

# Modeli analize varijance (ANOVA)

Josip Arnerić<sup>1</sup>, Kristijan Protrka<sup>2</sup>

## Sažetak

Cilj ovog članka je dati pregled modela analize varijance koji su najčešće zastupljeni u ekonomskim analizama. Analiza varijance, skraćeno ANOVA (engl. *ANalysis Of VAriance*), se upotrebljava kada se želi testirati postoji li statistički značajna razlika između aritmetičkih sredina više populacija (osnovnih skupova) ili tzv. grupa te se donosi zaključak pripadaju li uzorci (grupe) istoj populaciji. Najčešće je u primjeni analiza varijance s jednim promjenjivim faktorom (engl. *One-Way ANOVA*) i analiza varijance s dva promjenjiva faktora (engl. *Two-Way ANOVA*). Svaki od navedenih modela je specijalni slučaj poopćenog linearnog modela GLM (engl. *generalized linear model*) te svaki od njih zahtijeva određene pretpostavke koje trebaju biti ispunjene. U slučajevima kada se analizira dva ili više faktora, početni se model može proširiti za članove interakcije. Ono što je zajedničko svim modelima analize varijance jest da se značajnost utjecaja analiziranih faktora na promatranu slučajnu varijablu provodi empirijskim  $F$ -omjerom, tj.  $F$ -testom.

## Uvodni dio

ANOVA je metoda kojom se uspoređuju aritmetičke sredine više uzoraka te se na temelju  $F$ -testa donosi zaključak o (ne)postojanju značajnih razlika između sredina više populacija. Na taj se način analizira utjecaj jedne ili više kategorijalnih (nezavisnih) varijabli na jednu numeričku kontinuiranu (zavisnu) varijablu. Kategorijalne su varijable nominalne skale mjerenja i često se nazivaju faktorima pa se govori o jednofaktorskoj, dvofaktorskoj ili višefaktorskoj analizi varijance. Pri tome broj modaliteta kategorijalnih varijabli određuje broj grupa između kojih se testiraju razlike. Naime, radi se o usporedbi većeg broja prosječnih vrijednosti numeričke (zavisne) varijable  $Y$  odjednom. U skladu s time nultom se hipotezom pretpostavlja da su prosjeci svih populacija jednaki, a alternativnom hipotezom da nisu svi jednaki. S obzirom na broj analiziranih faktora može se testirati više hipoteza, tj. skup hipoteza. Nakon što se definiraju statističke hipoteze (nulta i alternativna) provodi se postupak testiranja, tj. izračunava se odgovarajuća testna veličina.

## Analiza varijance s jednim faktorom

Kod analize varijance s jednim promjenjivim faktorom cilj je ispitati odnos varijabilnosti podataka između uzoraka (grupa) s varijabilnosti podataka unutar uzoraka (grupa). Broj uzoraka određen je brojem grupa, tj. brojem modaliteta ili razina promatranog faktora. Odnos navedenih vrsta varijabilnosti naziva se empirijski  $F$ -omjer

<sup>1</sup> Autor je s Ekonomskog fakulteta u Zagrebu; e-pošta: jarneric@efzg.hr

<sup>2</sup> Autor je s Ekonomskog fakulteta u Splitu; e-pošta: kristijan.protrka@efst.hr

koji pripada Fisherovoj, tzv.  $F$ -distribuciji. Ako je on statistički značajan tada se može odbaciti nulta hipoteza da se aritmetičke sredine između grupa statistički značajno ne razlikuju. Drugim riječima, nultom hipotezom se pretpostavlja da određeni faktor značajno ne djeluje na promatranu slučajnu varijablu  $Y$ , dok se alternativnom hipotezom pretpostavlja da određeni faktor statistički značajno djeluje na promatranu slučajnu varijablu  $Y$ .

Korištenjem analize varijance moguće je raščlaniti i procijeniti varijabilnosti uvjetovane različitim faktorima. Pretpostavke ove analize nalažu da opažanja varijable  $Y$  moraju biti nezavisna i normalno distribuirana unutar svake grupe s jednakim varijancama (uvjet homogenosti). Pretpostavka da su opažanja  $y_{ij}$  nezavisna, za svaki  $j$ , znači da se radi o nezavisnim uzorcima.

U postupku primjene jednofaktorske analize varijance prvo treba izračunati aritmetičke sredine  $\bar{y}_j$  za svaku od  $k$  grupa (uzoraka) veličine  $n_j$ ,  $j = 1, 2, \dots, k$ , a zatim zajedničku aritmetičku sredinu  $\bar{y}$  svih grupa zajedno. Nakon toga se definiraju i izračunavaju zbrojevi kvadrata odstupanja  $SS$  (engl. *Sum of Squares*):

$$SSB = \sum_{j=1}^k n_j(\bar{y}_j - \bar{y})^2 - \text{zbroj kvadrata odstupanja između grupa,}$$

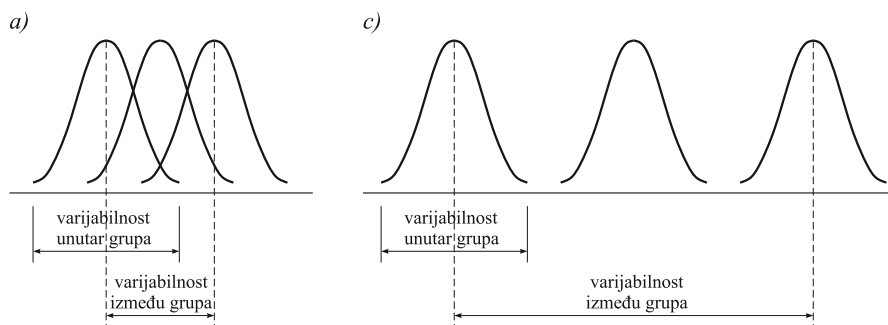
$$SSW = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 - \text{zbroj kvadrata odstupanja unutar grupa,}$$

$$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 - \text{ukupan zbroj kvadrata odstupanja,}$$

pri čemu je  $y_{ij}$   $i$ -ta vrijednost zavisne varijable  $j$ -tog uzorka (grupe) i  $n = n_1 + n_2 + \dots + n_k$ . Budući da su uzorci nezavisni, ne moraju biti jednakih veličina.

Potom se svaki zbroj kvadrata odstupanja  $SS$  podijeli s odgovarajućim brojem stupnjeva slobode  $df$  (engl. *degrees of freedom*) da bi se dobile sredine kvadrata  $MS$  (engl. *Mean Squares*) i konačno empirijski  $F$ -omjer.

Dakle, jednadžba analize varijance polazi od dekompozicije ukupne varijabilnosti na varijabilnost između grupa (uzoraka) i varijabilnost unutar grupa (uzoraka), tj.  $SST = SSB + SSW$ . Za odbacivanje nulte hipoteze  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k = \mu$  poželjno je da je veći zbroj  $SSB$  u odnosu na zbroj  $SSW$  (slika 1-C). Na dijelu slike 1-A može se primijetiti suprotno ( $SSB < SSW$ ).



Slika 1. Usporedba  $SSB$  i  $SSW$  između triju normalno distribuiranih populacija.

Ako je  $SSB$  značajno veći u odnosu na  $SSW$  tada će se nulta hipoteza odbaciti. Da bi to statistički testirali, koristi se  $F$ -test. Pri tome se u svim računalnim programima provođenje  $F$ -testa svodi na formiranje tablice ANOVA (tablica 1).

Tablica 1. Jednofaktorska analiza varijance.

izvor varijabilnosti	$SS$	$df$	$MS$	$F$
između grupa	$SSB$	$k - 1$	$MSB = \frac{SSB}{k - 1}$	$\frac{MSB}{MSW}$
unutar grupa	$SSW$	$n - k$	$MSW = \frac{SSW}{n - k}$	
ukupno	$SST$	$n - 1$		

Varijabilnost unutar grupa  $SSW$  se pripisuje slučajnom djelovanju i predočuje komponentu pogreške (engl. *Error*) odnosno neprotumačeni dio varijabilnosti.

Empirijski  $F$ -omjer se može zapisati u sljedećem obliku:

$$F = \frac{MSB}{MSW} = \frac{\frac{SSB}{k - 1}}{\frac{SSW}{n - k}} = \frac{n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_1 - \bar{y})^2 + \dots + n_k(\bar{y}_k - \bar{y})^2}{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_k - 1)s_k^2}{n - k}}$$

pri čemu je  $s_j$  standardna devijacija  $j$ -te grupe.

Gore navedeni izraz omogućuje provedbu  $F$ -testa bez formiranja tablice ANOVA, ako su poznate veličine grupa (uzoraka), te aritmetičke sredine i standardne devijacije svake grupe. Nakon izračunavanja empirijskog  $F$ -omjera računa se vjerojatnost da će slučajna varijabla  $F$ -distribucije, s  $df_1 = k - 1$  stupnjeva slobode u brojniku te  $df_2 = n - k$  stupnjeva slobode u nazivniku, premašiti empirijski  $F$ -omjer. Navedena vjerojatnost je poznata kao  $p$ -vrijednost koja predočuje empirijsku razinu značajnosti testa.

U sljedećem primjeru analizira se utrošak radnog vremena po proizvodu u tri pogona koji proizvode isti proizvod:  $A$ ,  $B$  i  $C$ . Iz pogona  $A$  u slučajni uzorak izabrano je 8, iz pogona  $B$  7, te iz pogona  $C$  9 proizvoda. Utrošeno vrijeme u minutama za proizvode je dano u tablici.

Tablica 2. Utrošeno vrijeme za proizvodnju istog proizvoda u tri pogona  $A$ ,  $B$  i  $C$ .

$A$	$B$	$C$
27	19	24
28	24	21
25	23	21
25	26	22
20	19	23
26	18	27
23	20	18
22		20
		25

Može li se prihvatiti pretpostavka kako ne postoji razlika u prosječnom utrošku radnog vremena izrade proizvoda u pogonima  $A$ ,  $B$  i  $C$ ? Testiranje navedene pretpostavke se vrši na razini teorijske značajnosti od 5% ( $\alpha = 0.05$ ). Rezultati dobiveni iz ispisa programa STATISTICA su u tablici 3.

Tablica 3. Jednofaktorska analiza varijance iz ispisa programa STATISTICA

izvor varijabilnosti	SS	df	MS	F	p-vrijednost
između grupa	41.0714	2	20.53571	2.606865	0.09744
unutar grupa	165.4286	21	7.877551		
ukupno	206.5000	23			

Vjerojatnost da će slučajna varijabla  $F$ -distribucije, s  $df_1 = 2$  stupnjeva slobode u brojniku te  $df_2 = 21$  stupnjeva slobode u nazivniku, premašiti vrijednost 2.606865 jednaka je  $p$ -vrijednosti, tj. formalno zapisano  $P(F_{[2,21]} > 2.606865) = 0.09744$ .  $p$ -vrijednost se definira kao empirijska razina značajnosti. Usporedbom empirijske razine značajnosti ( $p$ -vrijednost) i teorijske razine značajnosti ( $\alpha$ ) donosi se odluka. Budući da je empirijska razina značajnosti veća od teorijske ( $0.09744 > 0.05$ ) nulta se hipoteza ne može odbaciti. Zaključuje se da ne postoji statistički značajna razlika u prosječnom utrošenom radnom vremenu za proizvodnju istog proizvoda između triju pogona  $A$ ,  $B$  i  $C$  na razini značajnosti  $\alpha = 0.05$ . S druge strane nulta bi se hipoteza mogla odbaciti na razini značajnosti od 10% ( $\alpha = 0.1$ ).

Alternativno se empirijski  $F$ -omjer mogao izračunati na temelju pokazatelja deskriptivne statistike (tablica 4).

Tablica 4. Pokazatelji deskriptivne statistike

grupa	broj jedinica	aritmetička sredina	standardna devijacija
A	8	24.5000	2.6726
B	7	21.2857	3.0394
C	9	22.3333	2.7386

Zajednička je ponderirana aritmetička sredina za sva tri pogona (grupe) zajedno:

$$\bar{y} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2 + n_3\bar{y}_3}{n_1 + n_2 + n_3} = \frac{8 \cdot 24.5000 + 7 \cdot 21.2857 + 9 \cdot 22.3333}{8 + 7 + 9} = 22.75.$$

Empirijski je  $F$ -omjer, jednak je:

$$F = \frac{n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2 + n_3(\bar{y}_3 - \bar{y})^2}{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2}{n - k}}$$

$$= \frac{8(24.5 - 22.75)^2 + 7(21.2857 - 22.75)^2 + 9(22.3333 - 22.75)^2}{\frac{(8 - 1)2.6726.^2 + (7 - 1)3.0394.^2 + (9 - 1)2.7386.^2}{24 - 3}} = 2.6069.$$

Drugi je pristup provedbe analize varijance pomoću procjene poopćenog linearnog modela GLM. Specijalni slučaj općeg linearnog modela je jednačba jednostruke i/ili višestruke linearne regresije. U navedenom primjeru zavisna varijabla je  $Y$  = utrošeno vrijeme za proizvodnju istog proizvoda, dok je nezavisna varijabla  $X$  = pogon ( $A$ ,  $B$ ,  $C$ ). Način na koji se kategorijalna (nominalna) varijabla  $X$  može uključiti u regresijsku jednačbu su binarne varijable, tzv. *dummy* varijabli. Pod pojmom *dummy* varijabla podrazumijevamo umjetno konstruiranu varijablu, koja je rezultat postojanja ili nepostojanja nekog fenomena. Najčešće *dummy* varijabla poprima vrijednost 0 u slučaju

odsutnosti nekog atributa, a vrijednost 1 ako je atribut prisutan. Na taj se način u regresijsku jednadžbu mogu uključiti i kvalitativne varijable, npr. spol ( $M = 0$ ,  $\check{Z} = 1$ ), itd. Ako nominalna varijabla  $X$  ima 3 kategorije ili modaliteta (pogon A, B i C) tada je potrebno definirati dvije *dummy* varijable:

$$X_1 = \begin{cases} 1 & \text{za pogon A} \\ 0 & \text{inače} \end{cases} \quad X_2 = \begin{cases} 1 & \text{za pogon B} \\ 0 & \text{inače} \end{cases}$$

*Dummy* varijabla za pogon C je izostavljena, tj. u regresijsku jednadžbu s konstantnim članom uvijek je potrebno uključiti jednu *dummy* varijablu manje u odnosu na broj modaliteta kvalitativne varijable. Uključivanje *dummy* varijabli za svaki pogon bi rezultiralo sustavom normalnih jednadžbi koji nema rješenje. Jednadžba višestruke linearne regresije u konkretnom slučaju primjene je  $y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e_i$ , pri čemu su  $\beta_0$ ,  $\beta_1$  i  $\beta_2$  parametri koje treba procijeniti metodom najmanjih kvadrata (engl. *Least Squares Method*), dok je  $e_i$  član pogreške. Procjene pogrešaka nazivamo rezidualima, stoga se metodom najmanjih kvadrata minimizira zbroj kvadrata rezidualnih odstupanja. Za procjenu parametara višestruke regresije potrebno je podatke u Excel datoteci urediti na drugačiji način u odnosu na tablicu 2, tj. sva opažanja staviti u retke, a sve varijable u stupce (slika 2). Nakon toga se podaci iz Excel datoteke importiraju u program STATISTICA te se metodom najmanjih kvadrata dobiju procjene parametara u jednadžbi višestruke regresije (tablica 5).

	A	B	C	D
1	<b>Pogon</b>	<b>Y</b>	<b>X1</b>	<b>X2</b>
2	A	27	1	0
3	A	28	1	0
4	A	25	1	0
5	A	25	1	0
6	A	20	1	0
7	A	26	1	0
8	A	23	1	0
9	A	22	1	0
10	B	19	0	1
11	B	24	0	1
12	B	23	0	1
13	B	26	0	1
14	B	19	0	1
15	B	18	0	1
16	B	20	0	1
17	C	24	0	0
18	C	21	0	0
19	C	21	0	0
20	C	22	0	0
21	C	23	0	0
22	C	27	0	0
23	C	18	0	0
24	C	20	0	0
25	C	25	0	0

Slika 2. Matrica podataka u Excelu.

Tablica 5. Rezultati procjene jednadžbe višestruke regresije (ispis iz programa STATISTICA).

	koeficijent	standardna pogreška	t-test	p
konstanta	22.3333	0.9356	23.8715	0.0000
$X_1$	2.1667	1.3638	1.5887	0.1271
$X_2$	-1.0476	1.4144	-0.7407	0.4671

#### ANOVA

izvor	SS	df	MS	F-test	p
regresija	41.0714	2	20.5357	2.6069	0.0974
residual	165.4286	21	7.8776		
ukupno	206.5	23			

Na temelju procjenjenih parametara (koeficijenata) u tablici 5 može se zapisati regresijska jednadžba  $\bar{y}_j = 22.3333 + 2.1667x_1 - 1.0476x_2$ . Dobivenom jednadžbom opisuju se sredine utrošenog vremena proizvodnje istog proizvoda s obzirom na tri vrste pogona ( $A$ ,  $B$  i  $C$ ). Prosječna se vrijednost utrošenog vremena proizvodnje u pogonu  $A$  dobije uvrštavanjem vrijednosti  $x_1 = 1$  i  $x_2 = 0$ , tj.  $\bar{y}_1 = 22.3333 + 2.1667 = 24.5$ . Nadalje, prosječna se vrijednost utrošenog vremena proizvodnje u pogonu  $B$  dobije uvrštavanjem vrijednosti  $x_1 = 0$  i  $x_2 = 1$ , tj.  $\bar{y}_2 = 22.3333 - 1.0476 = 21.2857$ , dok je prosječna vrijednost utrošenog vremena proizvodnje u pogonu  $C$  jednaka konstantnom članu  $\bar{y}_3 = 22.3333$  jer je u tom slučaju  $x_1 = x_2 = 0$ . U kontekstu višestruke regresije, nulta se hipoteza o jednakosti sredina triju populacija može formalno zapisati kao  $H_0 : \beta_1 = \beta_2 = 0$ . Test kojim se testira navedena nulta hipoteza je također  $F$ -test. U tablici 5 se vidi da je empirijski  $F$ -omjer jednak 2.6069 što odgovara empirijskom  $F$ -omjeru iz tablice 3. Iz dosadašnjih razmatranja se može zaključiti da se model višestruke regresije može upotrijebiti kao model jednofaktorske analize varijance. Očito je da se model ANOVA i model višestruke regresije temelje na istim pretpostavkama koje je potrebno ispuniti.

Nadalje, kod analize varijance s jednim promjenjivim faktorom provode se *post hoc* testovi. Među testovima koji zahtijevaju jednake varijance najpoznatiji su sljedeći: Tukey, Scheffe, Gabriel, Bonferroni, Sidak, Hochberg, LSD, S-N-K i Duncan. Kod testova koji ne zahtijevaju jednake varijance najpoznatiji su sljedeći: Tamhane T2, Dunnett T3, Games-Howell i Dunnett C test. Navedeni testovi se neće ovdje detaljno razmatrati.

## Analiza varijance s dva faktora

Kod analize varijance s dva promjenjiva faktora ispituje se djelovanje promjenjivih faktora  $X$  i  $Z$  na numeričku vrijednost slučajne varijable  $Y$ . Uvjet za ovo testiranje je da uzorci potječu iz normalno distribuiranih populacija s jednakim varijancama. To znači da pretpostavke koje bi trebale biti ispunjene prigodom testiranja razlika u aritmetičkim sredinama vrijede kao i kod analize varijance s jednim promjenjivim faktorom. Analiza varijance s dva promjenjiva faktora također se svodi na testiranje skupa nulnih hipoteza o jednakosti aritmetičkih sredina više osnovnih skupova, ali sada po dva različita faktora koji djeluju na istu numeričku varijablu  $Y$ . Pri tome se u model dvofaktorske ANOVA-e dodatno može uključiti i efekt interakcije. Efekt interakcije se predočuje kao član umnoška  $X \cdot Z$ . Dvofaktorska analiza varijance s efektom interakcije je dana u tablici 6.

Hipoteza o jednakosti sredina svih  $k_1k_2$  populacija testira se pomoću 3  $F$ -testa. Zato su u tablici 6 predočena tri  $F$ -omjera. Svaki od  $F$ -omjera je kvocijent sredine kvadrata odstupanja pripadajućeg faktora i neprotumačenog zbroja kvadrata odstupanja (ostatak). Dvofaktorska analiza varijance s efektom interakcije također se može predočiti kao opći linearni model GLM. Na primjer ako se varijabla  $X$  pojavljuje s tri modaliteta ( $k_1 = 3$ ), te varijabla  $Z$  pojavljuje s dva modaliteta ( $k_2 = 2$ ), tada se model višestruke regresije s efektom interakcije može zapisati kao:  $\bar{y}_{ij} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3z + \beta_4x_1z + \beta_5x_2z$ . Pri tome su uvedene dvije *dummy* varijable  $x_1$  i  $x_2$  za opisivanje prvog faktora (varijable  $X$ ), jedna *dummy* varijabla  $z$  za opisivanje drugog faktora (varijable  $Z$ ) te članovi umnoška između *dummy* varijabli  $x_1z$  i  $x_2z$  za opisivanje efekta interakcije između dvaju faktora.

Tablica 6. Dvofaktorska analiza varijance s efektom interakcije.

izvor varijabilnosti	SS	df	MS	F-omjer
faktor 1 (varijabla X)	SSB <sub>1</sub>	k <sub>1</sub> - 1	$MSB_1 = \frac{SSB_1}{k_1 - 1}$	$F_1 = \frac{MSB_1}{MSW}$
faktor 2 (varijabla Z)	SSB <sub>2</sub>	k <sub>2</sub> - 1	$MSB_2 = \frac{SSB_2}{k_2 - 1}$	$F_2 = \frac{MSB_2}{MSW}$
faktor 1 * faktor 2 (varijabla X * Z)	SSB <sub>12</sub>	(k <sub>1</sub> - 1)(k <sub>2</sub> - 1)	$MSB_{12} = \frac{SSB_{12}}{(k_1 - 1)(k_2 - 1)}$	$F_3 = \frac{MSB_{12}}{MSW}$
pogreška (ostatak)	SSW	n - k <sub>1</sub> k <sub>2</sub>	$MSW = \frac{SSW}{n - k_1 k_2}$	
ukupno	SST	n - 1		

U skladu s navedenim mogu se testirati tri skupa nultih hipoteza pomoću tri F-testa:

(a)  $H_0 : \beta_1 = \beta_2 = 0 \rightarrow F_1$

(b)  $H_0 : \beta_3 = 0 \rightarrow F_2$

(c)  $H_0 : \beta_4 = \beta_5 = 0 \rightarrow F_3$

Prvom se nultom hipotezom testira značajnost djelovanja prvog faktora (varijable X) na numeričku slučajnu varijablu Y. Drugom se nultom hipotezom testira značajnost djelovanja drugog faktora (varijable Z) na numeričku slučajnu varijablu Y, dok se trećom nultom hipotezom testira značajnost interaktivnog efekta između varijabli X i Z na numeričku slučajnu varijablu Y.

Podaci iz tablice 2 mogu se dodatno analizirati s obzirom na dva dobavljača sirovine koja je potrebna za proizvodnju promatranog proizvoda (tablica 7). Vjeruje se da što je sirovina kvalitetnija prosječno utrošeno vrijeme za proizvodnju proizvoda je manje. Pri tome je sirovina dobavljača X kvalitetnija od dobavljača Z, ali je zato sirovina dobavljača Z nešto jeftinija.

Tablica 7. Utrošeno vrijeme za proizvodnju istog proizvoda u tri pogona A, B i C s obzirom na kvalitetu sirovine izabranog dobavljača X ili Z.

dobavljač	A	B	C
Z	23	23	23
	25	24	24
	25	26	25
	26		27
	27		
	28		
X	20	18	18
	22	19	20
		19	21
		20	21
			22

Ispis dvofaktorske analize varijance s efektom interakcije iz programa STATISTICA je dan u tablici 8.

Tablica 8. Dvofaktorska analiza varijance s efektom interakcije iz ispisa programa STATISTICA.

izvor varijabilnosti	SS	df	MS	F	p
pogon	8.98	2	4.49	1.926	0.174559
dobavljač	121.13	1	121.13	51.975	0.000001
pogon * dobavljač	0.95	2	0.47	0.204	0.817596
ostatak	41.95	18	2.33		
ukupno	206.50	23			

Na temelju dobivenih rezultata iz tablice 8 se zaključuje kako jedino izbor dobavljača, tj. kvaliteta sirovine značajno utječe na utrošeno vrijeme proizvodnje promatranog proizvoda na razini teorijske značajnosti od 5% ( $p$ -vrijednost  $< \alpha$ ). Ako se nabavlja kvalitetnija sirovina (od dobavljača  $X$ ) prosječno vrijeme utrošeno za proizvodnju proizvoda je 20 sati, a ako se nabavlja manje kvalitetna sirovina (od dobavljača  $Z$ ) prosječno vrijeme utrošeno za proizvodnju proizvoda je 25.08 sati.

## Zaključak

U ekonomskim se primjenama analiza varijance upotrebljava kada se želi testirati postoji li statistički značajna razlika između aritmetičkih sredina više populacija (osnovnih skupova) ili tzv. grupa te se donosi zaključak pripadaju li uzorci (grupe) istoj populaciji. Najčešće je u primjeni analiza varijance (ANOVA) s jednim promjenjivim faktorom i analiza varijance s dva promjenjiva faktora. ANOVA spada u zavisne tehnike multivarijatne analize, što znači da je *a priori* definirana uzročnost među varijablama. Pri tome je zavisna varijabla numerička, dok su nezavisne varijable kategorijalne ili nominalne. Kod analize varijance s dva ili više faktora (kategorijalnih varijabli) može se dodatno testirati značajnost interakcijskog efekta koji se predočuje članom umnoška promatranih faktora. U skladu s tim može se testirati više skupova hipoteza: onaj kojim se testiraju glavni efekti svakog faktora pojedinačno i onaj kojim se testiraju interakcijski efekti. Svaki od navedenih modela analize varijance rezultira odgovarajućim  $F$ -omjerom, tj. omjerom protumačenog dijela varijabilnosti i ostatka. Također, svaki od modela se može predočiti kao specijalni slučaj poopćenog linearnog modela GLM koristeći binarne, tzv. *dummy* varijable.

## Literatura i izvori

- [1] A. HALMI, *Multivarijatna analiza u društvenim znanostima*, Alinea, Zagreb, 2003.
- [2] S. PIVAC, B. ŠEGO, *Statistika*, Alka script d.o.o., Zagreb, 2014.
- [3] A. ROZGA, *Statistika za ekonomiste*, Ekonomski fakultet Split, 2003.
- [4] A. ROZGA, B. GRČIĆ, *Poslovna statistika*, Ekonomski fakultet Split, 2009.
- [5] I. ŠOŠIĆ, *Primijenjena statistika*, Školska knjiga Zagreb, Zagreb, 2004.
- [6] [www.statsoft.com](http://www.statsoft.com)