

## **PLANIRANJE EKSPERIMENTA METODOM ANOVA U PROGRAMSKOM JEZIKU R**

Damir Hodžić

Univerzitet u Bihaću, Tehnički fakultet, hodzicdamir@yahoo.com

**Ključne riječi: Eksperiment, ANOVA, varijansa, programski jezik R**

### **SAŽETAK:**

*U radu će se opisati planiranje eksperimenta metodom ANOVA. Ova metoda podrazumjeva analizu varijanse zadanih grupa podataka. ANOVA predstavlja jednu od najviše korištenih metoda kojom se analizira neki eksperiment, odnosno eksperimentalni podaci. Ti podaci mogu biti inženjerski ali i iz bilo kojeg drugog naučnog polja. Programska jezika R predstavlja programsko okruženje za statističke proračune i grafiku. Ima veoma široku primjenu kako u istraživanju tako i u praksi. U radu će se prikazati njegova primjena kod jednosmjerne ANOVA metode.*

### **1. UVOD**

Pod pojmom eksperimenta obično se podrazumjeva niz radnji ili postupaka da bi se došlo do određenih zaključaka. Planiranje eksperimenta je kao disciplina usko povezano sa teorijom uzorka. Teorija uzorka ocjenjuje karakteristike populacije na osnovu posmatranja jednog dijela te populacije bez nekih promjena populacije dok planiranje eksperimenta djeluje na jedan dio populacije kako bi se ocijenio efekat tog djelovanja na cijelu populaciju.

Jedan eksperiment može imati različita značenja. Tako u inženjerskim analizama eksperiment može predstavljati pripremanje komada nekog materijala i mjerjenje karakteristika tog komada. Može međutim značiti i proučavanje mjeseca slanjem mjesecnih snimaka ili proučavanje ozonskog omotača Zemlje. Isto tako eksperiment može značiti i proučavanje efekata nekih lijekova pri liječenju određenih bolesti kod ljudi. Kod svih ovih izučavanja pretpostavka je da je posmatrani dio populacije ili podataka reprezentativan.

Ako planiranje eksperimenta posmatramo sa inženjerskog aspekta onda se može reći da je posmatranje sistema ili procesa za vrijeme njegovog rada značajan dio izučavanja procesa i da je sastavni dio razumjevanja i učenja kako sistemi i procesi rade. Međutim da bi razumjeli šta se dešava u nekom procesu kada se mijenjaju ulazni faktori potrebno je da uradimo i neke druge aktivnosti osim posmatranja. To znači da je za potpuno razumjevanje uzročno posljedične veze u nekom sistemu neophodno namjerno mijenjati ulazne faktore i posmatrati i razumjeti promjene u izlazu iz sistema te razumjeti vezu između izlaznih i ulaznih faktora. Drugim riječima, potrebno je izvršiti eksperiment na tom sistemu.

Planiranje eksperimenta (engl. Design Of Experiment – DOE) predstavlja moćan instrument za dostizanje značajnih poboljšanja kvaliteta proizvoda i efikasnosti procesa. Za planiranje eksperimenta se može reći i da je metodologija primjene statistike u procesu izvođenja eksperimenta. Kao takva disciplina nije samo inženjerski problem nego se koristi u različitim sferama društva.

Osnovni dijelovi planiranja eksperimenta mogu se definisati na slijedeći način, bez obzira u kojim se naučnim disciplinama koristi:

- a) definisanje problema – potrebno je da se odredi koji je to zadatak našeg eksperimenta, koji materijal imamo i koji se postupci ili tretmani upoređuju.
- b) podjela eksperimentalnog materijala na sastavne dijelove – podrazumjeva eksperimentalne jedinice na koje se primjenjuje određeni tretman
- c) pridruživanje tretmana – potrebno je odrediti kako jedinicama pridružiti tretmane pri čemu se na svaku jedinicu može primjeniti samo jedan od tretmana.
- d) provođenje eksperimenta – eksperiment se prema ranije definisanim koracima uradi do kraja
- e) analiza i zaključci – dobivene rezultate potrebno je analizirati i na osnovu njih izvesti zaključke te razmotriti kako se dobiveni zaključci mogu proširiti na cijelu populaciju.

U nekom statističkom eksperimentu posmatra se uticaj jednog ili više faktora na određenu pojavu koja se može nazvati obilježje i predstavlja slučajnu promjenljivu. Ti faktori se zovu i kontrolisani faktori. Na posmatranu pojavu mogu uticati i slučajni faktori koji se zovu nekontrolisani ili eksperimentalne greške. Da bi iz nekog eksperimenta izveli konkretnе zaključke potrebno je izvođenje eksperimenta na određen način planirati i pri tome ispitati uslove pod kojima je moguće smanjiti eksperimentalnu grešku.

Dobro planiran eksperiment omogućava dobivanje jasnih interpretacija i izbjegavanje komplikovanih analiza a loše planiran eksperiment daje nam pogrešne zaključke nekog procesa.

## 2. ANALIZA VARIJANSE

Analiza varijanse predstavlja jednu od osnovnih i najvažnijih metoda planiranja eksperimenta koja se koristi prilikom analiza u različitim poljima društvenih i tehničkih nauka. U praksi i literaturi se ova metoda nalazi pod skraćenicom ANOVA. Ona predstavlja statističku metodu koja se koristi za upoređivanje srednjih vrijednosti dvije ili više grupa podataka. Pojam varijansa ili disperzija je pojam iz teorije vjerovatnoće i statistike i predstavlja matematičko očekivanje odstupanja slučajne promjenljive od njene srednje vrijednosti. U istraživačkoj praksi često se utvrđuje razlika između dvije, tri ili više statističkih serija. Te razlike se mogu posmatrati u odnosu na samo jedan ili više kriterija. Metodom ANOVA potrebno je dokazati da li je varijansa između grupa podataka veća od varijanse unutar tih istih grupa, odnosno postoje li razlike između nekoliko aritmetičkih sredina i da li su te razlike statistički značajne ili slučajne. Odnos varijanse između grupa i varijanse unutar grupa testira se pomoću F (Fišerovog) testa.

Osnovne veličine u ANOVI su faktori i nivoi. Najjednostavnije rečeno faktor može biti spol a nivoi su muški i ženski.

Da bi metodu ANOVA mogli uzeti kao mjerodavnu za analizu određenih rezultata potrebno je da se definišu određene pretpostavke:

- a) distribucija srednjih vrijednosti uzoraka je u obliku normalne distribucije (Gaussova krivulja). Smisao normalne raspodjele je u tome da stvari kad se mijenjaju nastoje ostati u blizini tačke prosjeka, pa se raspoređuju oko tog prosjeka po glatkoj, zvonolikoj krivulji,
- b) greške između pojedinih slučajeva su nezavisne jedna od drugih,
- c) podaci koji značajno odstupaju se uklanjuju iz analize,
- d) homogenost varijanse.

U zavisnosti od broja faktora u analizi eksperimenta kao i od nivoa svakog od faktora razlikujemo više tipova ANOVE. Osnovna a time i najjednostavnija analiza varijanse je jednosmjerna (One Way) ANOVA. Kod ove analize imamo jedan faktor sa najmanje dva nivoa pri čemu su nivoi nezavisni. Primjer za ovu analizu može biti slučaj kada na jednoj mašini izrađujemo cilindrične elemente sa

različitim prečnicima (npr. 30, 50 i 80 mm) i mjerimo hrapavost površine u [ $\mu\text{m}$ ] svakog elementa. Uzet ćemo da od svakog prečnika izradimo po 5 uzoraka (Tabela 1).

Tabela 1. Prikaz podataka za jednosmjernu ANOVU

30 mm	50 mm	80 mm
0,7	0,5	0,8
0,8	0,6	0,7
0,6	0,6	0,5
0,8	0,7	0,7
0,8	0,6	0,6

U primjeru jednosmjerne ANOVE faktor je prečnik cilindričnog elementa a nivoi su 30, 50 i 80 mm.

Ukoliko uzmem da na jednoj mašini izrađujemo cilindrične elemente jednakog prečnika u tri smjene i da iz svake smjene uzmem određeni broj uzoraka i izmjerimo hrapavost površine dobit ćemo ponavljače (Reapeted Measures) ANOVU (tabela 2).

Tabela 2. Prikaz podataka za ponavljače ANOVU

1.smjena	2.smjena	3.smjena
0,7	0,5	0,8
0,8	0,6	0,7
0,6	0,6	0,5
0,8	0,7	0,7
0,8	0,6	0,6

U ovom slučaju faktor je smjena, a imamo tri nivoa, 1., 2. i 3. smjena. Nivoi su zavisni jer mjerimo hrapavost na cilindričnim elementima jednakog prečnika.

Višefaktorsku (Factorial) ANOVU dobit ćemo ako imamo dvije mašine na kojima izrađujemo cilindrične elemente i na njima mjerimo hrapavost (tabela 3).

Tabela 3. Prikaz podataka za višefaktorsku ANOVU

	1.smjena	2.smjena	3.smjena
Mašina 1	0,7	0,5	0,8
	0,8	0,6	0,7
	0,6	0,6	0,5
Mašina 2	0,8	0,7	0,7
	0,8	0,6	0,6
	0,7	0,5	0,8

U prikazanom slučaju višefaktorske ANOVE imamo dva faktora, jedan je smjena a drugi je mašina. Svaki od faktora ima više nivoa, mašina ima dva nivoa a smjena ima tri nivoa.

U ovom radu prikazat će se analiza jednosmjerne ANOVE za primjer iz tabele 1. Postoji nekoliko koraka na osnovu kojih se radi analiza za sve metode ANOVE a to su:

1. definisanje nulte i alternativne hipoteze

$H_0 : \mu_{30mm} = \mu_{50mm} = \mu_{80mm}$  – nulta hipoteza, grupe su jednakе (postoji zavisnost)

$H_1$  – alternativna hipoteza, grupe nisu jednakе (ne postoji zavisnost)

2. definisanje nivoa signifikantnosti ( $\alpha$ ) - nivo signifikantnosti bira istraživač. To može biti bilo koja vrijednost u intervalu 0-100%, ali je uobičajeno da se bira nivo signifikantnosti od 5% ili 1%, između ostalog i zato što su za njih izračunate vrijednosti u statističkim tablicama. U našem primjeru uzet ćemo  $\alpha = 0,01$ .

3. računanje stepeni slobode (Degrees of Freedom) - dF

$$dF_{izmedu} = a - 1 = 3 - 1 = 2 \quad a = 3 \text{ (broj nivoa), } N = 15 \text{ (broj uzoraka)}$$

$$dF_{unutar} = N - a = 15 - 3 = 12$$

$$dF_{ukupno} = N - 1 = 15 - 1 = 14$$

4. određivanje F tabelarno – iz tabele Fišerove raspodjele za  $\alpha = 0,01$  odredi se vrijednost F za vrijednosti stepeni slobode unutar i između grupa izračunatih u prethodnom koraku (2,12),  $F_t=6,92$ .

5. statističko određivanje računske vrijednosti F i upoređivanje sa tabelarnim F.

Tabela 4. Tabela za određivanje statističkih vrijednosti jednosmjerne ANOVE

	SS	dF	MS
Između	0,049	2	0,0246
Unutar	0,104	12	0,0086
Ukupno	6,153	14	

U tabeli 4 veličina SS predstavlja sumu kvadrata (Sum of Squares). Suma kvadrata određuje se prema slijedećim formulama:

$$SS_{izmedju} = \frac{\Sigma(\sum a_i)}{n} - \frac{T^2}{N} = \frac{3,7^2 + 3^2 + 3,3^2}{5} - \frac{10^2}{15} = 0,049$$

$$SS_{unutar} = \Sigma Y^2 - \frac{\Sigma(\sum a_i)}{n} = (0,7^2 + 0,8^2 + \dots + 0,6^2) - \frac{3,7^2 + 3^2 + 3,3^2}{5} = 0,104$$

Tabela 5. Ulazni podaci i veličine koje se na osnovu njih računaju

	30 mm	50 mm	80 mm
$n = 5$	0,7	0,5	0,8
	0,8	0,6	0,7
	0,6	0,6	0,5
	0,8	0,7	0,7
	0,8	0,6	0,6
	$a_1=3,7$	$a_2=3$	$a_3=3,3$
	$T=3,7 + 3 + 3,3 = 10$		

Veličina MS u tabeli 4 predstavlja srednju vrijednost kvadrata (Mean Squares) i dobije se kada podijelimo sumu kvadrata i stepen slobode (MS = SS/dF).

Računska vrijednost F dobije se iz slijedećeg izraza:

$$F_r = \frac{MS_{izm}}{MS_{unu}} = \frac{0,0246}{0,0086} = 2,86$$

6. zaključak – upoređivanjem vrijednosti tabelarnog i računskog F vidimo da je računski manji od tabelarnog i u tom slučaju se prihvata nulta hipoteza.

### 3. METODA ANOVA U PROGRAMSKOM JEZIKU R

R je programski jezik i okruženje za statističke proračune i vizualizaciju. Slobodan je programski jezik, otvorenog je koda i pruža širok izbor statističkih metoda za linearno i nelinearno modeliranje, klasične statističke testove i dr.

U ovom radu prikazat će se kako se programskim jezikom R jednostavno može riješiti eksperiment jednosmjernom ANOVOM.

```
# Analiza varijanse (ANOVA) primjer
# RIM 2017 - Tehnički fakultet Bihać
precnik30 <- c(0.7,0.8,0.6,0.8,0.8)
precnik50 <- c(0.5,0.6,0.6,0.7,0.6)
precnik80 <- c(0.8,0.7,0.5,0.7,0.6)

# kombiniranje podatak u jedinstveni podatkovni skup
Combined_Groups <- data.frame(cbind(precnik30, precnik50, precnik80))

Combined_Groups # tabelarni prikaz podataka
summary(Combined_Groups) # min, median, mean, max

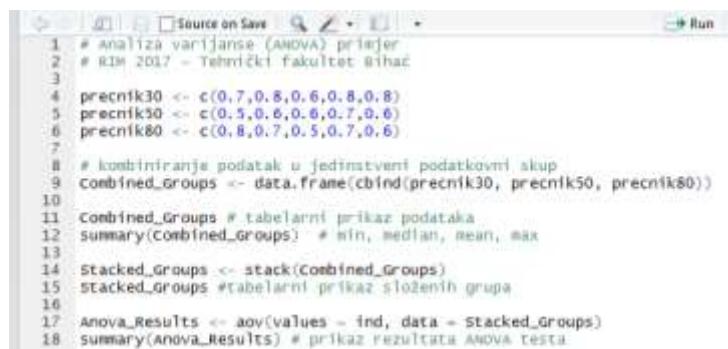
Stacked_Groups <- stack(Combined_Groups)
Stacked_Groups # tabelarni prikaz složenih grupa

Anova_Results <- aov(values ~ ind, data = Stacked_Groups)
summary(Anova_Results) # prikaz rezultata ANOVA testa
```

Slika 1. Prikaz programskog koda za primjer jednosmjerne ANOVE

U prozor za unošenje podataka direktno se ukucavaju vrijednosti hrapavosti površine za sve tri grupe. Nakon toga se grupe kombiniraju i slazu sa funkcijama Combined\_Groups i Stacked\_Groups (slika 1). Pozivanjem funkcije Anova\_Results dobivaju se sve neophodne vrijednosti koje smo u prethodnom poglavlju računali kao i računska vrijednost  $F$ .

U prozoru prikazanom na slici 2 mogu se i određeni komentari dodavati zbog lakšeg snalaženja ukoliko je problem složeniji.



The screenshot shows the RStudio interface with the code from Slika 1 pasted into the 'Code' tab. The code is identical to the one in Slika 1, starting with a comment '# Analiza varijanse (ANOVA) primjer' and ending with 'summary(Anova\_Results) # prikaz rezultata ANOVA testa'. The RStudio interface includes tabs for 'Code', 'Console', 'Output', 'Plots', and 'Help'. There are also buttons for 'Source on Save', 'Run', and other file operations.

Slika 2. Prozor u programskom jeziku R za unošenje podataka

Pored analize varijanse ispravnim unošenjem podatka i funkcija kako je prikazano na slici 2 dobit ćemo i osnovne statističke veličine za svaku grupu (slika 3).

```

prečnik30    prečnik50    prečnik80
Min. :0.80  Min. :0.5   Min. :0.50
1st Qu.:0.70 1st Qu.:0.6   1st Qu.:0.60
Median :0.80 Median :0.6   Median :0.70
Mean   :0.74 Mean   :0.6   Mean   :0.66
3rd Qu.:0.80 3rd Qu.:0.6   3rd Qu.:0.70
Max.  :0.80 Max.  :0.7   Max.  :0.80

> Stacked_Groups <- stack(combined_Groups)
> Stacked_Groups #tabelarni prikaz složenih grupa
  values   ind
1 0.7 prečnik30
2 0.8 prečnik30
3 0.6 prečnik30
4 0.8 prečnik30
5 0.8 prečnik30
6 0.5 prečnik50
7 0.6 prečnik50
8 0.8 prečnik50
9 0.7 prečnik50
10 0.6 prečnik50
11 0.8 prečnik80
12 0.7 prečnik80
13 0.5 prečnik80
14 0.7 prečnik80
15 0.6 prečnik80

> Anova_Results <- aov(values ~ ind, data = Stacked_Groups)
> summary(Anova_Results) # prikaz rezultata ANOVA testa
             Df Sum Sq Mean Sq F value Pr(>F)
ind          2 0.04933 0.024667  2.846 0.0974 .
Residuals 12 0.10400 0.008667

```

Slika 2. Prozor u programskom jeziku R sa izračunatim vrijednostima

Sa slikama 2 i 3 se vidi da vrlo lako i brzo dolazimo do konačnih rješenja za zadani problem. Naravno da postoje i drugi programski jezici pomoću kojih je moguće raditi slične analize ali prednost programskog jezika R je što je slobodan, odnosno može se besplatno skinuti sa interneta i koristiti, i što je jednostavniji za razumjevanje i korištenje od nekih drugih sličnih programskih jezika.

#### 4. ZAKLJUČAK

Planiranje eksperimenta predstavlja značajan aspekt inženjerskog rada i istraživanja. Da bi ispravno isplanirali neki eksperiment bitno je znati šta želimo sa podacima koje dobijemo, koji su nam ulazni faktori i koliko nivoa imaju. U teoriji i praksi postoji više metoda pomoću kojih se mogu obradivati eksperimentalni podaci u inženjerskim ali i društvenim ili medicinskim naukama. Analiza varijanse je metoda koja u svim segmentima eksperimentalnog planiranja predstavlja osnovnu metodu. Na primjeru jednosmjerne ANOVE mogli smo vidjeti na koji način se podaci iz različitih grupa upoređuju i kako se dolazi do zaključka o vezi između podataka u grupama. Ukoliko sami računamo ANOVU pomoću teorijskih formula moguće je doći do rješenja ali predstavlja značajan posao pogotovo ukoliko imamo više faktora i više nivoa. Programski jezik R nam olakšava taj posao, gdje samo u nekoliko koraka, unošenjem podataka iz različitih grupa te kombiniranjem i slaganjem grupa dolazimo do svih mogućih podataka i zaključka za eksperimentalnu analizu.

#### 5. LITERATURA

- [1] Montgomery D.C.: *Design and Analysis of Experiments (8<sup>th</sup> Edition)*, Arizona State University, John Wiley and Sons, 2015.
- [2] Petrović Lj.: *Teorija uzorka i planiranje eksperimenta*, Ekonomski fakultet Beograd, 2013.
- [3] Gerrard P., Johnson R.M.: *Mastering Scientific Computing with R*, Packt Publishing, Birmingham-Mumbai, 2015.
- [4] Kerns G.J.: *Introduction to Probability and Statistics using R*, GNU Free Documentation, 2012.
- [5] <http://statisticslectures.com/topics/introanova/>
- [6] <http://statisticslectures.com/topics/onewayanova/>