Anova

Ejemplo 5:

Se quiere determinar si la dosis de determinado tratamiento (Baja, Media, Alta) influye en el tiempo de sueño (en minutos) de los que la consumen.

Se plantea el problema de determinar si la tiempos de sueño varían o no, según la dosis del medicamento. ¿Puede afirmarse que los tiempos de sueño no varían según la dosis de medicamento?, use un nivel de significancia de 5%.

Para responder al problema planteado, se cuenta con los siguientes datos tomados de varios voluntarios expuestos a la aplicación de 3 dosis de medicamento (Alta, Media, Baja), registrando el número de minutos que duerme, una vez injerido la dosis:

DOSIS		
Baja	Media	Alta
67	96	74
69	98	24
72	130	15
79	65	33
		17

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

$H_1: \exists i, j \mid \mu_i \neq \mu_j$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.} \quad s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

n_i	4	4	5
T_i	287	389	163
$\frac{T_i^2}{n_i}$	20592,25	37830,25	5313,8

$N = 13$ $\quad \sum y_{ij}^2 = 68275$

$T = 839$ $\quad k = 3$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$$

$$SSA = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 = SST - SSA$$

$$SST = 68275 - \frac{839^2}{13} = 14127,23077$$

$$SSA = (20592,25 + 37830,25 + 5313,8) - \frac{839^2}{13} = 9588,530769$$

$$SSE = 14127,23077 - 9588,530769 = 4538,7$$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.} \quad s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$S_1^2 = \frac{9588,530769}{3-1} \qquad S^2 = \frac{4538,7}{13-3}$$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.}$$

$$f_{obs} = \frac{\frac{9588,530769}{3-1}}{\frac{9538,7}{13-3}} = 10,5631$$

$v = 2, 10$

$$f_c = f_{0.05,2,10} = 4,10282$$

$$\boxed{\begin{array}{l} \text{R/ Como } f_{obs} = 10,5631 > f_c = 4,10282 \\ \text{Se rechaza } H_0 \end{array}}$$

Valor P

$P(F > 10,5631) = 0,00342$ $\quad \alpha = 0,05$

$$\boxed{\begin{array}{l} \text{R/ Como } P = 0,00342 < \alpha = 0,05 \\ \text{Se rechaza } H_0 \end{array}}$$

Cada año, los miembros del equipo de atletismo de una universidad se dividen al azar en tres grupos que entrenan con métodos diferentes. El primer grupo realiza largos recorridos a ritmo pausado, el segundo grupo realiza series cortas de alta intensidad y el tercero trabaja en el gimnasio con pesas y en bicicleta estacionaria con pedaleo de alta frecuencia. Después de un mes de entrenamiento se realiza un test de rendimiento de la prueba de 110m con vallas. Seguidamente se muestran los tiempos obtenidos en el test en una muestra de 11 miembros:

MetodoI	*MetodoII*	*MetodoIII*
15	14	13
16	13	12
14	15	14
15	16	

A un nivel de significancia del 5%, ¿Puede considerarse que los tres métodos producen resultados equivalentes? (8 puntos)

$H_0: u_1 = u_2 = u_3$

$H_1: \exists\, i,j \mid u_i \neq u_j$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l. } \quad s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$$

$$SSA = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 = SST - SSA$$

n_i	4	4	3
T_i	60	58	39
$\frac{T_i^2}{n_i}$	900	841	507

$N = 11 \quad k = 3$

$T = 157 \quad y_{ij}^2 = 2257$

$$SST = 2257 - \frac{157^2}{11} = \frac{178}{11}$$

$$SSA = (900 + 841 + 507) - \frac{157^2}{11} = \frac{79}{11}$$

$$SSE = \frac{178}{11} - \frac{79}{11} = 9$$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l. } \quad s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$S_1^2 = \frac{\frac{79}{11}}{3-1} \qquad s^2 = \frac{9}{11-3}$$

$\alpha = 0{,}05$

$V = 3-1, 11-3$

$V = 2, 8$

$$f_{obs} = \frac{\frac{\frac{79}{11}}{3-1}}{\frac{9}{11-3}} \approx 3{,}19$$

$f_c = 0.05, 2, 8 = 4{,}45897$

$\boxed{\text{R/ Como } f_{obs} = 3{,}19 < f_c = 4{,}45897 \text{ No se rechaza } H_0}$

$f_{obs} \approx 3{,}19 \qquad v = 2{,}8 \qquad \alpha = 0{,}05$

$P(F > 3{,}19) = 0{,}09579$

R/ Como $P = 0{,}09579 > \alpha = 0{,}05$
No se rechaza H_0

Ejemplo

(Anova) Con el anuncio del cierre parcial de la carretera de Circunvalación los usuarios de esta vía se han visto en la necesidad de tomar rutas alternas. Los conductores pueden optar por seguir utilizando la ruta Circunvalación (CI), cruzar el centro de San José (SJ) o utilizar algún camino rural (CR). Para tener un panorama sobre los nuevos tiempos (en horas) de traslado desde Alajuela hasta Cartago se realizó un muestreo y se obtuvieron los siguientes datos

CI	2.2	2.1	1.8	1.5	1.9	1.7	
SJ	1.3	1.7	2.1	1.3	1.6	2.0	1.5
CR	2.6	2.4	1.9	2.5	2.1		

Con una significancia de 0.1, ¿puede asegurarse que los tiempos medios de traslado a Cartago desde Alajuela varían según la ruta que se utilice? (6 puntos)

$H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: \exists\, i,j \mid \mu_i \neq \mu_j$

CI	SJ	CR
2,2	1,3	2,6
2,1	1,7	2,4
1,8	2,1	1,9
1,5	1,3	2,5
1,9	1,6	2,1
1,7	2,0	
	1,5	

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.}$$

$$s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$$

$$SSA = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 = SST - SSA$$

n_i	6	7	5
T_i	11,2	11,5	11,5
$\frac{T_i^2}{n_i}$	$\frac{1568}{75}$	$\frac{529}{28}$	$\frac{529}{20}$

$N = 18 \qquad k = 3$

$T = 34{,}2 \qquad y_{ij}^2 = 67{,}52$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.}$$

n_i	6	7	5
T_i	11,2	11,5	11,5
$\frac{T_i^2}{n_i}$	$\frac{1568}{75}$	$\frac{529}{28}$	$\frac{529}{20}$

$N = 18$ $\quad k = 3$

$T = 34,2$ $\quad y^2_{ij} = 67,52$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$$

$$SSA = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 = SST - SSA$$

$$SST = 67,52 - \frac{34,2^2}{18} = 2,54$$

$$SSA = \left(\frac{1568}{75} + \frac{529}{28} + \frac{529}{20}\right) - \frac{34,2^2}{18} = \frac{1333}{1050}$$

$$SSE = 2,54 - \frac{1333}{1050} = \frac{667}{525}$$

$$s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$S_1^2 = \frac{\frac{1333}{1050}}{3-1} \qquad S^2 = \frac{\frac{667}{525}}{18-3}$$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.}$$

$$F_{obs} = \frac{\frac{\frac{1333}{1050}}{3-1}}{\frac{\frac{667}{525}}{18-3}} = 7,49$$

$\alpha = 0,10$

$v = 3-1,\ 18-3$

$v = 2,\ 15$

$$F_c = F_{0,10,\,2,\,15} = 2,69$$

$$\boxed{\text{B) Como } F_{obs} = 7,49 > F_c = 2,69 \text{ se rechaza } H_0}$$

Con valor P

$f_{obs} = 7,49$ $\qquad v = 2, 15$ $\qquad \alpha = 0,1$

$P(f > 7,49) = 0,00555$

$\boxed{\text{R/ Como } P = 0,00555 < \alpha = 0,01 \text{ Se rechaza } H_0}$

Ejercicio 49. Todos los domingos, durante dos horas, don Juan sale a pescar y utiliza una de sus tres cañas de pescar. Él desea determinar si el número de peces que pesca por domingo es independiente de la caña que utilice. Para ello registró durante 15 domingos el número de peces obtenidos y la caña utilizada; los datos son:

Caña 1	Caña 2	Caña 3
12	10	16
10	17	14
18	16	16
12	13	11
14		20
		21

Con un nivel de significancia de 0.05, ¿existe evidencia suficiente de que el número de peces obtenidos cada domingo depende de la caña utilizada?

$H_0: u_1 = u_2 = u_3$

$H_1: \exists i,j \mid u_i \neq u_j$

n_i	5	4	6
T_i	66	56	98
T_i^2	4356	784	4802
n_i	5		3

$N = 15 \qquad y^2_{ij} = 3392$

$T = 220 \qquad k = 3$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l. } s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$$

$$SSA = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 = SST - SSA$$

$N = 15 \quad y^2{}_{ij} = 3392$

$T = 220 \quad k = 3$

$$SST = 3392 - \frac{220^2}{15} = \frac{496}{3}$$

$$SSA = \left(\frac{9356 + 784}{5} + \frac{4802}{3}\right) - \frac{220^2}{15} = \frac{146}{5}$$

$$SSE = \frac{496}{3} - \frac{146}{5} = \frac{2042}{15}$$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.}$$

$$s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$S_1^2 = \frac{\frac{146}{5}}{3-1} \qquad S^2 = \frac{\frac{2042}{15}}{15-3}$$

$$F_{obs} = \frac{\frac{\frac{146}{5}}{3-1}}{\frac{\frac{2042}{15}}{15-3}} = 1{,}2869$$

$\alpha = 0{,}05$

$V = 3-1, 15-3$

$V = 2, 12$

$$F_c = F_{0.05, 2, 12} = 3{,}8529$$

R/ Como $F_{obs} = 1{,}2869 < F_c = 3{,}8529$
No se rechaza H_0

3. Durante un semestre un estudiante recibió calificaciones en diversas materias. Seguidamente se muestran algunas de las calificaciones obtenidas:

Matemáticas	72	80	83	75	
Ciencias	81	74	77		
Inglés	88	82	90	87	80
Economía	74	71	77	70	

Determine si existe diferencia significativa entre el rendimiento promedio de las materias que el estudiante cursó este semestre.

$H_0: u_1 = u_2 = u_3$

$H_1: \exists i, j \mid u_i \neq u_j$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.}$$

n_i	4	3	5	4
T_i	310	232	427	292
$\frac{T_i^2}{n_i}$	24025	$\frac{53824}{3}$	36465,8	21316

$N = 16$ $\quad y_{ij}^2 = 99947$

$T = 1261$ $\quad k = 4$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$$

$$SSA = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 = SST - SSA$$

$$SST = 99947 - \frac{1261^2}{16} = \frac{9031}{16}$$

$$SSA = \left(24025 + \frac{53824}{3} + 36465,8 + 21316\right) - \frac{1261^2}{16} = 365,5708333$$

$$SSE = \frac{9031}{16} - 365,5708333 = 198,8666667$$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.}$$

$$s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$f_{obs} = \frac{\frac{365,5708333}{4-1}}{\frac{198,8666667}{16-4}} = 7,353$$

$\alpha = 0,05$

$V = 3,12$

$f_c = f_{0.05,\,3,12} = 3,49029$

$\boxed{\text{Como } f_{obs} > f_c, \text{ se rechaza } H_0}$

5. Se compararon estudiantes avanzados en la carrera de Computación de tres universidades distintas con respecto a sus conocimientos en sistemas operativos. A cada estudiante se le aplicó una prueba y se registraron sus notas. Los datos son los siguientes:

Universidad		
U1	U2	U3
67	96	74
90	80	94
72	71	65
79	86	53
		87

¿Puede afirmarse que no hay diferencia entre los conocimientos sobre sistemas operativos en las tres universidades, a un nivel de significancia del 5 %? $R/\quad f_{obs} \approx 0.500742,\ f_c = 4.103$, se acepta H_0

$H_0: u_1 = u_2 = u_3$

$H_1: \exists i,j \mid u_i \neq u_j$

n_i	4	4	5
T_i	308	333	373
$\frac{T_i^2}{n_i}$	23716	27722,25	27825,8

$N = 13$

$T = 1014$

$y^2_{ij} = 80982$

$k = 3$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$$

$$SSA = \sum_{i=1}^{k} n_i(\bar{y}_i-\bar{y})^2 = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 = SST - SSA$$

$$SST = 80982 - \frac{1014^2}{13} = 1890$$

$$SSA = (23716 + 27722{,}25 + 27825{,}8) - \frac{1014^2}{13} = 172{,}05$$

$$SSE = 1890 - 172{,}05 = 1717{,}95$$

$$F = \frac{S_1^2}{S^2} \text{ con } (k-1, N-k) \text{ g.l.}$$

$$s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$f_{obs} = \frac{\frac{172{,}05}{3-1}}{\frac{1717{,}95}{13-3}} = 0{,}500742$$

$V = 2, 10$

$$f_c = F_{0{,}05,\,2,\,10} = 4{,}10282$$

$\boxed{\text{R/ Como } F_{obs} < F_c\text{, No se rechaza } H_0}$