

Analiza varijanse



PREDAVANJE BR.4

Osnovne ideje

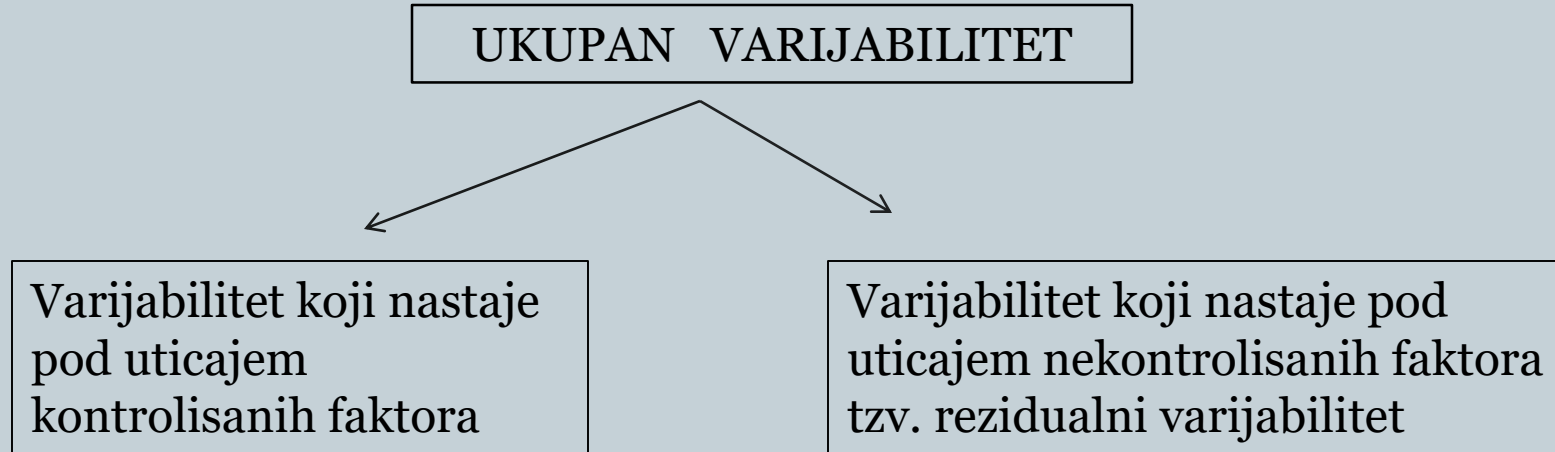


- Osnovne ideje i doprinose ANOVA dao je Fisher
- Kako testirati hipotezu o jednakosti aritmetičkih sredina više od dvije populacije?
- Zašto naziv analiza VARIJANSE kad se testiraju aritmetičke sredine?
- Nije korektno sprovesti testiranje Studentovim t testom sukcesivnom primjenom na različitim parovima aritmetičkih sredina
- Više razloga: veliki broj ponavljanja, ali i povećavanje rizika greške prve vrste.
- Cilj istraživanje uticaja jednog ili više faktora na varijabilitet određene pojave

Sušтина ANOVA



- Razlaganje ukupnog varijableta posmatrane pojave na sastavne komponente (izvore)



- Pravilo je da se ispitivani varijabilitet smatra značajnim ukoliko varijabilitet koji on prouzrokuje značajno premašuje rezidualni varijabilitet.

ANOVA sa jednim faktorom



- Razlaganje ukupnog varijabiliteta:

$$SK_{tot} = SK_{un} + SK_{iz}$$

$$SK_{tot} = \sum (Y - M_{tot})^2$$

$$SK_{un} = \sum (Y_g - M_g)^2$$

$$SK_{iz} = \sum n_g (M_g - M_{tot})^2$$

- Formula za skraćeno računanje varijanse

$$SD^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N - 1}$$

ANOVA sa jednim faktorom



- Uticaj jednog faktora (sa više nivoa/tretmana) na varijabilitet posmatrane pojave
- Model analize varijanse sa jednim faktorom

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

$$\alpha_i = \mu_i - \mu$$

- Efekat i-tog tretmana α_i
- Zajednička ar.sredina μ
- Slučajna greška ε_{ij}
- Opservacija izabrana iz populacije X_{ij}

ANOVA sa jednim faktorom



- Suma kvadrata unutar i između grupa nije dovoljna za ocjenu varijabiliteta
- Prava mjera varijabiliteta je varijansa i dobijamo je kada sume kvadrata podijelimo sa odgovarajućim brojem stepeni slobode

Izvor varijabiliteta	Suma kvadrata odstupanja	Broj stepeni slobode	Ocjena varijanse
Između grupa	SK_{iz}	$SS_{iz} = \sum g - 1$	PK_{iz}
Unutar grupa	SK_{un}	$SS_{un} = N_{tot} - g$	PK_{un}
Ukupno	SK_{tot}	$N_{tot} - 1$	

ANOVA sa jednim faktorom



- Hipoteze:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu$$

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = 0$$

- H_1 : aritmetičke sredine bar dva skupa se među sobom razlikuju ili efekat bar jednog tretmana se razlikuje od nule

- Statistika F testa i Snedecorov F raspored

$$F = \frac{PK_{iz}}{PK_{un}}$$

$$F \succ F_{\alpha, \nu_1, \nu_2}$$

ANOVA: pretpostavke



1. Normalnost
2. Homogenost varijansi
3. Slučajne greške u prosjeku jednake nuli
4. Slučajne greške su međusobno nezavisne
5. Aditivnost

Preporuka: uzorci da budu iste veličine!

Neparametarska alternativa: Kruskal-Wallis-ov test

Primjer 1



Pretpostavimo da nas interesuje da li postoji statistički značajna razlika u rezultatima između 3 metoda nastave statistike, odnosno da li se metodi nastave međusobno razlikuju po uticaju na uspjeh studenata. Metode nastave klasifikovane su na sljedeći način: A1-standardni način rada, A2-studentima se unaprijed daje materijal za nastavu sa odgovarajućim problemima koji se kasnije raspravljaju na času, a A3-osim dobijenog materijala, studenti imaju priliku da razrađuju primjere na računarima. Slučajnim izborom, formirane su jednake grupe od po 5 studenata. Nastavu u sva tri slučaja izvodi isti Profesor. Rezultati pismenog ispita su dati u narednoj tabeli:

Metod nastave		
A1	A2	A3
35	38	42
25	40	39
38	36	45
26	35	38
36	31	46

Primjer 1



	Metod nastave			(Y-Mtot) ²			(Yg-Mg) ²			ng(Mg-Mtot) ²		
	A1	A2	A3	1	2	3	1	2	3	1	2	3
	35	38	42	2.78	1.78	28.44	9.00	4.00	0.00	108.89	2.22	142.22
	25	40	39	136.11	11.11	5.44	49.00	16.00	9.00			
	38	36	45	1.78	0.44	69.44	36.00	0.00	9.00			
	26	35	38	113.78	2.78	1.78	36.00	1.00	16.00			
	36	31	46	0.44	32.11	87.11	16.00	25.00	16.00			
Suma	160	180	210	254.89	48.22	192.22	146.00	46.00	50.00	108.89	2.22	142.22
Mg	32	36	42	495.333			242.000			253.333		
Mtot	36.67			SKtot			SKun			SKiz		
Ntot	15											
Ng	5											
g	3											
				Izvor var	SK	SS	Varijansa	F				
				Između	253.33	2	126.665	6.280909				
				Unutar grupa	242	12	20.1666667					
				Ukupno	495.33	14						

Primjer 1



$$SK_{TOT} = \sum \sum y_{ij}^2 - \frac{(\sum \sum y_{ij})^2}{N} = 20662 - \frac{(550)^2}{15} = 495,33$$

$$SS = 15 - 1 = 14$$

$$SK_{IZ} = \sum \frac{(\sum y_j)^2}{N_j} - \frac{(\sum \sum y_{ij})^2}{N} = \left(\frac{25600}{5} + \frac{32400}{5} + \frac{44100}{5} \right) - \frac{(550)^2}{15}$$

$$= \frac{102100}{5} - \frac{(550)^2}{15} = 253,33$$

$$SS = 3 - 1 = 2$$

$$SK_{UN} = \sum \sum y_{ij}^2 - \sum \frac{(\sum y_j)^2}{N_j} = 20662 - 20420 = 242$$

$$SK_{UN} = SK_{TOT} - SK_{IZ}$$

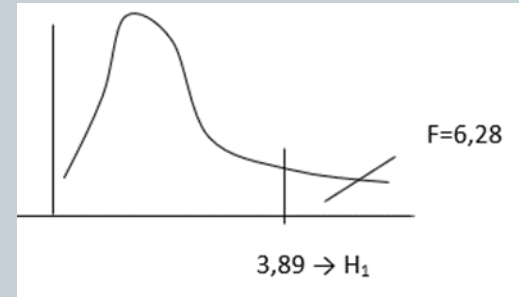
$$SS = N_{TOT} - g = 15 - 3 = 12$$

$$PK_{IZ} = \frac{SK_{IZ}}{2} = \frac{253,33}{2} = 126,67$$

$$PK_{UN} = \frac{SK_{UN}}{12} = \frac{242}{12} = 20,17$$

$$F = \frac{PK_{IZ}}{PK_{UN}} = \frac{242}{12} = 6,28$$

$$F_{0,05;2;12} = 3,89$$



	A1	A2	A3		Yij^2			
	35	38	42		1225	1444	1764	
	25	40	39		625	1600	1521	
	38	36	45		1444	1296	2025	
	26	35	38		676	1225	1444	
	36	31	46		1296	961	2116	
Suma	160	180	210	550	5266	6526	8870	20662
Suma^2	25600	32400	44100	102100				
Mg	32	36	42					
Mtot	36.67							
Ntot	15							
Ng	5							
g	3							

Primjer 1



Descriptives

uspjeh

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	5	32.00	6.042	2.702	24.50	39.50	25	38
2	5	36.00	3.391	1.517	31.79	40.21	31	40
3	5	42.00	3.536	1.581	37.61	46.39	38	46
Total	15	36.67	5.948	1.536	33.37	39.96	25	46

ANOVA

uspjeh

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	253.333	2	126.667	6.281	.014
Within Groups	242.000	12	20.167		
Total	495.333	14			

Složena ANOVA (sa više faktora)



- Kada postoje indicije da na posmatranu pojavu bitno utiče više faktora
- U dvofaktorskoj ANOVI postoje 3 različite nulte hipoteze (dvije o tzv. glavnim efektima faktora i treća o interakciji)
- Kod nezavisnih rezultata, suma kvadrata između grupa sastoji se iz tri dijela:
 - Variranje među grupama koje pripadaju kategorijama jedne varijable
 - Variranje među grupama koje pripadaju kategorijama druge varijable
 - Interakcija

Primjer 2



Prema jednu nacrtu istraživanja željelo se ispitati odnos intenziteta pušenja i starosne dobi pušača. Posmatrane su 3 grupe pušača starosnih dobi “30 g.”, “40 g.”, “50 g.” prema intenzitetu pušenja u 3 kategorije “jaki pušači”, “slabi pušači” i “nepušači”. Na taj način dobijeno je ukupno 9 grupa ispitanika, u svakoj po troje. Rezultati se nalaze u sljedećoj tabeli:

		Nezavisna varijabla A (starosna dob)		
		30	40	50
Nezavisna varijabla B (pušenje)	Jaki pušači	4	5	7
		5	6	9
		6	10	11
	Slabi pušači	3	2	3
		3	4	5
		6	6	4
	Nepušači	2	5	3
		2	3	6
		5	4	6

Primjer 2



		Nezavisna varijabla A (starosna dob)			
		30g.	40g.	50g.	
Nezavisna varijabla B (pušenje)	Jaki pušači	4	5	7	
		5	6	9	
		6	10	11	
	Suma Y	15	21	27	63
	Suma Y ²	77	161	251	489
	Slabi pušači	3	2	3	
		3	4	5	
		6	6	4	
	Suma Y	12	12	12	36
	Suma Y ²	54	56	50	160
	Nepušači	2	5	3	
		2	3	6	
		5	4	6	
	Suma Y	9	12	15	36
Suma Y ²	33	50	81	164	
		36	45	54	135
		164	267	382	813

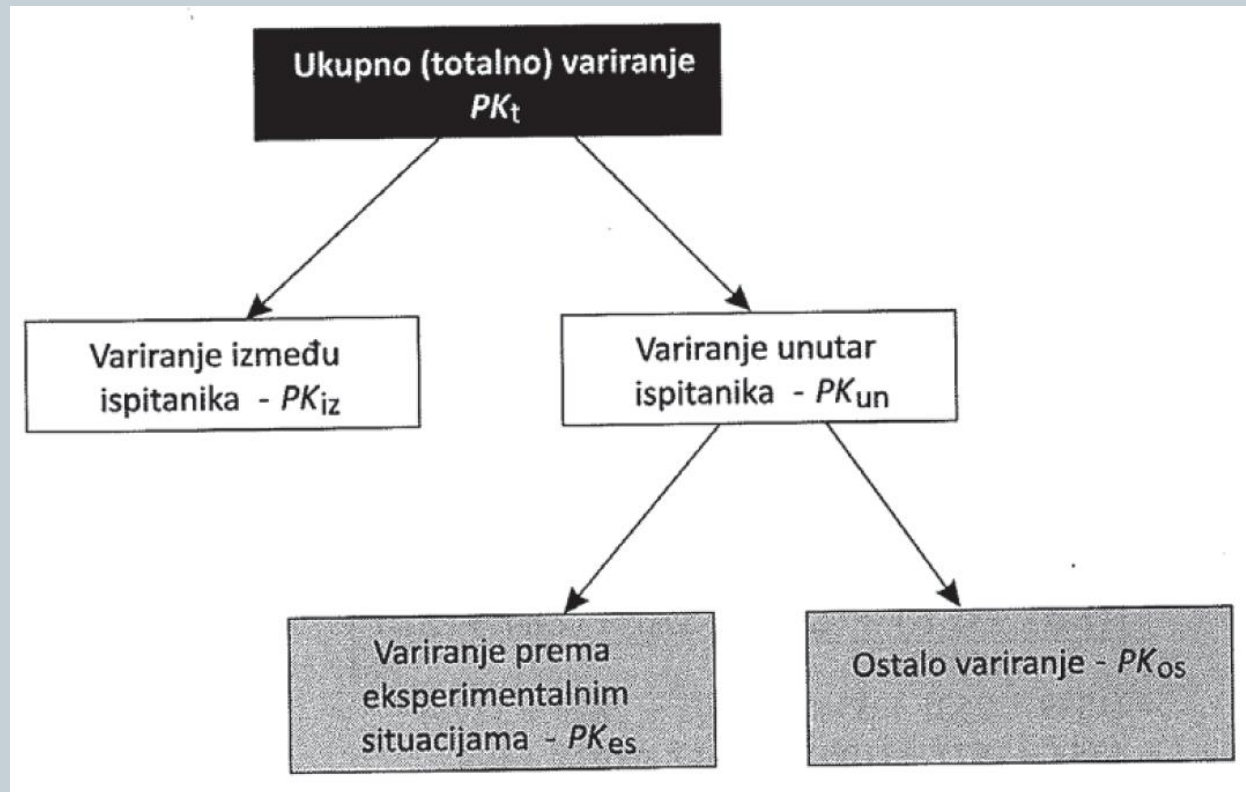
Tests of Between-Subjects Effects

Dependent Variable: disanje

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	84.000 ^a	8	10.500	3.500	.013
Intercept	675.000	1	675.000	225.000	.000
pušači	54.000	2	27.000	9.000	.002
starost	18.000	2	9.000	3.000	.075
pušači * starost	12.000	4	3.000	1.000	.433
Error	54.000	18	3.000		
Total	813.000	27			
Corrected Total	138.000	26			

a. R Squared = .609 (Adjusted R Squared = .435)

ANOVA (kod zavisnih uzoraka)



Primjer 3



Jedan proizvođač igračka želio je da ustanovi da li boja neke igračke utiče na njenu atraktivnost, pa je na 4 uzorka od po 10-oro djece mjerio minute koliko se pojedino dijete zadržalo u igri tom igračkom. Dobio je sljedeće rezultate:

Crveni	Žuti	Zeleni	Plavi
1	2	2	5
2	3	4	3
5	6	2	1
7	3	1	2
6	2	2	1
1	8	3	3
2	7	4	4
2	5	1	2
4	6	3	3
4	8	2	1
34	50	24	25
156	300	68	79

Primjer 3



ANOVA

Minuti

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	43.475	3	14.492	4.448	.009
Within Groups	117.300	36	3.258		
Total	160.775	39			

Šta nakon završenog računa ANOVE?



Na narednom času naučićemo više o post-hoc testovima...