

Analysis of Variance

ANOVA

Bariantzaren analisia

ZER DA ANOVA EDO BARIANTZAREN ANALISIA?

Bariantza iturri bat baino gehiago dugunean, bariantza iturri nagusia eta errore aleatorioa banatu eta estimatzen ditu.

KASU BATZUK:

- Konposatu baten fluoreszentzia pH desberdinetan neurtu da. Eragina al du baldintza ezberdinak erabiliz lortutako fluoreszentiaren batezbestekoak?
- Mn-a hiru metodo desberdinekin determinatu da, kontzentrazioen batezbestekoak desberdinak al dira?
- Fertilizante berri bat frogatu nahi da, eta konparatzeko lur sail batean aplikatuko da, baina beste "kontrol" batean ez. Egon daitezkeen ezberdintasunak fertilizanteagatik dira?
- Farmako berri bat frogatu nahi da, eta horretarako 5 pertsona talde egin dira. 4 taldeei farmakoaren kontzentrazio ezberdinak eman zaizkie, 5. taldeari, aldiz, plazeboa eman zaio. Eraginkorra da farmako berria edo bere eraginak plazeboarenarekin alderatu daitezke?
- Pertsona talde bati dieta ezberdinak jarri dizkiegu (adib 5 pertsonako 4 talde, eta talde bakoitzari dieta bat). Ezberdintasunik badago dieta ezberdinen artean?
- Danone adibidea (merkatuaren azterketa kasu bat):

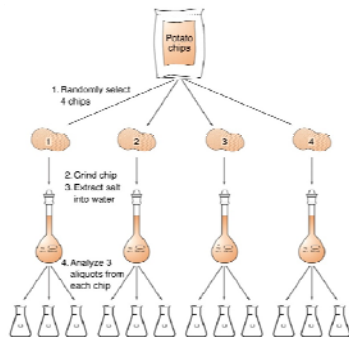
<https://www.youtube.com/watch?v=hCA0HBBc2Gs>

ANOVA

Batezbesteko bat baino gehiagok parte hartzen duten egoeretan egin beharreko konparaketak:

Adibidez:

✓Sodia determinatu nahi da patata frijituko poltsa batean. Ezberdina izango al da gatz edukia patata batean edo bestean?



ANOVA

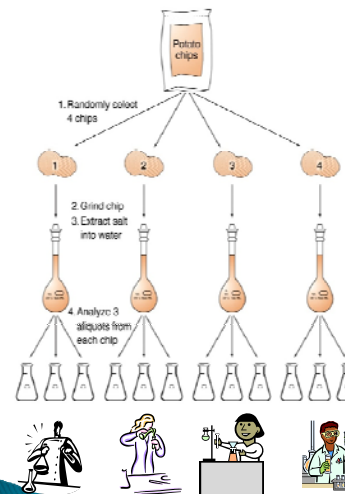
Hemen bi bariazio iturri posible aurki daitezke:

✓**Errore aleatorioa.**
Bariazio iturri hau beti aurkituko dugu. Errore honen eraginez baldintza ezberdinetan egindako neurketa emaitza ezberdinak emango dituzte.

✓**Kontrolatutako faktorea edo efektu finkoa.**
Aurreko adibideetan: metodo ezberdinak, hartutako lagina, erabilitako baldintza edota balorazio egin duten pertsona ezberdinei dagokien bariazioa.

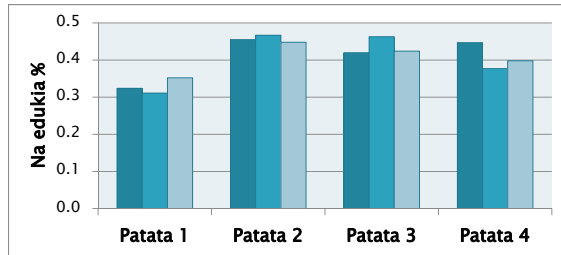
ANOVA

Bariazio iturri desberdinak banatu eta estimatzeko erabil daitezkeen teknika estatistikoa da



ANOVA

Patata poltsa				
	Patata 1	Patata 2	Patata 3	Patata 4
Na edukia %	0.324	0.455	0.420	0.447
	0.311	0.467	0.463	0.377
	0.352	0.448	0.424	0.398



ANOVA

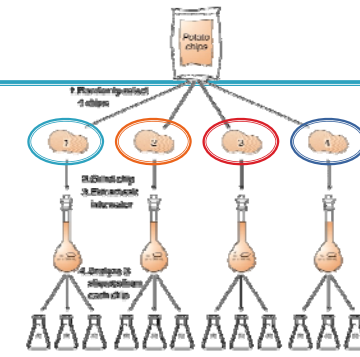


Table 29-2 Analysis of sodium (wt %) in potato chips

Statistic	Chip 1	Chip 2	Chip 3	Chip 4
FINDING STANDARD DEVIATION WITHIN SAMPLE:				
wt % Na	0.324	0.455	0.420	0.447
	0.311	0.467	0.463	0.377
	0.352	0.448	0.424	0.398
Average within sample	0.329 ₃	0.456 ₃	0.435 ₃	0.407 ₃
Standard deviation within sample	0.020 ₃	0.009 ₃	0.023 ₃	0.035 ₃
	\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4
	s_1	s_2	s_3	s_4

$n = \text{number of replicates} = 3$ $nh = \text{total measurements}$

ANOVA

DESBIDERATZE ESTANDARRA:
$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

BARIANTZA:
$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Bariantza lagin barruan (within samples, dentro de muestras, intra-grupos): patata bakoitzarekin lortutako repliken desbizioetatik (bariantzak) abiatuz egingo dugu 4 pataten bariantzen batezbestekoa.

Patata poltsa					
		Patata 1	Patata 2	Patata 3	Patata 4
Na edukia (pisu portzentaian)		0.324	0.455	0.420	0.447
		0.311	0.467	0.463	0.377
		0.352	0.448	0.424	0.398
LAGIN BARRUAN (within samples)	Batezbestekoa	0.329	0.457	0.436	0.407
	s	0.021	0.010	0.024	0.036
	s ²	4.390E-04	9.233E-05	5.643E-04	1.290E-03
	s ² lagin barruan	5.965E-04	lortutako bariantzen batezbestekoa		

ANOVA

Bariantza laginen artean (between samples, entre muestras, inter-grupos): bariantzaren formula erabiliko dugu eta batezbestekotzat patata bakoitzaren replikekin lortutako batezbestekoak erabiliko ditugu

Patata poltsa					
		Patata 1	Patata 2	Patata 3	Patata 4
Na edukia (pisu portzentaian)		0.324	0.455	0.420	0.447
		0.311	0.467	0.463	0.377
		0.352	0.448	0.424	0.398
LAGIN BARRUAN (within samples)	Batezbestekoa	0.329	0.457	0.436	0.407
	s	0.021	0.010	0.024	0.036
	s ²	4.390E-04	9.233E-05	5.643E-04	1.290E-03
	s ² lagin barruan	5.965E-04	lortutako bariantzen batezbestekoa		
LAGIN ARTEAN (between samples)	Batezbestekoa	0.407			
	s ² lagin artean	9.373E-03	lortutako bariantzen batezbestekoa		

$S^2_{lagin artean} = n s^2_{(lagin batezbestekoa)}$ lagin bakoitzaren batezbestekoa egiteko n replika egin ditugulako

Batezbesteko orokorra: lortutako 12 balioekin kalkulatu

s² (lagin batezbestekoen): patata bakoitzaren batezbestekoa eta batezbesteko orokorra kontuan hartuta

s² lagin artean: n replika egin ditugunez batezbesteko "individual" bakoitza lortzeko, horrekin erlazionatu behar da.

ANOVA

Bi bariantzak konparatzeko: F froga

$$F = \frac{S_{\text{laginartean}}^2}{S_{\text{laginbarruan}}^2}$$

Hypothesis testing with F test:

$k - 1 = 3$ degrees of freedom

$$F = \frac{9.37 \times 10^{-3}}{5.96 \times 10^{-4}} = 15.7 > 4.07 \text{ in Table 4-5}$$

$nk - k = 8$ degrees of freedom

Beraz, bi bariantzen artean ezberdintasun adierazgarriak daude

$$\rightarrow S_{\text{laginartean}}^2 \neq S_{\text{laginbarruan}}^2$$

ONDORIOA: patata bat edo beste hartu, ezberdintasun adierazgarriak egongo dira beraien gatz edukian!!

9

ANOVA

Adibidea. Konposatu fluoreszente bat baldintza ezberdinetan gorde da. Azterketa egin da, konposatuaren fluoreszentsia neurtuz eta ondorengo datuak lortu dira:

Baldintzak	Intentsitatea (I_f) (3 neurketa)
A, prestatu berria	102,100,101
B, ordubetez ilunpetan	101,101,104
C, ordubetez argi ahuletan	97,95,99
D, ordubetez argi distiratsutan	90,92,94

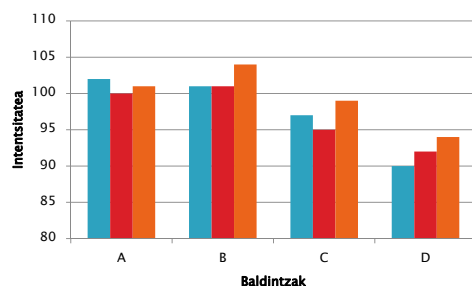
Baldintza ezberdinek eragina al dute erreaktiboaren egonkortasunean?

10

ANOVA

ANOVA aplikatzen hasi aurretik interesgarria izan daiteke irudikapen bat egitea eta baita A-D baldintzei dagozkien datuen batezbestekoa kalkulatzeko:

Baldintzak	I_f (X)
A	101
B	102
C	97
D	92



11

ANOVA

Baldintzak/ LAGINAK	I_f	Batezbestekoa
A	102,100,101	$X_A=101$
B	101,101,104	$X_B=102$
C	97,95,99	$X_C=97$
D	90,92,94	$X_D=92$

$$X = 98$$

Hipotesi nulua (H_0): lagin guztiak μ batezbestekoa eta σ_0^2 bariantza dituen populazioa batetik hartu dira.

12

ANOVA

Bi bariazio iturri:

ERRORE ALEATORIOA edo **LAGIN BARRUKO BARIANTZA.**

Laginen bariantzaren batezbestekoa (s^2 barruan)

KONTROLATUTAKO FAKTOREA edo **LAGIN ARTEKO BARIANTZA.**

Laginen batezbestekoen batezbestekoari dagozkion bariantza (s^2 artean)

13

ANOVA

Bi bariazio iturri:

LAGIN BARRUKO BARIANTZA. a.g.= h (n-1)

$$\sigma_{barruan}^2 = \frac{\sum_i \sum_j (X_{ij} - \bar{X}_i)^2}{h \cdot (n-1)}$$

LAGIN ARTEKO BARIANTZA. a.g.= h-1

$$\sigma_{artean}^2 = \frac{n \sum_i (\bar{X}_i - \bar{X})^2}{(h-1)}$$

$$F_{h(n-1), (h-1)} = \frac{\sigma_{artean}^2}{\sigma_{barruan}^2}$$

Konparatu F_{taula} -rekin irizpide berdina jarraituz

14

ANOVA

ANOVA taula:

Bariazio iturria	Karratuen batura (SC/SS)	Askatasun graduak (gl/df)	Batezbesteko Karratua (CM/MS)	F
Lagin artean-"Intergrupos" (Faktorearen efektua)	SS_F	h-1	$MS_F = \frac{SS_F}{h-1}$	$F = \frac{MS_F}{MS_E}$
Lagin barruan-"Intragrupos" (Errorea)	SS_E	h (n-1)	$MS_E = \frac{SS_E}{nh-h}$	
TOTALA	SS_T	nh-1		

$$SS_F = n \sum_i (\bar{X}_i - \bar{X})^2 \quad SS_T = \sum_i \sum_j (X_{ij} - \bar{X})^2$$

$$SS_E = \sum_i \sum_j (X_{ij} - \bar{X}_i)^2 = SCT - SCF$$

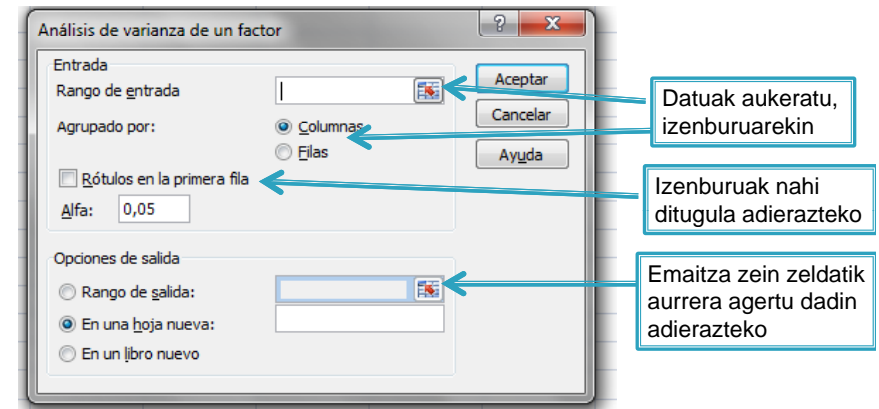
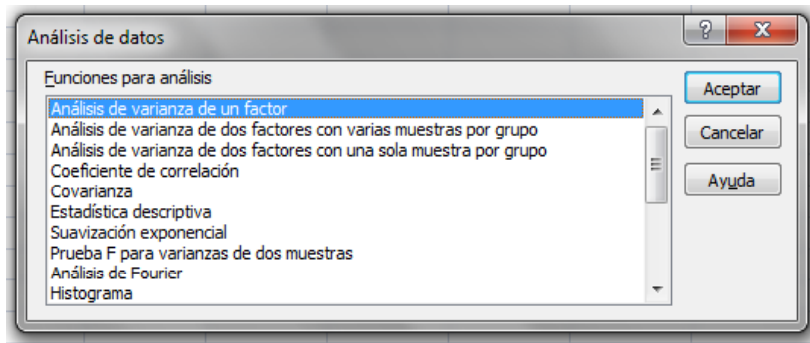
15

ANOVA

Fluoreszentiaren adibidearekin jarraituz.

Bariazio iturria	Karratuen batura (SC/SS)	Askatasun graduak (gl/df)	Batezbesteko Karratua (CM/MS)	F
Lagin artean-"Intergrupos" (Faktorearen efektua)	186	4-1=3	186/3=62	62/3=20.667
Lagin barruan-"Intragrupos" (Errorea)	24	12-4=8	24/8=3	
TOTALA	210	12-1=11		

16



Datuak Excel taula batean

Filetan

A	102	100	101
B	101	101	104
C	97	95	99
D	90	92	94

Zutabeetan

A	B	C	D
102	101	97	90
100	101	95	92
101	104	99	94

ANOVA EXCEL-ekin egina

Análisis de varianza de un factor				
RESUMEN				
Grupos	Cuenta	Suma	Promedio	Varianza
A	3	303	101	1
B	3	306	102	3
C	3	291	97	4
D	3	276	92	4

ANÁLISIS DE VARIANZA						
Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Entre grupos	186	3	62	20,6666667	0,00040015	4,06618056
Dentro de los grupos	24	8	3			
Total	210	11				

$$F_{kal} > F_{taula}$$

Diferentzia adierazgarria beraz, gordetzeko baldintzak eragina dute erreaktiboen egonkortasunean.



ZEIN DATU DIRA EZBERDINAK?

Ca edukia (mmol)					
Analista 1	Analista 2	Analista 3	Analista 4	Analista 5	
10.3	9.5	12.1	9.6	11.6	
9.8	8.6	13	8.3	12.5	
11.4	8.9	12.4	8.2	11.4	

$$DSM = t \sqrt{\frac{2 \cdot MS_E}{n}}$$

(Diferencia Significativa Mínima)

Análisis de varianza de un factor				
RESUMEN				
Grupos	Cuenta	Suma	Promedio	Varianza
Analista 1	3	31.5	10.5	0.67
Analista 2	3	27	9	0.21
Analista 3	3	37.5	12.5	0.21
Analista 4	3	26.1	8.7	0.61
Analista 5	3	35.5	11.83333333	0.34333333

ANÁLISIS DE VARIANZA						
Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Entre grupos	33.8026667	4	8.450666667	20.6786297	7.97096E-05	3.47804969
Dentro de los grupos	4.08666667	10	0.408666667			
Total	37.88933333	14				



ZEIN DATU DIRA EZBERDINAK?

$$DSM = t \sqrt{\frac{2 \cdot MS_E}{n}}$$

(Diferencia Significativa Mínima)

t (%95, nh-h)=2.228

DSM = 1.16

Batazbestekoak (ordenatuak)	Diferentzia		
x ₃ 12.5	x ₃ -x ₄ 3.8	> 1.16	Diferentzia ADIERAZGARRIA da
x ₅ 11.8	x ₂ -x ₄ 3.1	> 1.16	Diferentzia ADIERAZGARRIA da
x ₁ 10.5	x ₁ -x ₄ 1.8	> 1.16	Diferentzia ADIERAZGARRIA da
x ₂ 9.0	x ₅ -x ₄ 0.3	< 1.16	Diferentzia EZ da ADIERAZGARRIA
x ₄ 8.7			

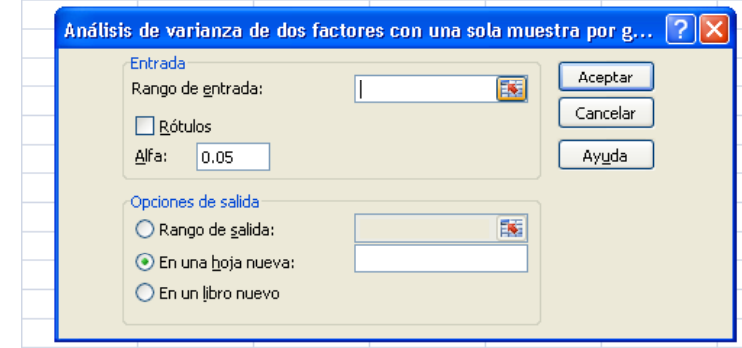
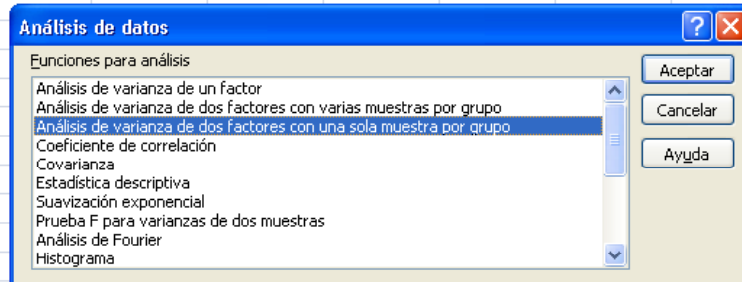


Aztertu nahi den prozesuaren konplexutasuna handituz gero, adibidez, bi aldagairen aldebereko eragina aztertu eta ondorioak ateratu nahi badugu, bi faktoreko ANOVA erabili behar da.

Adibidea. Mn-aren kontzentrazioa determinatu nahi da erreferentziako lagin batean. Determinazioak lau laborategitan aurrera eramane da eta, horretarako 3 prozesura erabili ditugu laborategian.

Prozedura analitikoa	Laborategia			
	1	2	3	4
1	2.01	1.96	1.99	2.03
2	1.97	2.05	2.04	1.99
3	2.05	2.06	2.11	2.12

Bildutako emaitza horiekin jakin nahi da ea erabilitako prozesuaren edo laborategien artean joera berezirik antzematen den.





Análisis de varianza de dos factores con una sola muestra por grupo

RESUMEN	Cuenta	Suma	Promedio	Varianza
1	4	7.99	1.9975	0.00089167
2	4	8.05	2.0125	0.00149167
3	4	8.34	2.085	0.00123333
1	3	6.03	2.01	0.0016
2	3	6.07	2.02333333	0.00303333
3	3	6.14	2.04666667	0.00363333
4	3	6.14	2.04666667	0.00443333

Zutabeak: $F_{kal} = 6.666 > F_{krit} = 5.143$

Diferentzia adierazgarria da prozedura horien artean, beraz zorizkoa ez den aldagarritasuna erakusten du.

ANÁLISIS DE VARIANZA

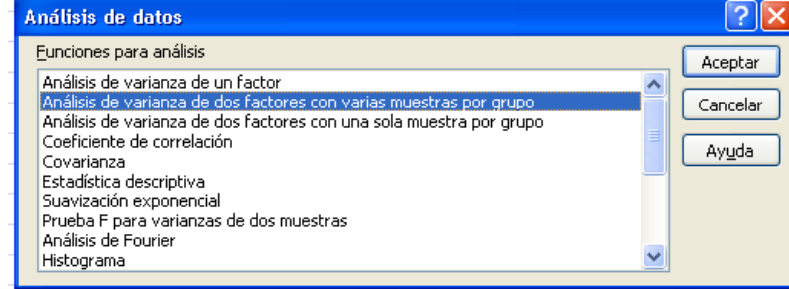
Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Filas	0.01751667	2	0.00875833	6.66596195	0.02989706	5.14325285
Columnas	0.00296667	3	0.00098889	0.75264271	0.55968236	4.75706266
Error	0.00788333	6	0.00131389			
Total	0.02836667	11				

Filak: $F_{kal} = 0.753 < F_{krit} = 4.757$

Beraz, laborategien arteko desberdintasunak zorizkoak dira.



Gerta daiteke taldean lagin bat baino gehiago egotea orduan hemen behekoa erabiliko genuke



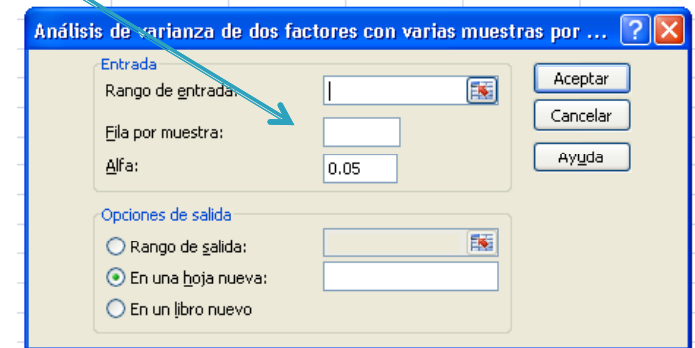
Adibidea. Laborategi arteko elkarlaneko esperimantu batean ikatzean dagoen artseniko edukia determinatu nahi da. Determinazioak 3 laborategitan egin eta bakoitzari 3 lagin desberdin bidali zaizkie. Laborategi bakoitzean analisi bikoitzak egin dira.

		Laborategiak		
		1	2	3
Ikatzak	1	5.1 ; 5.1	5.3 ; 5.4	5.3 ; 5.1
	2	5.8 ; 5.4	5.4 ; 5.9	5.2 ; 5.5
	3	6.5 ; 6.1	6.6 ; 6.7	6.5 ; 6.4

Konprobatu lagin eta laborategi artean interakzioak ez daudela.



Zenbat lerro lagineko





Datuak nola adierazi Excel orri batean.
Lagin bateko analisi edo datu kopurua lerrotan adieraziko da.
Segun eta zenbat datu, hainbeste lerro.

Laborategiak			
Ikatzak	1	2	3
1	5.1	5.3	5.3
	5.1	5.4	5.1
2	5.8	5.4	5.2
	5.4	5.9	5.5
3	6.5	6.6	6.5
	6.1	6.7	6.4



ANÁLISIS DE VARIANZA						
Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Muestra	5.06777778	2	2.53388889	62.4794521	5.281E-06	4.25649473
Columnas	0.18777778	2	0.09388889	2.31506849	0.15446863	4.25649473
Interacción	0.10222222	4	0.02555556	0.63013699	0.65331484	3.63308851
Dentro del grupo	0.365	9	0.04055556			
Total	5.72277778	17				

Filak: $F_{kal} = 62.48 > F_{krit} = 4.256$
Beraz, lagin arteko diferentziak adierazgarriak dira. Laginak eraba ezberdinak dira

Zutabeak: $F_{kal} = 2.315 < F_{krit} = 4.256$
Beraz, laborategien arteko desberdintasunak zorizkoak dira.

Interakzioa:
 $F_{kal} = 0.653 < F_{krit} = 3.633$

laborategi eta laginen artean ez dago interakziorik, zorizkoak dira,

