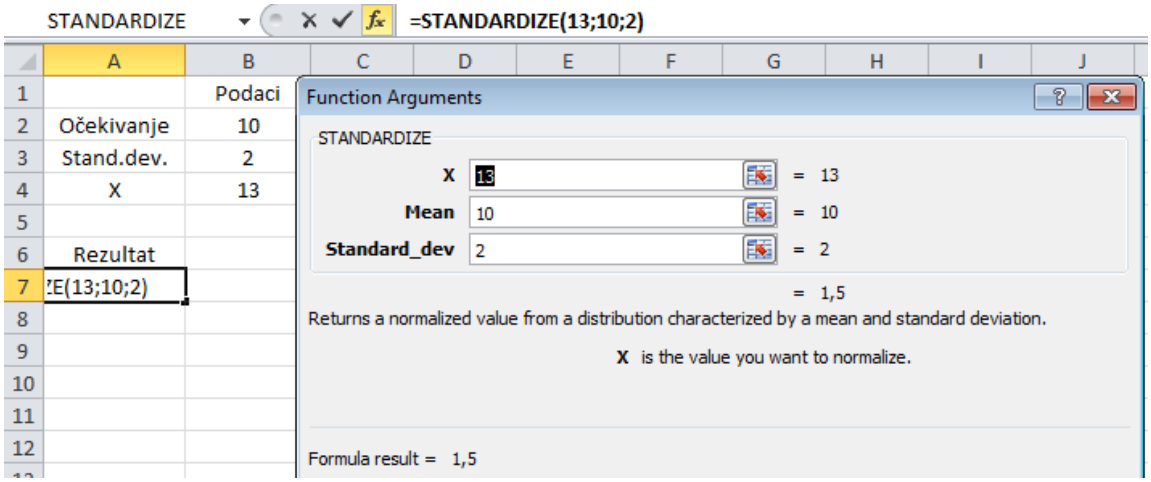
„Standard_dev“ je standardna devijacija razdiobe.

Jednadžba:

$$Z = \frac{X - \mu}{\sigma} \tag{3.4}$$

Primjer:

Pretpostavi li se da se izmjerena jakost struje u vodiču pokorava zakonu normalne razdiobe s očekivanjem μ = 10 A, standardnom devijacijom σ = 2 A i stvarnom vrijednošću X = 13 A, funkcija vraća normaliziranu vrijednost, kako je prikazano na slici dolje [Slika 13]:



Slika 13. Funkcija STANDARDIZE

NORM.DIST

Funkcija prikazuje vjerojatnost pojave nekog uzorka u normalno distribuiranoj populaciji za navedenu srednju vrijednost i standardnu devijaciju. Vrlo široko područje primjene u statistici, uključujući testiranje pretpostavke.

Sintaksa:

„X“ je vrijednost za koju se traži vjerojatnost pojave.

„Mean“ je očekivanje populacije iz koje je uzorak.

„Standard_dev“ je standardna devijacija populacije.

„Cumulative“ je logička vrijednost koja određuje oblik funkcije. Ako je njena vrijednost TRUE (tj. 1), funkcija vraća kumulativnu vrijednost vjerojatnosti, a ako je FALSE (tj. 0) rezultat vraća vjerojatnost pojave zadanog X uzorka.

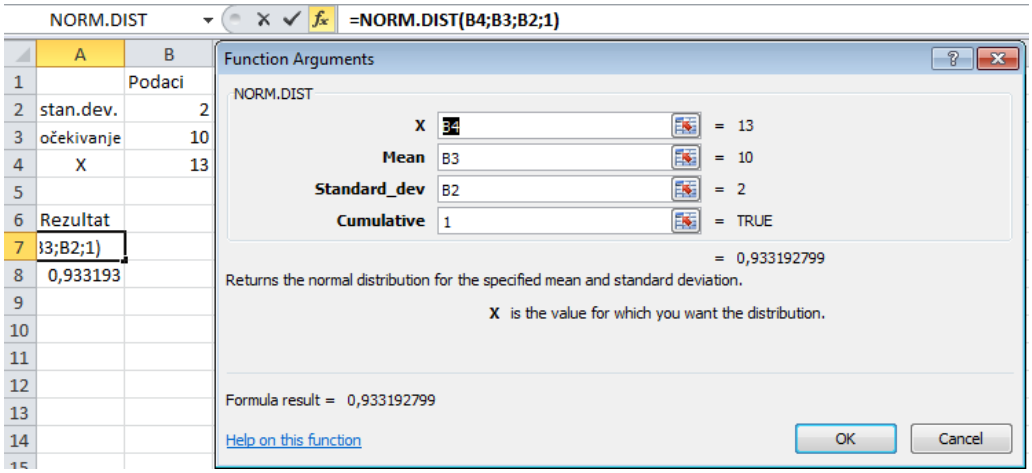
Jednadžba funkcije gustoće normalne razdiobe, kada je „cumulative“ FALSE je :

$$f(x) = \frac{1}{\sigma \times \sqrt{2\pi}} \times e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)} \tag{3.5}$$

naspram integrala zadane formule od minus beskonačno do x kada je njegova vrijednost TRUE.

Primjer:

Primjer će se temeljiti na zadatku korištenom u primjeru funkcije „standardize“ te se dakle treba izračunati vjerojatnost da će struja premašiti iznos od 13 A uz očekivanje od 10 A i standardnu devijaciju od 2 A. Rezultati su prikazani na slici dolje [Slika 14]:



Slika 14. Funkcija NORM.DIST

Kako funkcija uz vrijednost „cumulative“ TRUE vraća površinu od -∞ do X, a potreban je upravo suprotni dio ispod krivulje normalne razdiobe, tražena vrijednost je 1 – vrijednost dobivena funkcijom[14]. Dakle, vjerojatnost da će struja premašiti iznos od 13 A je 1 – 0,933192 = 0,066808. Iz toga slijedi da je vjerojatnost da struja premaši zadanu vrijednost 6,68%.

VAR.P

Funkcija izračunava varijancu na temelju cijele populacije.

Sintaksa:

„Number1;Number2....“ su elementi, tj. populacija na kojoj se izračunava varijanca.

Funkcija VAR.P pretpostavlja da su njezini argumenti cijela populacija.

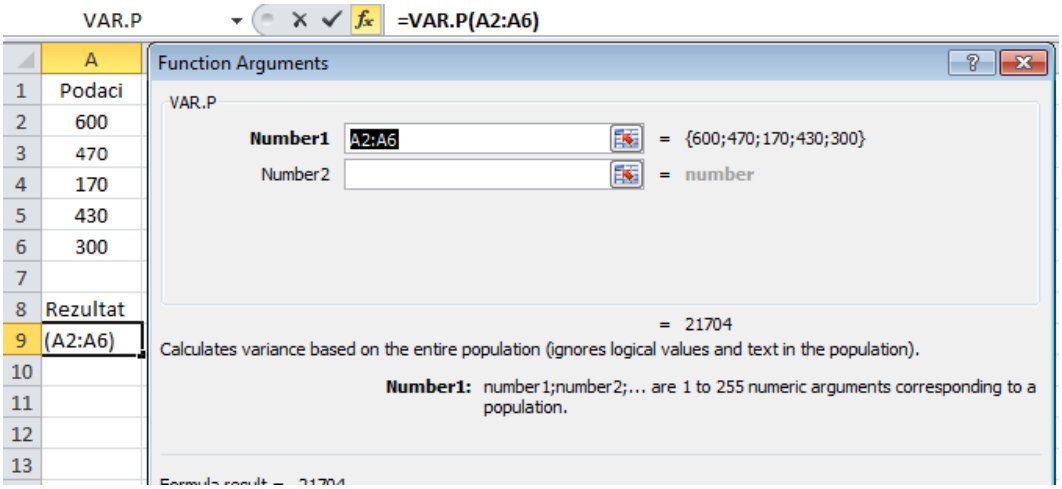
Ako podaci koji se koriste nisu cijela populacija već uzorak treba izračunati procjenu varijance S^2 funkcijom VAR.S.

Jednadžba:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \tag{3.6}$$

gdje je x̄ aritmetička sredina populacije, a n veličina populacije.

Primjer:



Slika 15. Funkcija VAR.P

STDEV.P

Izračunava standardnu devijaciju temeljenu na cijeloj populaciji. Standardna devijacija mjera je širine raspršenosti vrijednosti.

Sintaksa:

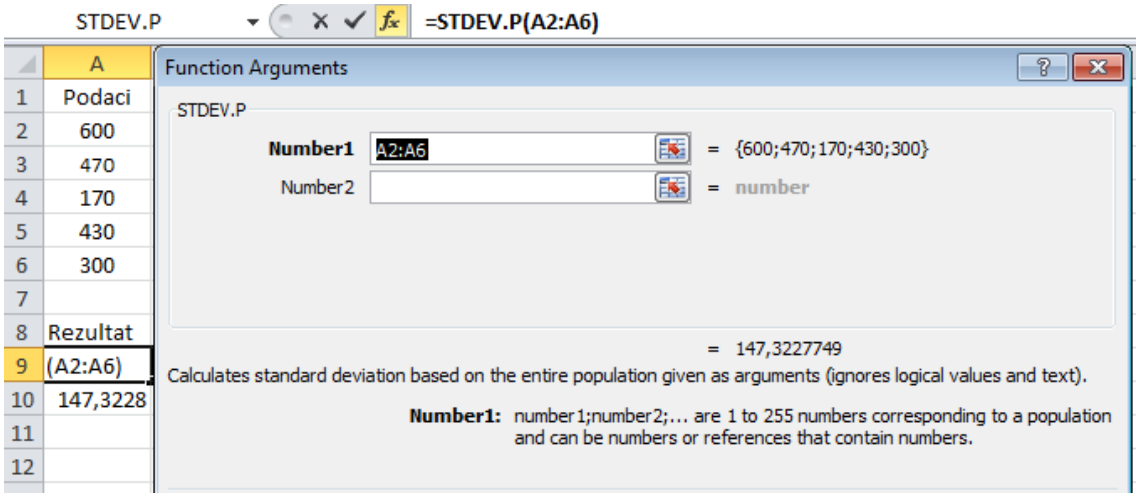
„Number1;Number2...“ su elementi, tj. populacija za koju će se izračunati standardna devijacija.

Ova funkcija izračunava standardnu devijaciju za cijelu populaciju. Ako su podaci za koje se želi izračunati standardna devijacija samo uzorak neke populacije mora se izračunati procjena standardne devijacije S pomoću funkcije STDEV.S.

Jednadžba:

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \tag{3.7}$$

Primjer:



Slika 16. Funkcija STDEV.P

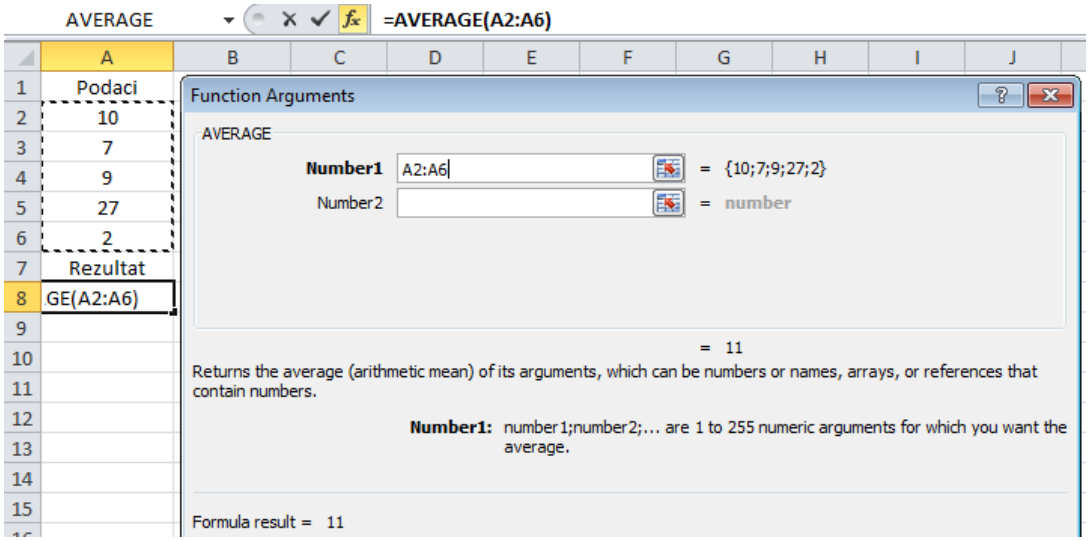
AVERAGE

Funkcija vraća aritmetičku sredinu niza podataka.

Sintaksa:

„Number1;Number2...“ su elementi za koje će funkcija vratiti aritmetičku sredinu. Kod računanja mora se voditi računa da postoji razlika između praznih ćelija i ćelija koje sadrže vrijednost nula.

Primjer:



Slika 17. Funkcija AVERAGE

DEVSQ

Prikazuje zbroj kvadrata odstupanja točaka od srednje vrijednosti uzorka.

Sintaksa:

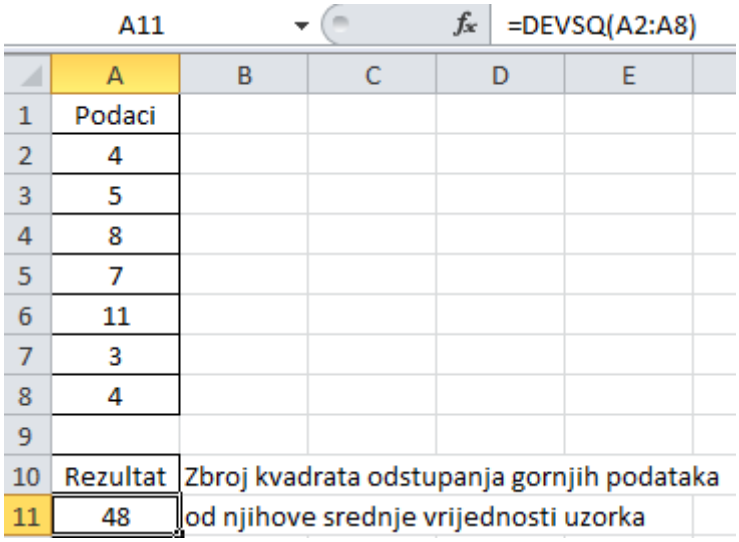
„Number1, Number2...“ su podaci za koje se izračunava zbroj kvadrata odstupanja.

Napomene:

- ćelije s vrijednošću nula su uključene
- jednadžba:

$$DEVSQ = \sum (x - \bar{x})^2 \tag{3.8}$$

Primjer:



Slika 18. Funkcija DEVSQ

HARMEAN

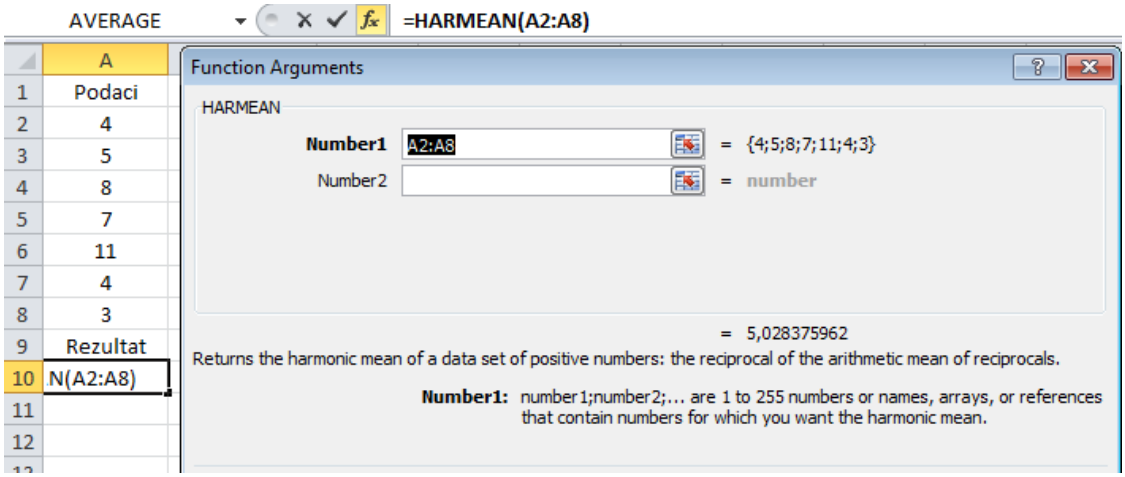
Funkcija vraća harmonijsku sredinu niza podataka. Harmonijska sredina je recipročna vrijednost aritmetičke sredine recipročnih vrijednosti.

Sintaksa:

„Number1;Number2...“ su brojevi ili niz podataka za koje se želi izračunati harmonijska sredina. Harmonijska sredina je uvijek manja od geometrijske sredine, koja je uvijek veća od aritmetičke sredine. Jednadžba:

$$\frac{1}{H_y} = \frac{1}{n}\sum \frac{1}{y_j} \tag{3.9}$$

Primjer:



Slika 19. Funkcija HARMEAN

MEDIAN

Funkcija vraća medijan zadanih brojeva. Medijan je broj koji se nalazi u sredini zadanog skupa brojeva, što znači da će polovica brojeva imati vrijednost veću od medijana, a polovica manju.

Sintaksa:

„Number1, number2“ su podaci za koje se želi izračunati medijan

Napomene:

- ako se u skupu nalazi paran broj vrijednosti, funkcija MEDIAN izračunava prosjek dvaju brojeva u sredini

Primjer:

Slika 20. Funkcija MEDIAN

MODE.SNGL

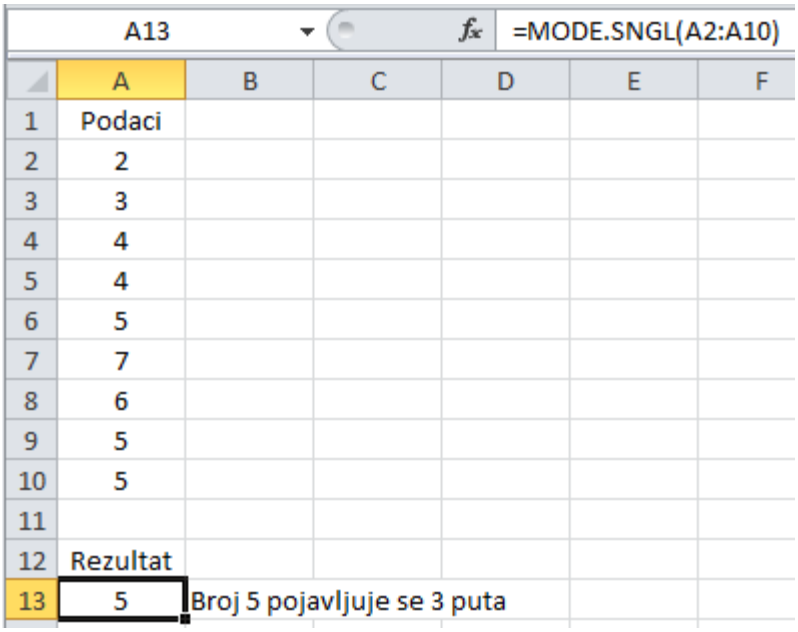
Vraća vrijednost koja se najčešće javlja, odnosno ponavlja u zadanom polju ili rasponu podataka. Kao i MEDIAN, mod je mjera mjesta.

Sintaksa:

„Number1, number2“ su podaci za koje se izračunava mod.

Napomene:

- ako zadani skup podataka ne sadrži vrijednosti koje se ponavljaju, funkcija vraća vrijednost pogreške



Slika 21. Funkcija MODE.SNGL

U skupu podataka, mod je vrijednost koja se najčešće javlja, medijan je vrijednost koja se nalazi u sredini, a srednja vrijednost je prosječna vrijednost.

QUARTILE.INC

Funkcija koja izbacuje kvartil niza podataka. Često se koristi kada se treba podijeliti populaciju u skupine, npr. kada se želi naći 25% donjih vrijednosti u populaciji.

Sintaksa:

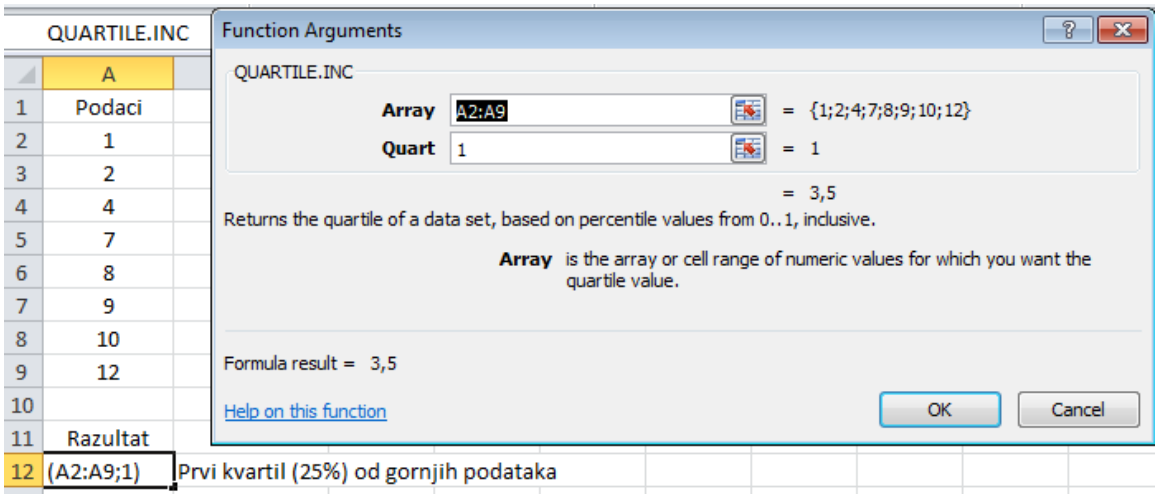
„Array“ je raspon ćelija za koje se želi dobiti vrijednost kvartila, dok „Quart“ definira koju vrijednost će funkcija vratiti, kako prikazuje tablica dolje [Tablica 12].

Tablica 12. Izlazne vrijednosti „quarta“

Vrijednost „quarta“	Izlaz
0	Minimalna vrijednost
1	Prvi kvartil (25%)
2	Medijan (50%)
3	Treći kvartil (75%)
4	Maksimalna vrijednost

Iz navedene tablice vidi se da kada je quart jednak 0, 2 i 4, funkcija izbacuje vrijednosti minimuma, medijana odnosno maksimuma.

Primjer:



Slika 22. Funkcija QUARTILE.INC

Nakon obrade jednih od bitnijih statističkih funkcija, sljedeće poglavlje opisat će mogućnosti paketa „Analysis ToolPak“ koji se koristi za složene statističke analize.

3.2.3. Paket za statističku analizu „Analysis ToolPak“

Korištenjem navedenog paketa u složenoj statističkoj ili tehničkoj analizi može se uštedjeti mnogo koraka u postupku i vremena. Potrebno je navesti podatke i parametre za svaku analizu i alat zatim koristi odgovarajuće funkcije makronaredbi i prikazuje rezultate u izlaznoj tablici. Neki alati osim tablica imaju mogućnost i izrade grafikona.

Funkcije analize podataka istodobno se mogu koristiti samo na jednom radnom listu. Prilikom izvođenja analize podataka na grupiranim radnim listovima rezultati će se pojaviti na prvom

radnom listu, a na ostalima će se pojaviti prazno oblikovane tablice[15]. Da bi se analizirali podaci i na ostalim radnim listovima potrebno je za svaki od njih ponovno pokrenuti alat. Paket je potrebno uključiti i aktivirati prije korištenja. U nastavku će se prikazati i ukratko opisati svaki od alata paketa.

3.2.3.1. ANOVA: jednofaktorska analiza varijance

Ovaj alat izvodi jednostavnu analizu varijance na podacima za dva ili više uzoraka. Analiza omogućuje testiranje hipoteze prema kojoj je svaki uzorak uzet iz iste osnovne distribucije vjerojatnosti u odnosu na drugu hipotezu prema kojoj osnovne distribucije vjerojatnosti nisu iste za sve uzorke.

3.2.3.2. ANOVA: dvofaktorska analiza varijance s replikacijom

Alat se koristi kada se podaci mogu svrstati u dvije različite dimenzije. Na primjer, u pokusu pojave prvih pukotina kod čelika, ako se čelik obrađuje različitim metodama (A, B, C) i izlaže se različitim eksploatacijskim uvjetima (niska ili visoka naprezanja). Za svaki od 6 mogućih parova (način obrade, eksploatacijski uvjeti) postoji jednak broj opažanja u vezi s pojavom prvih pukotina. Pomoću Anove može se testirati:

- Jesu li se pukotine za različite vrste obrade čelika pojavile na čeliku iz iste osnovne populacije. Ovdje se zanemaruju temperature.
- Jesu li se pukotina za različite eksploatacijske uvjete pojavile na čeliku iz iste osnovne populacije. Ovdje se zanemaruje vrsta obrade.

Alternativna hipoteza govori da postoje utjecaji nastali zbog specifičnih parova koji nadilaze razlike nastale samo zbog jednog faktora.

3.2.3.3. ANOVA: dvofaktorska analiza varijance bez replikacije

Alat je jednak prethodnom alatu samo što se pretpostavlja da postoji samo jedno opažanje za svaki par.

3.2.3.4. Korelacija

Alat za analizu korelacije koristi se kada postoje više od dvije mjerne varijable za svaki od n elemenata, nasuprot funkciji CORREL koja izračunava koeficijent korelacije između samo dvije mjerne varijable. Alat daje matricu korelacije koja prikazuje vrijednost funkcije CORREL kada se primjeni na svaki mogući par mjernih varijabli. Koeficijent korelacije je skaliran pa njegova vrijednost ne ovisi o jedinicama u kojima su izražene mjerne varijable. Ovim se alatom, dakle može ispitati svaki par mjernih varijabli za određivanje da li su blizu, odnosno, ako je korelacija:

- Pozitivna, velike vrijednosti jedne varijable pridružene su velikim vrijednostima druge
- Negativna, male vrijednosti jedne varijable pridružene su velikim vrijednostima druge
- Nula, vrijednosti u oba skupa nisu povezane.

3.2.3.5. Kovarijanca

Ovaj alat sličan je alatu korelacije uz razliku u tome što se koeficijenti korelacije iskazuju u rasponu između -1 i 1, uključujući i -1 i 1, a odgovarajuće kovarijance nisu skalirane. I korelacija i kovarijanca izračunavaju u kojoj se mjeri dvije varijable razlikuju, ali zbog toga što kovarijanca nije skalirana iz rezultata se ne može očitati jakost sličnosti kretanja varijabli već samo njeno postojanje i predznak.

3.2.3.6. Deskriptivna statistika

Ovaj alat stvara statističko izvješće za podatke iz ulaznog raspona. U tablici vraća aritmetičku sredinu, standardnu devijaciju i pogrešku, varijancu podataka, raspon, minimalnu i maksimalnu vrijednost, koeficijent spljoštenosti i interval pouzdanosti, kako je prikazano na slici dolje [Slika 23] .

1	Podaci			
2	4		*Podaci*	
3	7			
4	9		Mean	42,875
5	56		Standard Error	16,0762
6	34		Median	34
7	56		Mode	56
8	143		Standard Deviation	45,47036
9	34		Sample Variance	2067,554
10			Kurtosis	3,588046
11			Skewness	1,748674
12			Range	139
13			Minimum	4
14			Maximum	143
15			Sum	343
16			Count	8
17			Confidence Level(95,0%)	38,01417

Slika 23. Tablica rezultata alata deskriptivne statistike

3.2.3.7. Eksponencijalno zaglađivanje

Ovaj alat predviđa vrijednost na temelju prognoze iz prethodnog razdoblja, prilagođenu za pogrešku iz te prethodne prognoze. Koristi se konstanta zaglađivanja čija veličina određuje u kojoj se mjeri prognoze odazivaju na pogreške iz prethodne prognoze. Vrijednosti konstante zaglađivanja od 0,2 do 0,3 su prihvatljive te označavaju da se trenutna prognoza treba prilagoditi za 20 do 30 posto kod pogreške u prethodnoj prognozi.

3.2.3.8. F-test s dva uzorka za varijance

Ovaj alat izvodi F-test s dva uzorka za usporedbu dviju varijanci populacija. Vraća rezultate testiranja nulte hipoteze prema kojoj ta dva uzorka dolaze iz distribucija s jednakim varijancama u odnosu na zamjensku hipotezu prema kojoj se varijance razlikuju od osnovnih distribucija. Alat izračunava vrijednost $F_{rač}$ nekog F omjera. Ako je vrijednost $F_{rač}$ približno 1, znači da su varijance osnovne populacije jednake.

3.2.3.9. Fourierova analiza

Ovaj alat rješava probleme u linearnim sustavima i analizira periodične podatke pomoću postupka brze Fourierove transformacije podataka. Također, mogu se raditi i inverzne transformacije, tj. inverz pretvorenih podataka prikazuje izvorne podatke.

3.2.3.10. Histogram

Alat histogram računa pojedinačne i kumulativne frekvencije raspona ćelije podataka i niza podataka. Stvara podatke o broju pojava određenih vrijednosti u skupu podataka.

3.2.3.11. Pomični prosjeci

Alat „Moving average“ predviđa vrijednosti u razdoblju prognoziranja na temelju prosječne vrijednosti varijable tijekom određenog broja prethodnih razdoblja. Pomoću ove funkcije korisnik dobiva informaciju o trendu koja bi bila sakrivena da se napravio prosjek svih proteklih podataka.

3.2.3.12. Generiranje slučajnog broja

Ovim alatom korisnik dobiva raspon slučajnih brojeva izvučenih iz jedne od nekoliko distribucija (normalna, binomna, poissonova...).

3.2.3.13. Rangiranje i percentili

Pomoću rangiranja i percentila alat stvara tablicu koja sadrži redni i postotni rang svake vrijednosti u skupu podataka te može analizirati položaj relativnih vrijednosti.

3.2.3.14. Regresija

Korištenjem metode najmanjih kvadrata za prilagodbu pravca kroz skup opažanja alat izvodi linearnu regresijsku analizu. Uz pomoć ovog alata moguće je istražiti kako vrijednosti jedne ili više nezavisnih varijabli utječe na zavisnu varijablu. Na primjer, kako će na ponašanje ležaja utjecati njegova starost, podmazanost ili veličina.

3.2.3.15. Uzorkovanje

Alat „Sampling“ slučajno odabire uzorak iz populacije odnoseći se prema ulaznom rasponu kao prema populaciji. Dobivanjem reprezentativnog uzorka, veliki setovi podataka lakše se mogu analizirati i prikazati grafički.

3.2.3.16. t-test

T-test za analizu s dva uzorka ispituje jesu li jednake srednje vrijednosti na kojima se temelje uzorci. Dostupna su tri alata:

- Test uparena dva uzorka za srednje vrijednosti

- Test dva uzorka s pretpostavkom jednakih varijanci
- Test dva uzorka s pretpostavkom različitih varijanci.

Test uparena dva uzorka za srednje vrijednosti koristi se kada u uzorcima dolazi do prirodnog uparivanja opažanja, na primjer kada se grupa uzoraka testira dvaput, prije i poslije pokusa. Pokušava se utvrditi dolaze li opažanja prije i poslije pokusa iz distribucije s jednakim srednjim vrijednostima populacija. Ovaj oblik t-testa ne pretpostavlja da su varijance obje populacije jednake.
Test dva uzorka s pretpostavkom jednakih varijanci izvodi studentov test s dva uzorka. Pretpostavlja se da oba skupa dolaze iz raspodjele s jednakim varijancama. Još je poznat kao homoscedastični t-test [15]. Određuje vjerojatnost da li su dva uzorka došla iz raspodjela jednakih srednjih vrijednosti.
Test dva uzorka s pretpostavkom različitih varijanci također izvodi studentov t-test na dva uzorka u kojem se pretpostavlja da su dva skupa podataka uzeta iz distribucija s različitim varijancama. Poznat je i pod nazivom heteroscedastični t-test [15]. Koristi se kada u dva uzorka postoje različiti subjekti.

3.2.3.17. Z-test

Alat izvodi z-test s dva uzorka za srednje vrijednosti s poznatim varijancama. Koristi se za provjeru nulte hipoteze o nepostojanju razlike između dvije srednje vrijednosti populacija u odnosu na jednostranu ili dvostranu zamjensku hipotezu. Ako varijance nisu poznate, umjesto ovog testa treba koristiti funkciju Z.TEST.

3.2.4. Alat „Solver“

Alat „Solver“ rješava najsloženije oblike „what-if“ analize i probleme linearnog i nelinearnog optimiranja. Ideja ovog alata je u dostizanju optimalnih vrijednosti na temelju zadanih ograničenja (fiksne ulazne veličine, razne nepromjenjive vrijednosti, „uska grla“) i niza parametarskih veličina koje se mogu mijenjati – to mogu biti vrijednosti ali i odabrane metode kojima će složeni matematički model biti podvrgnut. Time se definira model problema i zadaje dostizanje optimalne vrijednosti (minimum, maksimum ili unaprijed zadana vrijednost). Ograničenje alata je veće od realnih potreba većine korisnika[16]:

- Može se definirati do 200 ulaznih parametara
- Može se definirati 100 ograničenja.

Alat je potrebno uključiti i dolazi u paketu s više nego dovoljno zanimljivih primjera. U poglavlju optimizacije bit će prikazan primjer korištenja alata i optimizacija stvarnog proizvodnog poduzeća.

4. STATISTIČKA ANALIZA

U ovom poglavlju biti će prikazano nekoliko različitih testova koji se provode u svrhu dobrog vođenja i praćenja proizvodnje s ciljem povećanja efikasnosti i rane detekcije nepravilnosti u proizvodnji. Podaci su uzeti iz stvarnih proizvodnih pogona.

4.1. T-test uparenih uzoraka

Analizom uzoraka od 17 komada testnih epruveta nodularnog lijeva dobivene su vrijednosti vlačne čvrstoće R_m. Nakon što su vrijednosti zabilježene uzorak je poslan na toplinsku obradu nakon koje se također vršio test vlačne čvrstoće i podaci su pohranjeni. Cilj ovakvog pokusa je utvrditi da li je toplinska obrada imala utjecaja na vlačnu čvrstoću testnih epruveta. U tablici su dani podaci vrijednosti vlačne čvrstoće prije i poslije toplinske obrade[17].

Tablica 13. Podaci vrijednosti vlačnih čvrstoća [17]

Komad	R_m prije T.O.	R_m poslije T.O.
1	512	518
2	549	549
3	517	502
4	493	503
5	488	491
6	490	486
7	507	485
8	499	490
9	517	518
10	496	485
11	518	521
12	436	499
13	488	494
14	461	483
15	503	478

16	478	487
17	596	611

Kako se ovdje radi o jedno te istom uzorku ali s mjerenjima prije i poslije pokusa, te kako je broj elemenata u uzorku manji od 30, najprimjereniji test biti će t-test uparenih uzoraka. Nakon provedenog testa u Excelu rezultati su dostupni odmah te je lako donijeti zaključak o potrebi toplinske obrade.

t-Test: Paired Two Sample for Means		
	Rm prije T.O.	*Rm poslije T.O.*
Mean	502,8235294	505,8823529
Variance	1196,279412	1065,110294
Observations	17	17
Pearson Correlation	0,826085428	
Hypothesized Mean Difference	0	
df	16	
t Stat	-0,633422568	
P(T<=t) one-tail	0,267703986	
t Critical one-tail	1,745883676	
P(T<=t) two-tail	0,535407973	
t Critical two-tail	2,119905299	

Slika 24. T-test uparena dva uzorka

U dobivenim rezultatima vidi se da su očekivanje i varijanca približno slični u oba mjerenja. Red „df (eng. degree of freedom)“ govori koliko ima stupnjeva slobode, dok P vrijednost predstavlja našu izračunatu vrijednosti pomoću koje se dolazi do zaključka. Iz slike se vidi da postoje P i t vrijednosti za jednu i za obje strane krivulje razdiobe. P i t vrijednosti za jednu stranu krivulje razdiobe koriste se kada postoji definirani smjer kretanja pretpostavljene razlike očekivanja (npr. tražena vrijednost mora biti veća od nekog broja). Kako je u našem pokusu pretpostavljena razlika očekivanja jednaka 0 (nulta hipoteza), možemo zaključiti da toplinska obrada nije utjecala na vlačnu čvrstoću testnih epruveta jer je P vrijednost veća od pogreške α. Također vidimo da je izračunata vrijednost t, „t stat“, unutar granica „t critical two tail“.

4.2. T-test dva uzorka jednakih varijanci

U sljedećem problemu provjerit će se veza dva uzorka sa strojevima na temelju dimenzija uzoraka [18] uz pretpostavku jednakih varijanci.

Dva stroja proizvode istovrsne proizvode kod kojih je važna dimenzija koja je funkcija samo podešenosti stroja. Analizom dva uzorka treba se provjeriti da li su strojevi različito podešeni.

Tablica 14. Dimenzije proizvoda dva stroja [18]

Stroj	Dimenzije proizvoda u mm							
A	84,2	84,1	84,5	84,3	84,4	84,3	84,4	84,2
B	84,3	83,9	84,1	84,2	84,0			

Pretpostavljaju se hipoteze te se podaci ubace u Excel.

t-Test: Two-Sample Assuming Equal Variances

	Stroj A	*Stroj B*
Mean	84,3	84,1
Variance	0,01714286	0,025
Observations	8	5
Pooled Variance	0,02	
Hypothesized Mean Difference	0	
df	11	
t Stat	2,48069469	
P(T<=t) one-tail	0,01526801	
t Critical one-tail	1,79588482	
P(T<=t) two-tail	0,03053602	
t Critical two-tail	2,20098516	

Slika 25. T-test dva uzorka jednakih varijanci

Uz α=0,05, vidi se da izračunati t („t stat“) pada u područje odbacivanja nulte hipoteze, tj. odbacuje se hipoteza H_0: $\mu_1 - \mu_1 = D$ (D=0). Vidi se da je vrijednost „t Critical two-tail“ manja od izračunate vrijednost t odnosno da je P vrijednost manja od vrijednosti pogreške α. Stoga je razumno prihvatiti hipotezu H_1, dakle zaključiti da strojevi od kojih su uzeti uzorci nisu jednako podešeni.

4.3. F-test

Prethodni problem analizirat će se i F testom koji nam omogućava testiranje hipoteze o jednakosti varijanci osnovnih normalnih distribucija.

Uz iste podatke kao u prethodnom problemu pretpostavljaju se hipoteze navedene u jednadžbi (2.25):

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \tag{4.1}$$

$$H_1: \frac{\sigma_1^2}{\sigma_2^2} \geq 1 \; ili \; \frac{\sigma_1^2}{\sigma_2^2} \leq 1$$

Nakon unošenja podataka u radni list Excela program vraća sljedeće vrijednosti:

F-Test Two-Sample for Variances

	Stroj B	*Stroj A*
Mean	84,1	84,3
Variance	0,025	0,017142857
Observations	5	8
df	4	7
F	1,458333	
P(F<=f) one-tail	0,310533	
F Critical one-tail	4,120312	

Slika 26. F-test dva uzorka

Za stupanj slobode brojnika 4 i nazivnika 7, iz tablica se može očitati vrijednost 4,12 navedena u ćeliji „F critical one-tail“. Zbog toga što je procjena varijance uzoraka sa stroja B veća od procjene varijance uzoraka sa stroja A, u brojnik jednadžbe izraza za izračunatu vrijednost F:

$$F = \frac{S_1^2}{S_2^2} \tag{4.2}$$

stavit će se S_B^2, dok će se u nazivnik pisati S_A^2:

$$F = \frac{S_B^2}{S_A^2} \tag{4.3}$$

Zbog toga je izračunati F uvijek veći od 1. Dobiva se vrijednost za F=1,458 što je manje od tablične vrijednosti kritičnog F. S obzirom na takav rezultat hipoteza H_0 prihvaća se kao istinita. Drugim riječima, procjene varijanci ne razlikuju se značajno, pa ne možemo na bazi gornjih uzoraka zaključiti da uzorci potječu iz normalnih razdioba s različitim varijancama.

4.4. Z – test dva uzorka poznatih varijanci

Kada se u uzorku koji će se testirati nalazi više od 30 elemenata, da bi test bio valjan ne smije se koristiti t test već z test. Ako se radi o dva uzorka kojima su poznate varijance tada se koristi z test dva uzorka poznatih varijanci.

U sljedećem problemu radi se o dva stroja za pakiranje sipkih proizvoda. Kako ti strojevi pakiraju iste proizvode bitno je provjeriti da li to rade jednako precizno.

z-Test: Two Sample for Means		
	Stroj 1	*Stroj 2*
Mean	299,588235	299,852941
Known Variance	38,7344	30,7359
Observations	34	34
Hypothesized Mean Difference	0	
z	-0,185184	
P(Z<=z) one-tail	0,42654236	
z Critical one-tail	1,64485363	
P(Z<=z) two-tail	0,85308472	
z Critical two-tail	1,95996398	

Slika 27. Z-test dva uzorka

Kako je izračunata z vrijednost manja od kritične z vrijednosti, tj. P vrijednost veća je od pogreške α, možemo reći da ne postoje značajne razlike između stroja 1 i stroja 2 te se nulta hipoteza prihvaća.

5. OPTIMIZACIJA PROIZVODNOG SUSTAVA

Metode optimizacije prvi put su se pojavile 30-tih godina prošlog stoljeća kao metode rješavanja problema planiranja proizvodnje [19]. Omogućuju nalaženje najboljih rješenja različitih vrsta problema te su samim time vrlo pogodne u proizvodnom i poslovnom svijetu. Tipični proizvodni problemi vezani su za korištenje ograničenih resursa (ljudi, opreme, materijala) kojima se nastoji ostvariti najveća moguća dobit, pritom osigurati što veću kvalitetu uz što niži trošak. Jedna od najčešće korištenih metoda je linearno programiranje koje omogućuje nalaženje najpovoljnijih rješenja problema u kojima su i funkcija cilja (npr. maksimizacija dobiti) i utrošci resursa (npr. materijala ili vremena) linearno proporcionalni vrijednostima nezavisnih varijabli (npr. broj proizvedenih komada). Područje primjene linearnog programiranja je široko: proizvodnja, transport, distribucija, marketing, planiranje i raspored zaposlenika. U ovom poglavlju bit će prikazan primjer linearnog programiranja na stvarnom proizvodnom poduzeću proveden pomoću alata „Solver“ u MS Excelu.

Zadatak:

Proizvodno poduzeće proizvodi različite metalne proizvode. Proizvodnja svakog od njih obuhvaća strojnu obradu, brušenje i montažu. Vrijeme obrade za svaki od proizvoda u pojedinim fazama, te raspoloživo vrijeme za tjedan dana prikazani su u tablici[20]:

Tablica 15. Proizvodni podaci [20]

	Strojna obrada [h/kom]	Brušenje [h/kom]	Montaža [h/kom]
Proizvod 1	3	1	2
Proizvod 2	2	1	1
Proizvod 3	2	2	2
Proizvod 4	4	3	1
Raspoloživo vrijeme	480	400	400

Definiranje problema:

Dobit za svaki od proizvoda iznosi redom 6, 4, 6 i 8 novčanih jedinica po komadu. Poduzeće je s jednim kupcem ugovorilo tjednu isporuku 45 komada proizvoda 1 i 80 komada proizvoda

2 i 4 zajedno. Svim ostalim kupcima poduzeće može tjedno isporučiti koliko god može proizvesti proizvoda 1, 2 i 4, a proizvoda 3 maksimalno 25 komada.

Potrebno je odrediti proizvodni program koji omogućava maksimalnu dobit na mjesečnoj razini.

Rješavanje problema:

U prvom dijelu rješavanja ovog proizvodnog problema na temelju opisnog modela postavlja se matematički model. Ako sa x_1, x_2, x_3, x_4 označimo proizvode 1,2,3 i 4, funkcija cilja ima oblik:

$$\max z = 6x_1 + 4x_2 + 6x_3 + 8x_4$$

Za definiranje ograničenja, zadane parametre za postupke proizvodnje upisuje se u lijevoj strani jednadžbe, a zadana maksimalna ograničenja ukupnog broja sati za svaki postupak upisuje se na desnoj strani. Ograničenja raspoloživog vremena redom za strojnu obradu, brušenje i montažu glase:

$$3x_1 + 2x_2 + 2x_3 + 4x_4 \leq 480$$

$$1x_1 + 1x_2 + 2x_3 + 3x_4 \leq 400$$

$$2x_1 + 1x_2 + 2x_3 + 1x_4 \leq 400$$

Naravno, uz ograničenja raspoloživog vremena operacija moraju se uključiti i ograničenja proizvodnje koje je poduzeće prihvatilo ispuniti pri ugovaranju posla:

$$x_1 \geq 45$$

$$x_2 + x_4 \geq 80$$

$$x_3 \leq 25$$

Također, s obzirom da proizvodnja ne može biti negativna (količine proizvoda ne mogu imati negativni predznak) potrebno je definirati i ograničenja nenegativnosti:

$$x_1 \geq 0$$

$$x_2 \geq 0$$

$$x_3 \geq 0$$

$$x_4 \geq 0$$

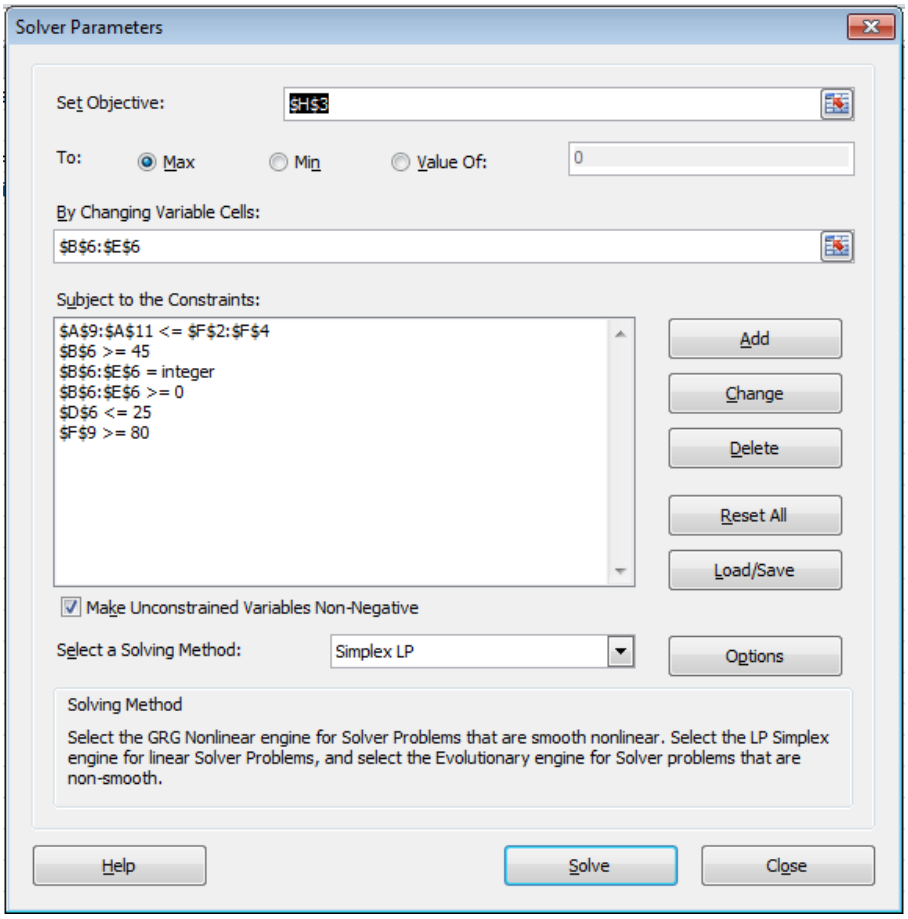
Zadaća u rješavanju problema proizvodnje je maksimizirati funkciju dobiti. Visina dobiti ovisi o vrijednosti varijabli u funkciji cilja te se pokušava maksimizirati pronalaskom optimalnih vrijednosti količina proizvoda koje će se proizvoditi.

Kada se postavi matematički model potrebno je unijeti podatke u Excel i oblikovati Strukturni model kako je prikazano na slici:

	A	B	C	D	E	F	G	H	I	J	K
1		proizvod 1	proizvod 2	proizvod 3	proizvod 4	raspoloživost		Funkcija cilja			
2	str.obrada	3	2	2	4	480		Maksimum			
3	brušenje	1	1	2	3	400		24	"=SUMPRODUCT((B5:E5;B6:E6))		
4	montaža	2	1	2	1	400			ili B5*B6+C5*C6+D5*D6+E5*E6		
5	cijena	6	4	6	8						
6	količina	1	1	1	1						
7											
8	Iskorištenost vremena										
9	11	"=SUMPRODUCT(B2:E2;$B$6:$E$6)			x2+x4=	2					
10	7	"=SUMPRODUCT(B3:E3;$B$6:$E$6)									
11	6	"=SUMPRODUCT(B4:E4;$B$6:$E$6)									

Slika 28. Strukturni model

Kada se podaci unesu u novi radni list Excela uz pomoć jednostavnih matematičkih funkcija izračuna se funkcija cilja i stupac iskorištenosti vremena strojne obrade, brušenja i montaže. Zatim će se te ćelije uzimati kao reference za postavljanje ograničenja kao što je prikazano u sljedećem koraku na slici ispod:



Slika 29. Parametri

Iz slike vidi se da je pod rubrikom „Set objective“ odabrana ćelija H3 koja predstavlja funkciju cilja tj. maksimalnu dobit, koja se postiže mijenjanjem niza ćelija od B6 do E6. Maksimalna dobit mora se postići uz ograničenja navedena u prozoru „Subject to the Constraints“. Uz gore navedena ograničenja, dodano je i ograničenje „B6:E6 = integer“ koje ograničava proizvodne količine na cijele brojeve, tj. eliminira mogućnost pojave rezultata optimalne količine od npr. 3,5 proizvoda. Također, vidi se da je kao metoda rješavanja odabrana „Simplex LP“ tj., Simplex metoda linearnog programiranja. Pritiskom tipke „Solve“, alat izbacuje optimalno rješenje te se otvara novi radni list s tumačenjem rješenja i prikazom cilja, varijabli i ograničenja.

Result: Solver found an integer solution within tolerance. All Constraints are satisfied.

Solver Engine

Engine: Simplex LP

Solution Time: 0,016 Seconds.

Iterations: 1 Subproblems: 4

Solver Options

Max Time Unlimited, Iterations Unlimited, Precision 0,000001, Use Automatic Scaling

Max Subproblems Unlimited, Max Integer Sols Unlimited, Integer Tolerance 1%, Assume NonNegative

Objective Cell (Max)

Cell	Name	Original Value	Final Value
$H$3	brušenje Maksimum	24	1010

Variable Cells

Cell	Name	Original Value	Final Value	Integer
$B$6	količina proizvod 1	1	46	Integer
$C$6	količina proizvod 2	1	146	Integer
$D$6	količina proizvod 3	1	25	Integer
$E$6	količina proizvod 4	1	0	Integer

Slika 30. Rezultati

Rezultat, tj. maksimalna dobit za zadane uvjete i ograničenja vidi se u stupcu „Final Value“ prve tablice i iznosi 1010. Varijable, tj. količina proizvoda koja će se proizvoditi vidi se u drugoj tablici također u stupcu „Final Value“. Iz danih rezultata vidi se da su ograničenja zadovoljena te se mogu uočiti neke zanimljivosti. Vidi se da se proizvod 1 i proizvod 3 nalaze na samoj granici svojih ograničenja. Također, proizvod 4 uopće ne doprinosi povećanju dobiti i razlog se može tražiti u dugotrajnoj obradi. Na slici ispod vidi se stanje ograničenja, da li su iskorištena do kraja kod ostvarivanja maksimalne dobiti ili ne.

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
$A$9	Iskorištenost vremena	480	$A$9<=$F$2	Binding	0
$A$10	Iskorištenost vremena	242	$A$10<=$F$3	Not Binding	158
$A$11	Iskorištenost vremena	288	$A$11<=$F$4	Not Binding	112
$F$9	x2+x4= raspoloživost	146	$F$9>=80	Not Binding	66
$B$6	količina proizvod 1	46	$B$6>=45	Binding	0
$B$6	količina proizvod 1	46	$B$6>=0	Binding	0
$C$6	količina proizvod 2	146	$C$6>=0	Not Binding	146
$D$6	količina proizvod 3	25	$D$6>=0	Binding	0
$E$6	količina proizvod 4	0	$E$6>=0	Binding	0
$D$6	količina proizvod 3	25	$D$6<=25	Binding	0
$B$6:$E$6=Integer					

Slika 31. Stanje ograničenja

Može se primijetiti da je raspoloživo vrijeme za strojnu obradu u potpunosti iskorišteno dok operacije brušenja i montaže nisu u potpunosti iskoristile raspoloživo vrijeme. U tom podatku može se tražiti prostor za poboljšanje u vidu pronalaska boljeg načina iskorištenja strojeva za obradu ili nabave dodatnog stroja kako operacije brušenja i montaže ne bi čekale na izvršenje rada i imale puno praznog hoda.

6. ZAKLJUČAK

Statistika u inženjerstvu vrlo je bitna kao grana znanosti. Praćenje procesa i njihovo poboljšavanje, kontrola kvalitete i poznavanje preciznosti strojeva ne bi bilo moguće bez sposobnosti primjene teorijskog znanja iz statistike. Međutim, samo poznavanje teorijskih znanja i metoda vrlo često nije dovoljno za efikasno vođenje i planiranje procesa proizvodnje. Zbog toga se traži brži i jednostavniji način statističke analize i optimizacije. Jedan od tih načina upravo je MS Excel. Širokodostupan (relativno niska cijena cijelog paketa Microsoft Office) i izrazito jednostavan za korištenje pruža mnoštvo mogućnosti analize podataka te prikaza rezultata. Izbor statističke analize te njene provedbe vrlo je raznolika. Korištenje raznih funkcija i paketa alata kao što su „Analysis ToolPak“ vrlo je jednostavno i gotovo trenutno daje sve potrebne rezultate za donošenje zaključka i grafički prikaz rješenja. Optimizacija proizvodnog sustava sa zadanim ograničenjima lako se provodi pomoću alata „Solver“ te odabirom funkcije cilja vraća rezultate koje je potrebno namjestiti kako bi postigli zadani cilj.

LITERATURA

[1] Cajner H.: Osnove statistike, skripte s predavanja, 2013

[2] Kocеić Bilan N,: Primijenjena statistika, PMF, Split, 2011

[3] Lulić, I.: Završni rad, Uporaba metode regresijske analize u rješavanju problema vezanih uz inženjersku praksu, Zagreb, 2014

[4] Cajner H.: Analiza varijance, Skripte s predavanja, Zagreb, 2013

[5] Grubišić N.: Diplomski rad, Karakterizacija procesa uporabom metode regresijske analize i analize varijance, Zagreb, 2014

[6] http://mathnstats.com/index.php/hypothesis-testing/

[7] Cajner H.: Osnove teorije uzoraka, skripte s predavanja, 2013

[8] Montgomery D.C., Runger G.C.: Applied Statistics and Probability for Engineers, John Wiley & Sons, 2010

[9] http://www.statsoft.com/Products/STATISTICA/Product-Index

[10] http://www.minitab.com/en-us/products/minitab/features/

[11] http://www.ss-otocac.skole.hr/dokumenti?dm_document_id=114&dm_dnl=1

[12] http://www.ilri.org/biometrics/TrainingResources/Documents/University%20of%20Reading/SSCother%20resources/GuidelinesGoodStatistGraphics.pdf

[13] https://support.office.microsoft.com/en-US/article/Excel-functions-by-category-7fd9655a-4e87-400a-ae5c-c48f16afde0c?ui=en-US&rs=en-US&ad=US

[14] Cajner H.: Raspodjele, skripte s predavanja, 2013

[15] https://support.office.microsoft.com/hr-HR/article/Slo%C5%BEena-analiza-podataka-pomo%C4%87u-skupa-alata-za-analizu-6c67ccf0-f4a9-487c-8dec-bdb5a2cefab6#__toc340479285

[16] http://www.ic.ims.hr/faq/office2007/excel2007/solver.html

[17] Grubešić N.: Završni rad, 2012

[18] Pavlić I.: Statistička teorija i primjena, Tehnička knjiga Zagreb, Zagreb, 2005

[19] http://matematika.fkit.hr/staro/izborna/referati/Daniela%20Petkovicek%20-%20Linearno%20programiranje.pdf

[20] Štefanić N.: Skripte s predavanja, 2013

Za statističku analizu i grafičke prikaze korišten je MS Excel 2010.

PRILOG

I. CD-R disc