LA RAZÓN F DE FISHER CON EXCEL, WINSTATS Y GEOGEBRA

A diferencia de otras pruebas de medias que se basan en la diferencia existente entre dos valores, el análisis de varianza emplea la razón de las estimaciones, dividiendo la estimación intermediante entre la estimación interna

$$Razón\ F = \frac{s_x^2}{s_w^2} = \frac{ns_{\bar{x}}^2}{(s_1^2 + s_2^2 + s_3^2 + \cdots s_k^2)/k}$$

Esta razón F fue creada por Ronald Fisher (1890-1962), matemático británico, cuyas teorías estadísticas hicieron mucho más precisos los experimentos científicos. Sus proyectos estadísticos, primero utilizados en biología, rápidamente cobraron importancia y fueron aplicados a la experimentación agrícola, médica e industrial. Fisher también contribuyó a clarificar las funciones que desempeñan la mutación y la selección natural en la genética, particularmente en la población humana.

El valor estadístico de prueba resultante se debe comparar con un valor tabular de F, que indicará el valor máximo del valor estadístico de prueba que ocurría si H_0 fuera verdadera, a un nivel de significación seleccionado. Antes de proceder a efectuar este cálculo, se debe considerar las características de la distribución F

i) Características de la distribución F

- Existe una distribución F diferente para cada combinación de tamaño de muestra y número de muestras. Por tanto, existe una distribución F que se aplica cuando se toman cinco muestras de seis observaciones cada una, al igual que una distribución F diferente para cinco muestras de siete observaciones cada una. A propósito de esto, el número distribuciones de muestreo diferentes es tan grande que sería poco práctico hacer una extensa tabulación de distribuciones. Por tanto, como se hizo en el caso de la distribución t, solamente se tabulan los valores que más comúnmente se utilizan. En el caso de la distribución F, los valores críticos para los niveles 0,05 y 0,01 generalmente se proporcionan para determinadas combinaciones de tamaños de muestra y número de muestras.

- La distribución es continua respecto al intervalo de 0 a + ∞. La razón más pequeña es 0. La razón no puede ser negativa, ya que ambos términos de la razón F están elevados al cuadrado. Por otra parte, grandes diferencias entre los valores medios de la muestra, acompañadas de pequeñas variancias muestrales pueden dar como resultado valores extremadamente grandes de la razón F.

- La forma de cada distribución de muestreo teórico F depende del número de grados de libertad que estén asociados a ella. Tanto el numerador como el denominador tienen grados de libertad relacionados.

ii) Determinación de los grados de libertad

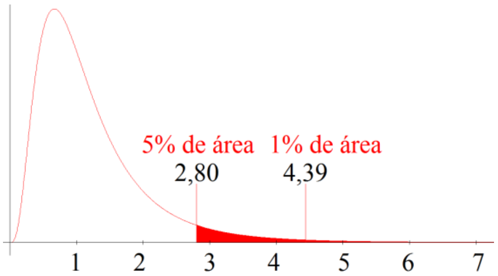
Los grados de libertad para el numerador y el denominador de la razón F se basan en los cálculos necesarios para derivar cada estimación de la variancia de la población. La *estimación intermediante* de variancia (numerador) comprende la división de la suma de las diferencias elevadas al cuadrado entre el número de medias (muestras) menos uno, o bien, k - 1. Así, **k - 1** *es el número de grados de libertad para el numerador.*

En forma semejante, el calcular cada variancia muestral, la suma de las diferencias elevadas al cuadrado entre el valor medio de la muestra y cada valor de la misma se divide entre el número de observaciones de la muestra menos uno, o bien, n - 1. Por tanto, el promedio de las variancias muestrales se determina dividiendo la suma de las variancias de la muestra entre el número de muestras, o k. *Los grados de libertad para el denominador son entonces*, **k(n -l).**

iii) Uso de la tabla de F del análisis de variancia (ANOVA)

En la tabla 5 se ilustra la estructura de una tabla de F para un nivel de significación de 0,01 o 1% y 0,05 o 5%. Se obtiene el valor tabular, localizando los grados de libertad del numerador n_1 (que se listan en la parte superior de la tabla), así como los del denominador n_2 (que se listan en una de las columnas laterales de la tabla) que corresponden a una situación dada. Utilizando el nivel de significación de 0,05 para $n_1 = 7\ y\ n_2 = 3$ grados de libertad, el valor de F es 8,89

TABLA Nº 5
DISTRIBUCIÓN F DE FISHER



Ejemplos:

$\boldsymbol{Para\ n_1 = 9\ ;\ n_2 = 12\ grados\ de\ libertad}$

$P(F > 2{,}80) = 0{,}05 = 5\%$

$\boldsymbol{P(F > 4{,}39) = 0{,}01 = 1\%}$

5% (normal) **y 1% (negritas)**

$\boldsymbol{n_1}$ **= grados de libertad del numerador**

$\boldsymbol{n_2}$ **= grados de libertad del denominador**

n_2 \ n_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**15**	**20**	**25**	**50**	**100**
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	245,95	248,01	249,26	251,77	253,04
	4052,2	**4999,5**	**5403,4**	**5624,6**	**5763,6**	**5859,0**	**5928,4**	**5981,1**	**6022,5**	**6055,8**	**6157,3**	**6208,7**	**6239,8**	**6302,5**	**6334,1**
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,43	19,45	19,46	19,48	19,49
	98,50	**99,00**	**99,17**	**99,25**	**99,30**	**99,33**	**99,36**	**99,37**	**99,39**	**99,40**	**99,43**	**99,45**	**99,46**	**99,48**	**99,49**
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66	8,63	8,58	8,55
	34,12	**30,82**	**29,46**	**28,71**	**28,24**	**27,91**	**27,67**	**27,49**	**27,35**	**27,23**	**26,87**	**26,69**	**26,58**	**26,35**	**26,24**

iv) Cálculo de la razón F a partir de datos muestrales

$$F_{prueba} = \frac{estimación\ intermediante\ de\ variancia}{estimación\ interna\ de\ variancia}$$

$$F_{prueba} = \frac{s_x^2}{s_w^2} = \frac{ns_{\bar{x}}^2}{(s_1^2 + s_2^2 + s_3^2 + \cdots s_k^2)/k}$$

Para calcular F se debe seguir el siguiente procedimiento

1) Calcular la estimación interna (Denominador)

1.1) Determinar la variancia de cada muestra, utilizando la fórmula

$$variancia = s^2 = \frac{\sum(x_{i-}\bar{x})^2}{n-1}$$

1.2) Obtener la estimación interna de variancia (variancia promedio de la muestra), mediante la fórmula

$$s_w^2 = \frac{s_1^2 + s_2^2 + s_3^2 + \ldots \ldots \ldots \ldots s_k^2}{k}$$

2) Calcular la estimación intermediante (Numerador)

2.1) Calcular la variancia de la medias muestrales, utilizando la fórmula

$$s_{\bar{x}}^2 = \frac{\sum(\bar{x} - \bar{\bar{x}})^2}{k - 1}$$

2.2) Multiplicar la variancia de la medias muestrales por n

$$ns_{\bar{x}}^2$$

3) Razón F

$$F_{prueba} = \frac{s_x^2}{s_w^2}$$

Las hipótesis Nula y Alternativa son:
H_0: Todas las proporciones de la población son iguales.
H_1: No todas las proporciones de la población son iguales.

Ejemplo ilustrativo

Los pesos en kg por 1,7 m de estatura se ilustran en la siguiente tabla. La finalidad es determinar si existen diferencias reales entre las cuatro muestras. Emplear un nivel de significación de 0,05

Observación	Muestra			
	1	2	3	4
1	70	74	68	75
2	75	77	70	70
3	74	70	65	73
4	72	80	60	72
5	68	72	72	71
6	59	76	73	72

Solución:

Las hipótesis Nula y Alternativa son:
H_0: Todas las proporciones de la población son iguales.
H_1: No todas las proporciones de la población son iguales.

Calculando los grados de libertad de numerador se tiene:
$k - 1 = 4 - 1 = 3$
Calculando los grados de libertad del denominador se tiene:
$k(n - 1) = 4(6 - 1) = 20$
Con 3 grados de libertad en el numerador, 20 grados de libertad en el denominador y con un nivel de significación $\alpha = 0{,}05$ con lectura la tabla se obtiene $F_{tabla} = 3{,}10$
Para calcular F_{prueba} se procede de la siguiente manera:

Calculando las medias aritméticas se obtiene:

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x}_1 = \frac{70 + 75 + 74 + 72 + 68 + 59}{6} = \frac{418}{6} = 69{,}667$$

$$\bar{x}_2 = \frac{74 + 77 + 70 + 80 + 72 + 76}{6} = \frac{449}{6} = 74{,}833$$

$$\bar{x}_3 = \frac{68 + 70 + 65 + 60 + 72 + 73}{6} = \frac{408}{6} = 68$$

$$\bar{x}_4 = \frac{75 + 70 + 73 + 72 + 71 + 72}{6} = \frac{433}{6} = 72{,}167$$

Se llena la siguiente tabla para calcular las varianzas muestrales:

	Muestra				

Observación	1	2	3	4	$(x_1-\bar{x}_1)^2$	$(x_2-\bar{x}_2)^2$	$(x_3-\bar{x}_3)^2$	$(x_4-\bar{x}_4)^2$
1	70	74	68	75	0,111	0,694	0	8,026
2	75	77	70	70	28,441	4,696	4	4,696
3	74	70	65	73	18,775	23,358	9	0,694
4	72	80	60	72	5,443	26,698	64	0,028
5	68	72	72	71	2,779	8,026	16	1,361
6	59	76	73	72	113,785	1,362	25	0,028
Total	418	449	408	433	169,334	64,834	118	14,833

Remplazando los datos en la fórmula de la varianza se obtienen las varianzas de las 4 muestras.

$$s^2=\frac{\sum(x_i-\bar{x})^2}{n-1}$$
$$s_1^2=\frac{169{,}334}{5}=33{,}867$$
$$s_2^2=\frac{64{,}834}{5}=12{,}967$$
$$s_3^2=\frac{118}{5}=23{,}6$$
$$s_4^2=\frac{14{,}833}{5}=2{,}967$$

Calculando la estimación interna de varianza se obtiene:

$$s_w^2=\frac{s_1^2+s_2^2+s_3^2+\cdots+s_k^2}{k}$$

$$s_w^2=\frac{33{,}867+12{,}967+23{,}6+2{,}967}{4}=\frac{73{,}401}{4}=18{,}35$$

Para calcular la estimación intermediante de varianza primero se calcular la varianza de las medias aritméticas

$$s_{\bar{x}}^2=\frac{\sum(\bar{x}-\bar{\bar{x}})^2}{k-1}$$

Para calcular la varianza de las medias aritméticas se calcula la media aritmética de las medias aritméticas, la cual es:

$$\bar{\bar{x}}=\frac{\sum\bar{x}_i}{k}=\frac{69{,}667+74{,}833+68+72{,}167}{4}=\frac{284{,}667}{4}=71{,}167$$

Se llena la siguiente tabla:

$\bar{x}$	$(\bar{x}-\bar{\bar{x}})^2$
69,667	2,25
74,833	13,44
68	10,03
72,167	1
Total	26,72

Se remplaza los datos de la tabla para calcular varianza de las medias aritméticas

$$s_{\bar{x}}^2=\frac{\sum(\bar{x}-\bar{\bar{x}})^2}{k-1}=\frac{26{,}72}{3}=8{,}907$$

Calculando la estimación intermediante de varianza se obtiene:
$s_x^2 = n \cdot s_{\bar{x}}^2 = 6 \cdot 8{,}907 = 53{,}44$

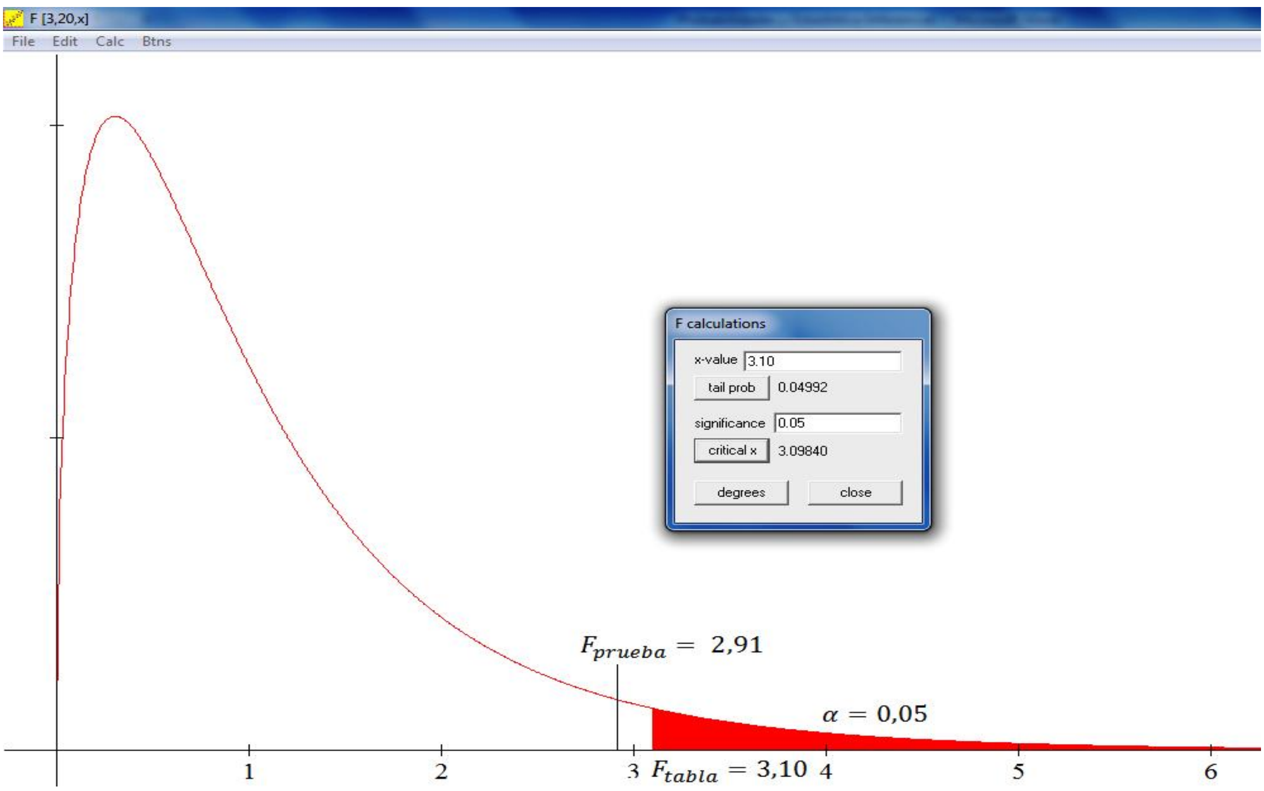
Finalmente calculando F_{prueba} se tiene:

$$F_{prueba} = \frac{s_x^2}{s_w^2} = \frac{53{,}44}{18{,}35} = 2{,}91$$

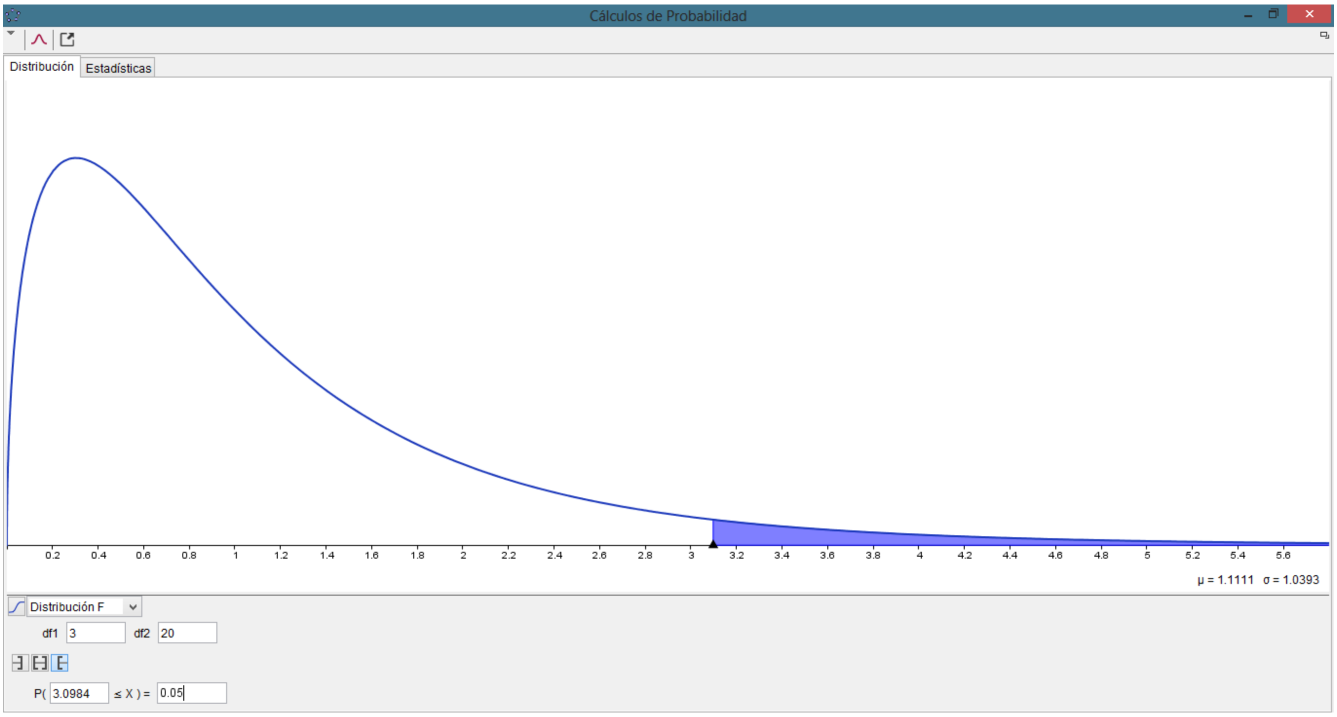
Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H
1		Muestra k						
2	Observación n	1	2	3	4	$\bar{x}$		
3	1	70	74	68	75	69,667	=PROMEDIO(B3:B8)	
4	2	75	77	70	70	74,833	=PROMEDIO(C3:C8)	
5	3	74	70	65	73	68,000	=PROMEDIO(D3:D8)	
6	4	72	80	60	72	72,167	=PROMEDIO(E3:E8)	
7	5	68	72	72	71			
8	6	59	76	73	72			
9	s^2	33,87	12,97	23,60	2,97			
10		=VAR.S(B3:B8)	=VAR.S(C3:C8)	=VAR.S(D3:D8)	=VAR.S(E3:E8)			
11								
12	Estimación interna		18,35	=PROMEDIO(B9:E9)	$s_w^2 = \frac{s_1^2 + s_2^2 + s_3^2 + \ldots\ldots\ldots s_k^2}{k}$			
13	Varianza de las medias muestrales		8,91	=VAR.S(F3:F6)	$s_{\bar{x}}^2 = \frac{\sum(\bar{x} - \bar{\bar{x}})^2}{k-1}$			
14	n		6	=CONTAR(A3:A8)				
15	Estimación intermediante		53,44	=C14*C13	$s_x^2 = n \cdot s_{\bar{x}}^2$			
16	F_{prueba}		2,91	=C15/C12	$F_{prueba} = \frac{s_x^2}{s_w^2}$			
17								
18	k		4	=CONTAR(B2:E2)				
19	$k-1$		3	=C18-1				
20	$k(n-1)$		20	=C18*(C14-1)				
21	α		0,05					
22	F_{tabla}		3,10	=INV.F.CD(C21;C19;C20)				

La gráfica elaborada en Winstats y Paint se muestra en la siguiente figura:



La gráfica elaborada en GeoGebra se muestra en la siguiente figura:



Decisión: Como F_{prueba} es menor que F_{tabla}, H_0 se aprueba, por lo tanto no existen diferencias reales en los pesos de las 4 muestras, es decir, todas las proporciones de la población son iguales**.**