Инжењерска статистика
Вежба 6

Анализа варијансе

- Назив области
- Дисперзиона анализа
- Анализа варијансе
- Пројектовање и анализа факторијелних експерименталних планова: ANOVA (ANalysis Of VArince)

Експеримент

Кораци у поступку тестирања

Сваки експеримент укључује следеће кораке:

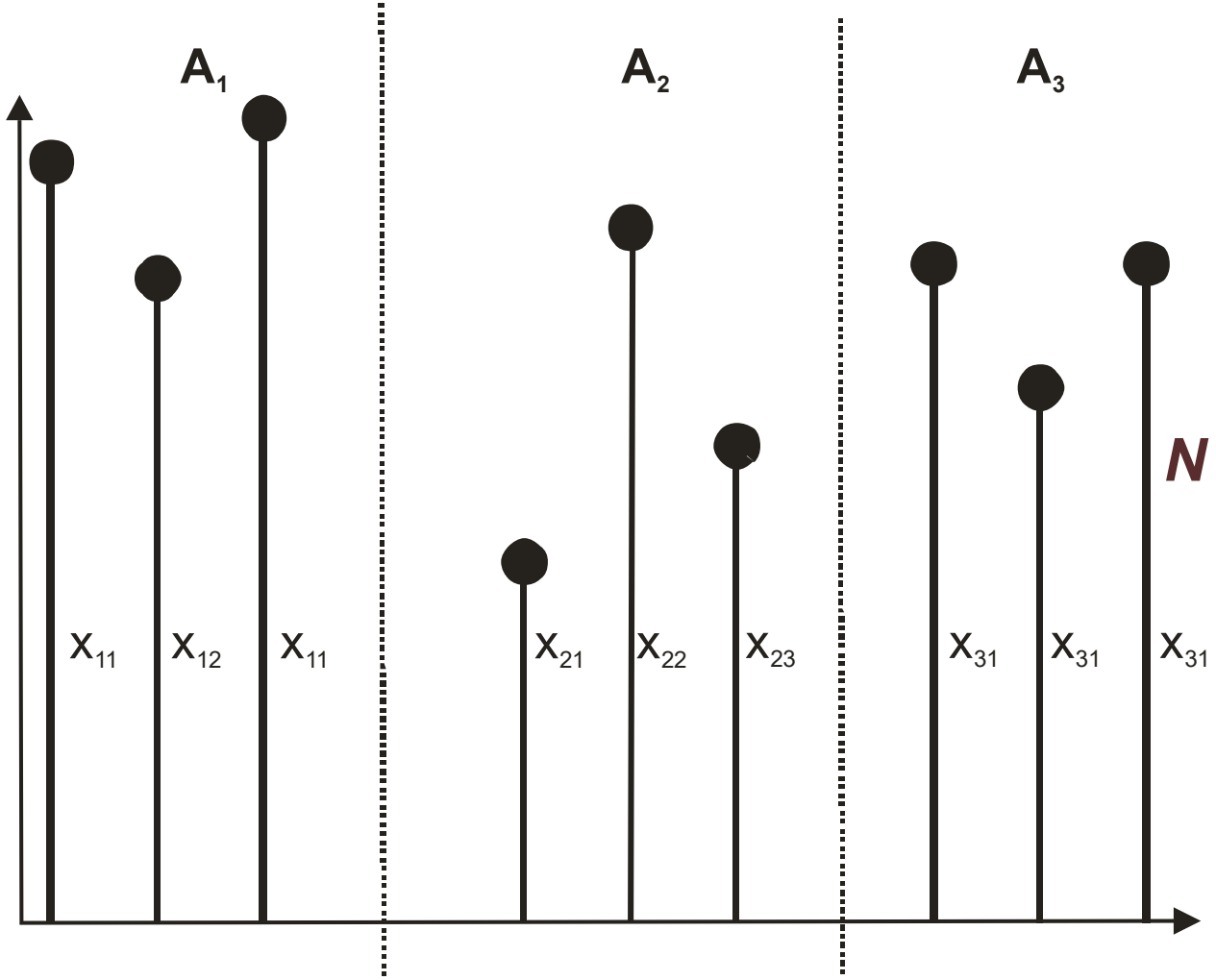
- **постављање почетне хипотезе** која мотивише експеримент
- **експеримент**, спровођење да би се добили резултати
- **анализа** –статистичка анализа података експеримента
- **закључци** о томе шта се добија на основу почетних хипотеза

Позвати статистичара пошто је експеримент спроведен је тражити од њега да изврши post mortem: моћи ће само да због чега експеримент није исправан.

(*Fisher, RA, 1938*)

Основни појмови

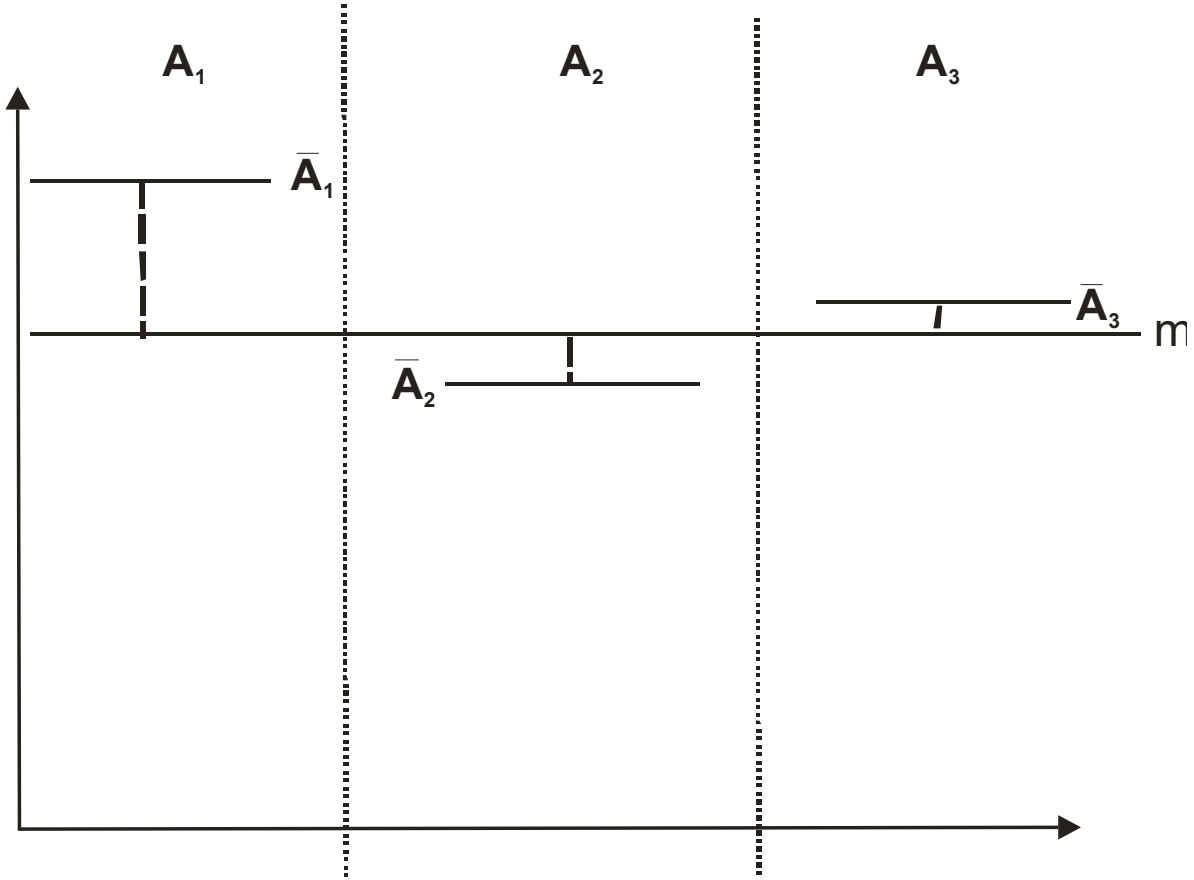
- Ниво (третман) фактора $A_j, j = 1...a$
- Репликација $x_{ji}, j = 1...a, i = 1...n$
- Рандомизација
- Експериментални план
  - Балансиран $\left(n_1 = ... = n_j = ... = n_a = n\right)$
  - Небалансиран $\left(n \neq const\right)$
- Фиксни ефекти
- Са блокирајућим фактором



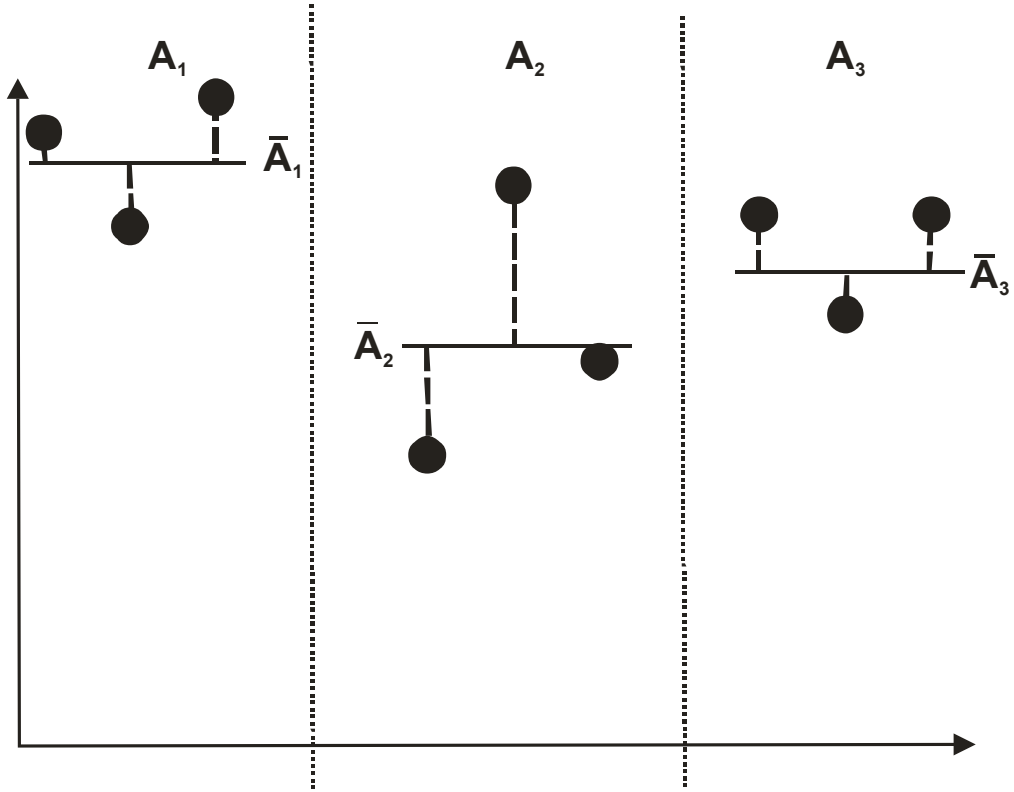
$$T = \sum_{j=1}^{a} \sum_{i=1}^{n} x_{ji}$$

$$N = a \cdot n = \sum_{j=1}^{a} n_j$$

$$\bar{x} = \frac{T}{N}$$



$$A_j = \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \; j = 1 \ldots a$$



$$e_{ij} = \left(x_{ji} - \bar{x}_j\right), j = 1...a, i = 1...n$$

Постављање модела

$$x_{ji} = \mu + \tau_j + e_{ji} = \bar{x} + \bar{x}_j + e_{ji} \quad \mathbb{N}\left(\mu, \sigma^2\right)$$

Хипотезе

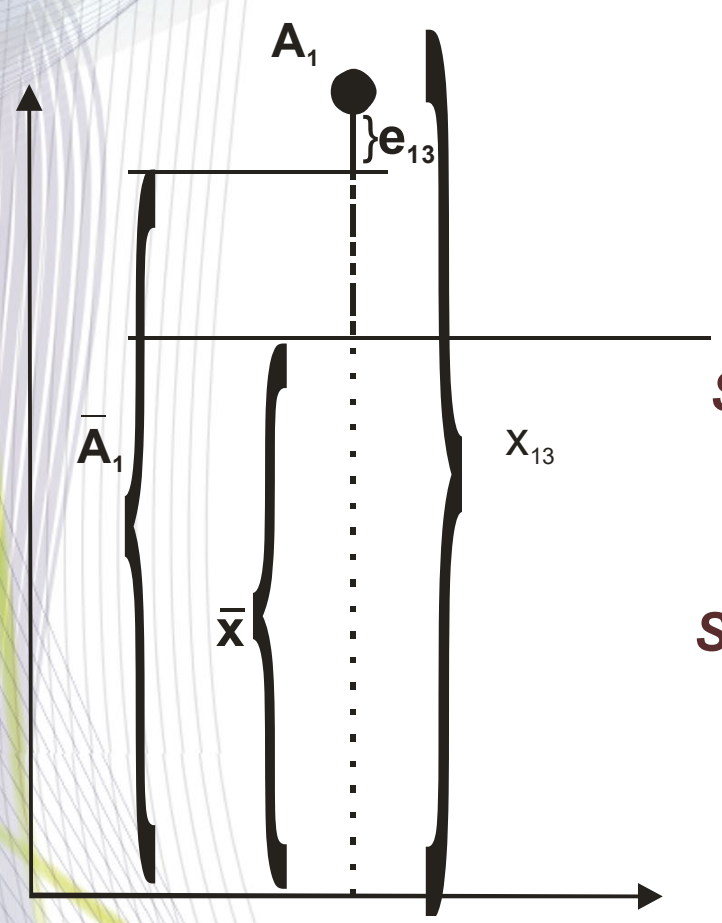
$$H_0\left(\tau_j = 0, j = 1...a\right) \quad H_0\left(\mu_1 = \mu_2 = ... = const\right)$$

$$H_1\left(\exists \tau_i \neq 0, j = 1...a\right) \quad H_1\left(\exists \mu_i \neq const, j = 1...a\right)$$

Поступак:

1. Суме квадрата
2. Број степени слободе
3. Оцене варијанси
4. Прорачунска вредност F теста однос оцена варијанси
5. Формирање ANOVA табеле

Варијације се дефинишу преко сума квадрата



$$SK_T = \sum_{j=1}^{a}\sum_{i=1}^{n}\left(x_{ji} - \frac{T}{N}\right)^2 = \sum_{j=1}^{a}\sum_{i=1}^{n}\left(x_{ji} - \bar{x}\right)^2$$

$$SK_A = \sum_{j=1}^{a}\left(\bar{A}_j - \mu\right)^2 = \sum_{j=1}^{a}\left(\bar{x}_j - \bar{x}\right)^2$$

$$SK_e = \sum_{j=1}^{a}\sum_{i=1}^{n}\left(x_{ji} - \bar{A}_j\right)^2 = \sum_{j=1}^{a}\sum_{i=1}^{n}\left(x_{ji} - \bar{x}_j\right)^2$$

Суме квадрата

$$SK_T = \sum_{j=1}^{a}\sum_{i=1}^{n}\left(x_{ji} - \bar{x}\right)^2 = \sum_{j=1}^{a}\sum_{i=1}^{n}\left(x_{ji} - \bar{x} - \bar{x}_j + \bar{x}_j\right)^2$$

$$= \sum_{j=1}^{a}\sum_{i=1}^{n}\left[\left(x_{ji} - \bar{x}_j\right) + \left(\bar{x}_j - \bar{x}\right)\right]^2$$

$$T = \sum_{j=1}^{a}\sum_{i=1}^{n} x_{ji} \qquad N = a \cdot n \qquad A_j = \sum_{i=1}^{n} x_{ij}$$

$$SK_T = SK_A + SK_e$$

$$SK_T = \sum_{j=1}^{a}\sum_{i=1}^{n} x_{ij}^2 - \frac{T^2}{N}$$

$$SK_A = \frac{1}{n}\sum_{j=1}^{a}\left(\sum_{i=1}^{n} x_{ij}\right)^2 - \frac{T^2}{a \cdot n} = \frac{1}{n}\sum_{j=1}^{a} A_j^2 - \frac{T^2}{a \cdot n}$$

$$SK_e = SK_T - SK_A$$

Број степени слободе

$$f_T = N - 1 = a \cdot n - 1$$

$$f_T = f_A + f_e$$

$$f_A = a - 1$$

$$f_e = f_T - f_A$$

Оцена варијансе

$$S_A^2 = \frac{SK_A}{f_A}$$

$$S_e^2 = \frac{SK_e}{f_e}$$

$$\Rightarrow$$

Прорачунска вредност *F*-теста

$$F = \frac{S_A^2}{S_e^2}$$

Таблична вредност F-теста

$$F_{\alpha, f_A, f_e}$$

Фактор има утицај ако је: $F_{\alpha, f_A, f_e} < F_0$

Табела анализе варијансе

Извор варијације	Сума квадрата	Број степени слободе	Оцена варијансе	F
Фактор	SK_A	f_A	S_A^2	F_0
Грешка	SK_e	f_e	S_e^2	
Укупно	SK_T	f_T		

Претпоставке:

- x_{ij} су независне случајне променљиве, *i* - тог мерења *j* - тог нивоа
- Случајне грешке e_{ji} су независно и нормално распоређене по: $\mathbb{N}\left(0,\sigma^2\right),\ \sigma = const$
- ***Последица:*** сваки ниво фактора је случајна величина са нормалном расподелом
- Сва мерења су случајна и спроведена су случајним редоследом (рандомизација)

$$\mathbb{N}\left(0,\sigma^2\right)$$

Услови:

- Све суме квадрата су *увек* веће од 0
- Број степени слободе је *увек* већи од 0

Пример 1

Под идентичним условима тестирани су случајни узорци 4 врсте аутобуских гума. Резултати су приказани у табели. Потребно је, за ниво значајности 0.01, утврдити да ли постоји разлика у квалитету између типова аутобуских гума.

Тип гума / Мерење	I	II	III	IV
1	27	22	28	25
2	31	20	30	25
3	25	21	28	26
4	22	21	35	28

Испитивани фактор: *A* – врста (тип) аутобуских гума (нпр. траг кочења)

Број нивоа фактора: $a=4$

Број мерења у оквиру нивоа (репликација): $n=4$

Тип експеримента: балансиран

Тип гума / мерење	A_1	A_2	A_3	A_4
1	27	22	28	25
2	31	20	30	25
3	25	21	28	26
4	22	21	35	28
Σ	105	84	121	104

Хипотезе

$$H_0\left(\bar{A}_1 = \bar{A}_2 = \bar{A}_3 = \bar{A}_4 = \bar{A}\right)$$

$$H_1\left(\exists \bar{A}_j \neq \bar{A}\right)$$

Припремни подаци

$$A_1 = \sum_{i=1}^{4} x_{i1} = 27 + 31 + 25 + 22 = 105$$

$$A_2 = \sum_{i=1}^{4} x_{i2} = 22 + 20 + 21 + 21 = 84$$

$$A_3 = \sum_{i=1}^{4} x_{i3} = 28 + 30 + 28 + 35 = 121$$

$$A_4 = \sum_{i=1}^{4} x_{i4} = 25 + 25 + 26 + 28 = 104$$

$$T = \sum_{j=1}^{a}\sum_{i=1}^{n} x_{ij} = \sum_{j=1}^{a} A_j = 105 + 84 + 121 + 104 = 414$$

$$N = a \cdot n = 4 \cdot 4 = 16$$

Суме квадрата

$$SK_T = \sum_{j=1}^{a}\sum_{i=1}^{n} x_{ij}^2 - \frac{T^2}{a \cdot n}$$

$$SK_T = \left(27^2 + 31^2 + \cdots + 26^2 + 28^2\right) - \frac{414^2}{16} = 255.75$$

$$SK_A = \frac{1}{n}\sum_{j=1}^{a}\left(\sum_{j=1}^{n} x_{ij}\right)^2 - \frac{T^2}{a \cdot n} = \frac{1}{n}\sum_{j=1}^{a} A_j^2 - \frac{T^2}{a \cdot n}$$

$$SK_A = \frac{1}{4}\left(105^2 + 84^2 + 121^2 + 104^2\right) - \frac{414^2}{16} = 172.25$$

$$SK_e = SK_T - SK_A = 255.75 - 172.25 = 83.5$$

Број степени слободе

$$f_T = N - 1 = a \cdot n - 1 = 16 - 1 = 15$$

$$f_T = f_A + f_e$$

$$f_A = a - 1 = 4 - 1 = 3$$

$$f_e = f_T - f_A = 15 - 3 = 12$$

Оцене варијанси

$$S_A^2 = \frac{SK_A}{f_A} = \frac{172.25}{3} = 57.417$$

$$S_e^2 = \frac{SK_e}{f_e} = \frac{83.5}{12} = 6.958$$

Прорачунска вредност *F*-теста

$$F = \frac{S_A^2}{S_e^2} = \frac{57.4167}{6.9583} = 8.2515$$

Таблична вредност *F*-теста

$$F_{\alpha, f_A, f_e} = F_{0.01,3,12} = 5.95$$

Табела анализе варијанси

извор варијације	*SK*	*f*	OV	*F*	*p*
фактор	172.25	3	57.4167	8.2515**	0.003013
грешка	83.50	12	6.95833		
Укупно	255.75	15			

Пример 2

Студенти су слушали вежбе из истог предмета код 5 различитих наставника. Сваки наставник је имао одређени број група. На крају семестра су посматрани критеријуми оцењивања наставника, на основу просечних оцена постигнутих у појединим групама. Резултати су приказани у табели. Потребно је утврдити да ли наставници имају исти критеријум оцењивања.

I	II	III	IV	V
6.5	7.5	6.9	9.4	8.3
7.7	6.9	7.8	8.9	6.1
7.3	8.3	6.7	8.7	7.2
7.9	8.1	6.2	8.8	6.7
7.1	7.2	8.3		8.6
6.9	7.9	7.6		
	9.0			

Испитивани фактор: A - наставник

Број нивоа фактора: $a=5$

Број репликација: $n_1=6$, $n_2=7$, $n_3=6$, $n_4=4$, $n_5=5$,

Тип експеримента: небалансиран

	A_1	A_2	A_3	A_4	A_5
	6.5	7.5	6.9	9.4	8.3
	7.7	6.9	7.8	8.9	6.1
	7.3	8.3	6.7	8.7	7.2
	7.9	8.1	6.2	8.8	6.7
	7.1	7.2	8.3		8.6
	6.9	7.9	7.6		
		9.0			
Σ	43.4	54.9	43.5	35.8	36.9
n	6	7	6	4	5

Хипотезе

$$H_0\left(\bar{A}_1 = \bar{A}_2 = \bar{A}_3 = \bar{A}_4 = \bar{A}_5 = \bar{A}\right)$$

$$H_1\left(\exists \bar{A}_j \neq \bar{A}\right)$$

Припремни подаци

$$A_1 = \sum_{i=1}^{6} x_{i1} = 6.5 + \cdots + 6.9 = 43.4$$

$$A_2 = \sum_{i=1}^{7} x_{i2} = 7.5 + \cdots + 9.0 = 54.9$$

$$A_3 = \sum_{i=1}^{6} x_{i3} = 6.9 + \cdots + 7.6 = 43.5$$

$$A_4 = \sum_{i=1}^{4} x_{i4} = 9.4 + \cdots + 8.8 = 35.8$$

$$A_5 = \sum_{i=1}^{5} x_{i5} = 8.3 + \cdots + 8.6 = 36.9$$

$$T = \sum_{j=1}^{5}\sum_{i=1}^{n_j} x_{ij} = \sum_{j=1}^{5} A_j = 43.4 + 54.9 + 43.5 + 35.8 + 36.9 = 214.5$$

$$N = \sum_{j=1}^{a} n_j = 6 + 7 + 6 + 4 + 5 = 28$$

Суме квадрата

$$SK_T = \sum_{j=1}^{a}\sum_{i=1}^{n} x_{ij}^2 - \frac{T^2}{N}$$

$$SK_T = \left(6.5^2 + 7.7^2 + \cdots + 6.7^2 + 8.6^2\right) - \frac{214.5^2}{28} = 21.567$$

$$SK_A = \sum_{j=1}^{a} \frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2 - \frac{T^2}{N} =$$

$$= \sum_{j=1}^{a} \frac{A_j^2}{n_j} - \frac{T^2}{N} = \frac{43.4^2}{6} + \frac{54.9^2}{7} + \frac{43.5^2}{6} + \frac{35.8^2}{4} + \frac{36.9^2}{5} - \frac{214.5^2}{28} =$$

$$= 9.383$$

$$SK_e = SK_T - SK_A = 21.567 - 9.383 = 12.184$$

Број степени слободе

$$f_T = N - 1 = 28 - 1 = 27$$

$$f_T = f_A + f_e$$

$$f_A = a - 1 = 5 - 1 = 4$$

$$f_e = f_T - f_A = 27 - 4 = 23$$

Оцене варијанси

$$S_A^2 = \frac{SK_A}{f_A} = \frac{9.383}{4} = 2.346$$

$$S_e^2 = \frac{SK_e}{f_e} = \frac{12.1834}{23} = 0.5297$$

Прорачунска вредност *F*-теста

$$F = \frac{S_A^2}{S_e^2} = \frac{2.3458}{0.5297} = 4.428$$

Таблична вредност *F*-теста

$$F_{\alpha, f_A, f_e} = F_{0.05,4,23} = 2.8$$

Табела анализе варијансе

Извор варијације	*SK*	*f*	OV	*F*	*p*
Фактор	9.383	4	2.346	4.4285**	0.00844
Грешка	12.183	23	0.5297		
Укупно	21.567	27			