

- Test hipoteza Analize varijanse (ANOVA)
- F raspodjela
- Računanje ANOVE
- Primjeri
- Modeli i faktori

10-1 ANOVA

- **ANOVA** (ANaliza VArijanse) je statistički metod za određivanje postojanja razlika između više od dva prosjeka populacija.
 - ANOVA otkriva razliku između prosjeka populacija koje su podvrgnute različitim *tretmanima*

10-2 Test hipoteza ANOVE

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots \mu_r$$

H_1 : Nijesu svi μ_i ($i = 1, \dots, r$) jednaki

10-3 F raspodjela

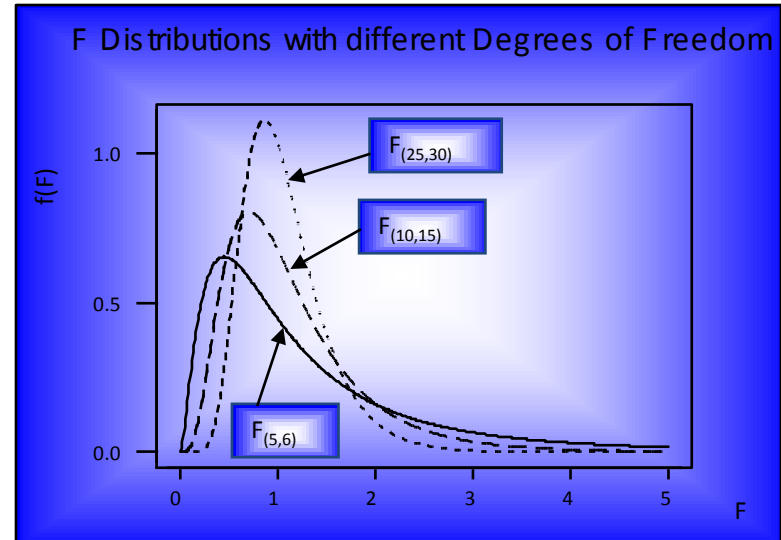
F raspodjela je raspodjela količnika dvije sl. promjenljive koje imaju hi-kvadrat raspodjelu, a nezavisne su međusobno, a svaka je podijeljena svojim brojem stepeni slobode.

F sl. promjenljiva sa k_1 i k_2 stepena slobode:

$$F_{(k_1, k_2)} = \frac{\frac{\chi_1^2}{k_1}}{\frac{\chi_2^2}{k_2}}$$

F raspodjela

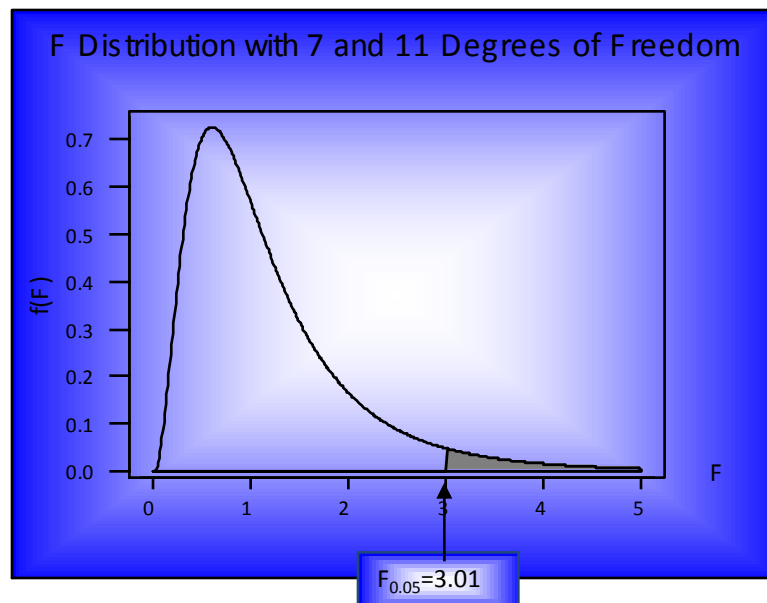
- F sl. promjenljiva ne može biti negativna, pa je ograničena nulom sa lijeve strane.
- F raspodjela je asimetrična udesno.
- F raspodjela se određuje brojem stepeni slobode brojica, k_1 , i stepena slobode imenioca, k_2 .



Tablice F raspodjele

Kritične vrijednosti F raspodjele koja odvaja
Prostor desnog repa od 0.05

k_1	1	2	3	4	5	6	7	8	9
k_2									
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59



Računanje ANOVE: F statistika

$$F_{(r-1, n-r)} = \frac{V_A}{V_R}$$

Računanje ANOVE: Varijanse

Faktorska varijansa:

$$V_A = \frac{S_A}{(r-1)}$$

Rezidualna varijansa:

$$V_R = \frac{S_R}{(n-r)}$$

Ukupna varijansa:

$$V_T = \frac{S_T}{(n-1)}$$

Ovdje ne važi pravilo aditivnosti!

Primjer 10-1

Slučajno izabrane grupe potrošača služene su različitim tipovima kafe i tražilo se da rangiraju vrstu kafe na skali od 0 do 100: 21 je služeno čistom brazilskom kafom, 20 kolumbijskom, i 22 afričkom kafom.

Rezultujuća test statistika je iznosila $F = 2.02$

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

H_1 : Nijesu sva tri prosjeka ista

$$n_1 = 21 \quad n_2 = 20 \quad n_3 = 22 \quad n = 21 + 20 + 22 = 63$$

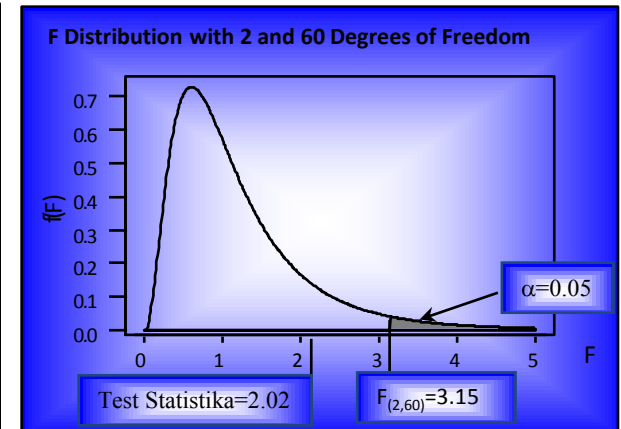
$$r = 3$$

Kritična vrijednost t za $\alpha = 0.05$ je:

$$F_{(r-1, (n-r))} = F_{(3-1, 63-3)} = F_{(2, 60)} = 3.15$$

$$F = 2.02 < F_{(2, 60)} = 3.15$$

H_0 se ne može odbaciti, i ne možemo zaključiti da postoji značajna razlika



5. zadatak

Pogoni V, VI i VII proizvode isti proizvod. Ispituje se utrošak materijala po proizvodu, i to po 8 proizvoda u svakom od pogona. Na osnovu podataka o potrošnji materijala dobijene su sledeće vrijednosti: suma kvadrata ukupnih odstupanja $ST=206,5$ i suma faktorskih odstupanja 41,07. Ako je teorijska vrijednost F distribucije 3,47, zaključuje se: $r=3$, $n=8*3=24$ (ne postoji razlika u utrošku materijala)

$$F = \frac{V_A}{V_R} = \frac{\frac{S_A}{r-1}}{\frac{S_R}{n-r}} = 2.605$$

$$S_R = S_T - S_A = 165.43$$

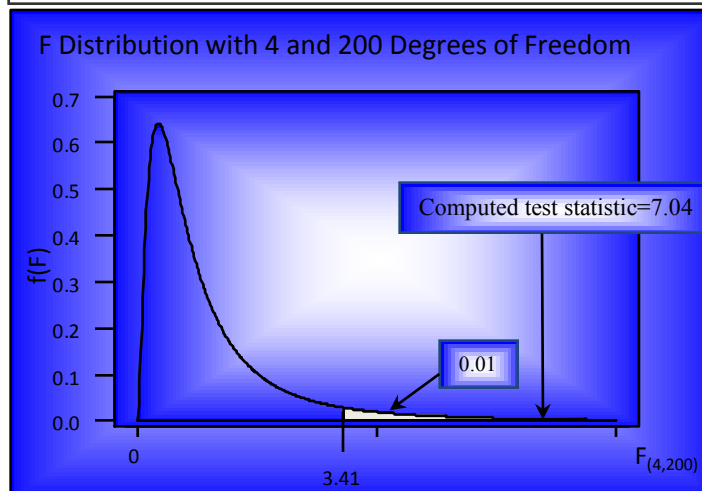
$$F > 3.47$$

Primjer 10-3: Club Med

Club Med sprovodi test da bi odredio da li su Karipska ostrva jednako atraktivna za odmor od članova kluba. Analiza je bazirana na upitniku (opšte zadovoljstvo, na skali od 0 do 100) ispunjenom u slučajnom uzorku od 40 ispitanika za svaku od 5 destinacija.

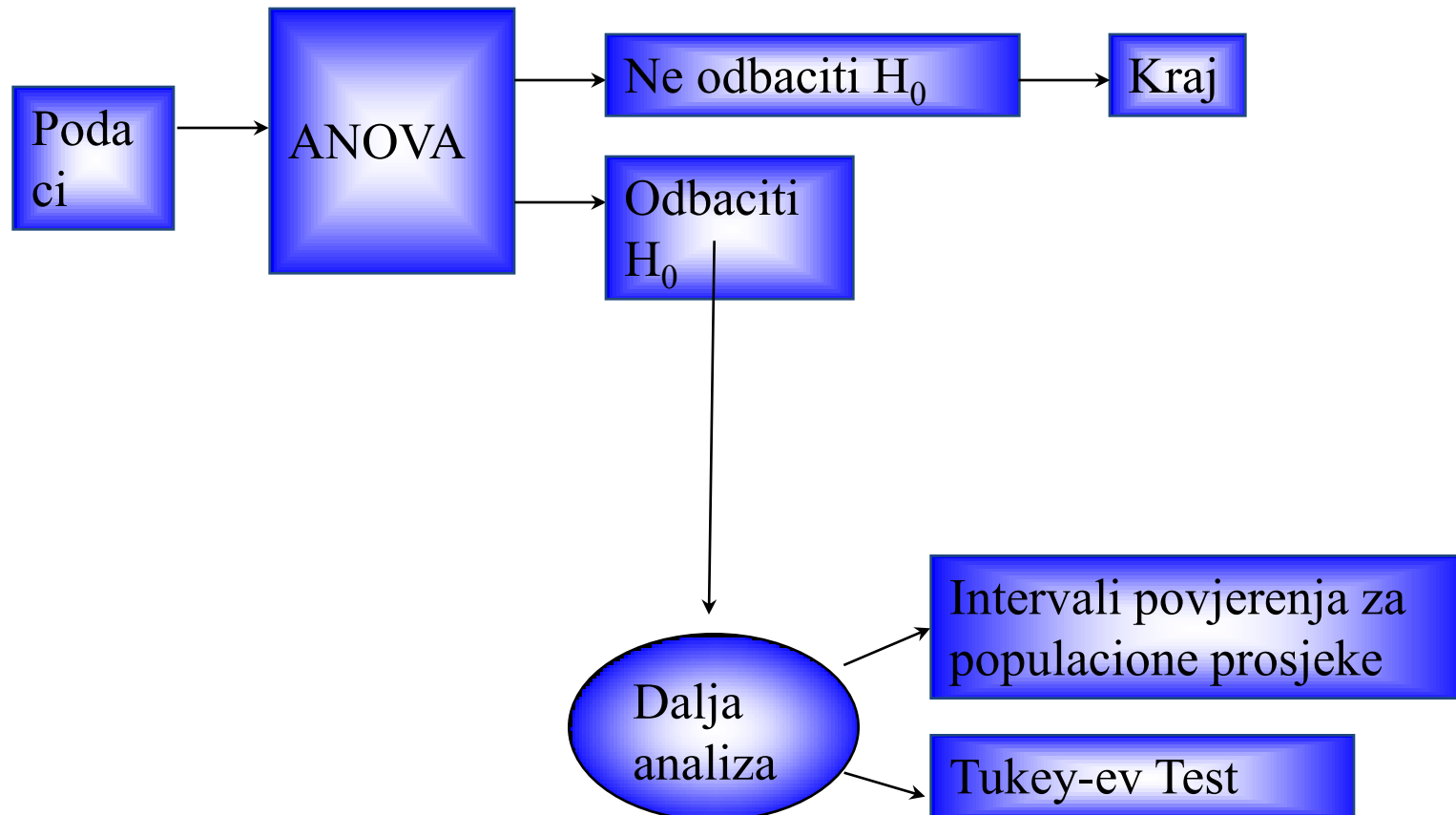
Destinacija	Prosječni odgov. ($\bar{x}_i$)
Guadeloupe	89
Martinique	75
Eleuthra	73
Paradise Island	91
St. Lucia	85
$S_T=112564$ $S_R=98356$	

Izvor varijacija	Suma kvadrata	Stepeni slobode	Varijansa	F
Tretman	$S_A= 14208$	$(r-1)= 4$	$V_A= 3552$	7.04
Greška	$S_R=98356$	$(n-r)= 195$	$V_R= 504.39$	
Ukupno	$S_T=112564$	$(n-1)= 199$	$V_T= 565.65$	



F količnik je veći od tablične vrijednosti za $\alpha = 0.01$, pa se nulta hipoteza odbacuje.

10-6 Dalja analiza



ANOVA Dijagram

Tukey-ev test

Tukey-ev test, omogućava nam da uporedimo svaki par populacionih prosjeka sa jedinstvenim nivoom značajnosti.

Bazira se na **studentovoj raspodjeli ranga**, q , sa r i $(n-r)$ stepena slobode.

Tablična vrijednost u Tukey-evom testu je u **Tukey kriterijumu**:

$$T = q_{\alpha} \frac{\sqrt{V_R}}{\sqrt{n_i}}$$

Gdje je n_i *najmanja* od r veličina uzoraka.

Statistika testa je *apsolutna vrijednost razlike* između odgovarajućih uzoračkih prosjeka, i **nulta hipoteza se odbacuje** ako je test statistika veća od tablične vrijednosti Tukey-evog kriterijuma.

Tukey-ev test: Club Med primjer

Statistika testa za test svakog para je apsolutna razlika između odgovarajućih prosjeka uzoraka.

i	Destinacija	Prosjek
1	Guadeloupe	89
2	Martinique	75
3	Eleuthra	73
4	Paradise Is.	91
5	St. Lucia	85

Tablična vrijednost $T_{0.05}$ za $r=5$ i $(n-r)=195$ stepena slobode je:

$$T = q_{\alpha} \frac{\sqrt{MSE}}{\sqrt{n_i}}$$

$$= 3.86 \frac{\sqrt{504.4}}{\sqrt{40}} = 13.7$$

Odbaciti nultu hipotezu ako je apsolutna vrijednost razlike između uzoračkih prosjeka veća od tablične vrijednosti T. (hipoteze označene sa * se odbacuju.)

I. $H_0: \mu_1 = \mu_2$
 $H_1: \mu_1 \neq \mu_2$
 $|89-75|=14 > 13.7^*$

II. $H_0: \mu_1 = \mu_3$
 $H_1: \mu_1 \neq \mu_3$
 $|89-73|=16 > 13.7^*$

III. $H_0: \mu_1 = \mu_4$
 $H_1: \mu_1 \neq \mu_4$
 $|89-91|=2 < 13.7$

IV. $H_0: \mu_1 = \mu_5$
 $H_1: \mu_1 \neq \mu_5$
 $|89-85|=4 < 13.7$

V. $H_0: \mu_2 = \mu_3$
 $H_1: \mu_2 \neq \mu_3$
 $|75-73|=2 < 13.7$

VI. $H_0: \mu_2 = \mu_4$
 $H_1: \mu_2 \neq \mu_4$
 $|75-91|=16 > 13.7^*$

VII. $H_0: \mu_2 = \mu_5$
 $H_1: \mu_2 \neq \mu_5$
 $|75-85|=10 < 13.7$

VIII. $H_0: \mu_3 = \mu_4$
 $H_1: \mu_3 \neq \mu_4$
 $|73-91|=18 > 13.7^*$

IX. $H_0: \mu_3 = \mu_5$
 $H_1: \mu_3 \neq \mu_5$
 $|73-85|=12 < 13.7$

X. $H_0: \mu_4 = \mu_5$
 $H_1: \mu_4 \neq \mu_5$
 $|91-85|=6 < 13.7$

10-4. zadatak

Na uzorcima od po 5 studenata četvrte godine Ekonomskog fakulteta, koji su pohađali tri različita metoda nastave iz jednog predmeta, vrši se ispitivanje uticaja metoda nastave na uspjeh studenata na ispitu. Na osnovu podataka dobijena je ukupna suma kvadrata odstupanja koja iznosi 2923,33 i faktorska suma kvadrata odstupanja 869,73. Uz rizik greške 0,05 (odgovarajuća F-statistika je 3,89) metod nastave značajno utiče na uspjeh studenata na ispitu:

$$r=3, n=5*3=15$$

$$F = \frac{V_A}{V_R} = \frac{\frac{S_A}{r-1}}{\frac{S_R}{n-r}} = 2.54$$

$$S_R = S_T - S_A = 2053.6$$

$$F > 3.89 \text{ _ne!}$$

10-5. zadatak

Posmatra se produktivnost fabrike u tri proizvodne smjene, na uzorku od po 5 radnika. Uz rizik greške 5%, i vrijednosti za sume kvadrata $S_A=19.6$ i $S_R=8.8$ zaključuje se da ($F=3,89$): (izbor smjene sistematski utiče na varijabilitet pojave)

$$r=3, n=5*3=15 \qquad F = \frac{V_A}{V_R} = \frac{\frac{S_A}{r-1}}{\frac{S_R}{n-r}} = 13.36$$

$$F > 3.89$$