

Diseño completamente al azar



Dr. Jesús Alberto Mellado Bosque

Partes de la varianza

La varianza se va a separar en dos partes, la suma de cuadrados y los grados de libertad.

$$\sum \frac{(x_i - \bar{x})^2}{n - 1} \begin{matrix} \longrightarrow \sum (x_i - \bar{x})^2 & \text{Suma de Cuadrados (SC)} \\ \longrightarrow n - 1 & \text{Grados de libertad (GL)} \end{matrix}$$

Varianza

ACLARACIÓN: Estrictamente una suma de cuadrados sería: $\sum x^2$, y la suma de cuadrados ajustada sería: $\sum (x_i - \bar{x})^2$, Pero para fines prácticos, en los diseños experimentales se le ha llamado solamente Suma de cuadrados a la suma de cuadrados ajustada

La suma de cuadrados del total se separa en la suma de cuadrados de los tratamientos y la suma de cuadrados del error.

$$\sum (x_i - \bar{x})^2 \begin{matrix} \longrightarrow \sum (x_i - \bar{x})^2 & \text{Suma de cuadrados (SC) del tratamiento} \\ \longrightarrow \sum (x_i - \bar{x})^2 & \text{Suma de cuadrados (SC) del error} \end{matrix}$$

Suma de cuadrados del total

Se calculan los grados de libertad que corresponden a la separación de la suma de cuadrados.

- $n - 1$ Grados de libertad del total (GL)
- $t - 1$ Grados de libertad de tratamientos, donde "t" es el número de tratamientos
- $(n - 1) - (t - 1)$ Grados de libertad del error experimental

La Suma de cuadrados del tratamiento se divide entre sus grados de libertad y se forman los Cuadrados medios del tratamiento

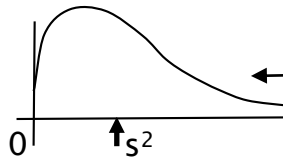
$$\text{Cuadrados medios tratamiento} = \frac{\text{SC tratamiento}}{\text{GL tratamiento}}$$

La Suma de cuadrados del error se divide entre sus grados de libertad y se forman los Cuadrados medios del error

$$\text{Cuadrados medios del error} = \frac{\text{SC error}}{\text{GL error}}$$

Distribución "F" de Fisher

La varianza se distribuye ji-cuadrada



Los cuadrados medios son una varianza y se distribuye ji-cuadrada

El objetivo es saber si la varianza de los tratamientos es mayor a la varianza del error. La forma de hacer la comparación es dividir los cuadrados medios de tratamientos entre los cuadrados medios del error. El resultado se llama F_c ("F" calculada)

"F" calculada

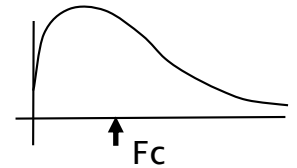
$$F_c = \frac{CM_{Tratamientos}}{CM_{Error}}$$

Si $F_c > 1$, significa que CM de tratamientos es mayor que los CM del error

Si $F_c = 1$, significa que CM de tratamientos es igual que los CM del error

Si $F_c < 1$, significa que CM de tratamientos es menor que los CM del error

Cuando se divide una distribución ji-cuadrada (de tratamientos) entre otra ji-cuadrada (error) el resultado es una gráfica similar, pero de nombre "F de Fisher" (Su nombre se lo debe a su autor)



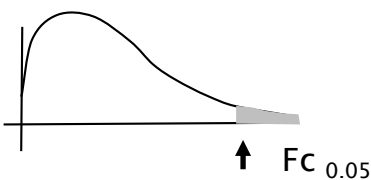
La distribución "F" tiene dos tipos de grados de libertad. Los de tratamientos y los del error. En la tabla "F" se tienen que buscar su valor seleccionando en las columnas los grados de libertad de tratamientos y en el renglón los del error.

	Grados de libertad de tratamientos
Grados de libertad del error	

Ejemplo:

	2 GL trat
9 GL error	4.26

En la tabla "F" del 0.05, proporciona el valor en el eje "x" cuando el área (probabilidad) a la derecha es 0.05.



Cuadro ANOVA (ANVA)

La forma para reportar los resultados de un experimento es mediante un cuadro llamado ANOVA (Analysis of Variance) o ANVA en español.

En el lado izquierdo se anota la denominación de cada una de las varianzas, nótese que la suma de los dos primeros renglones se van a sumar y formar el valor del tercero. A estas etiquetas también se llaman “Fuentes de variación”

En la primera columna se anotan los grados de libertad (GL) que corresponden a cada renglón

En la segunda columna se anotan la suma de cuadrados (SC) que corresponden a cada renglón

	GL	SC	CM	Fc	Ft
tratamientos					
Error					
Total					

JM

En la tercera columna se anotan los cuadrados medios (CM). El tercer renglón no se usa

En la cuarta columna se anota el valor de la F calculada. Solo se usa el primer renglón

En la quinta columna se anota el valor de la F de tablas. **IMPORTANTE:** Si la tabla la hace un paquete estadístico, aquí se pone la probabilidad de que Fc ocurra si H_0 sea cierta

Introducción

El diseño completamente al azar es un prueba basada en el análisis de varianza, en donde la varianza total se descompone en la “varianza de los tratamientos” y la “varianza del error”. El objetivo es determinar si existe un diferencia significativa entre los tratamientos, para lo cual se compara si la “varianza del tratamiento” contra la “varianza del error” y se determina si la primera es lo suficientemente alta según la distribución F.

La distribución debe ser normal, los errores deben ser independientes, los efectos deben ser aditivos, los tratamientos se asignan aleatoriamente.

Históricamente este modelo se llamaba “ONE WAY ANOVA”, así que ese nombre aparece en muchos libros y videos de internet que se basan en libros antiguos.

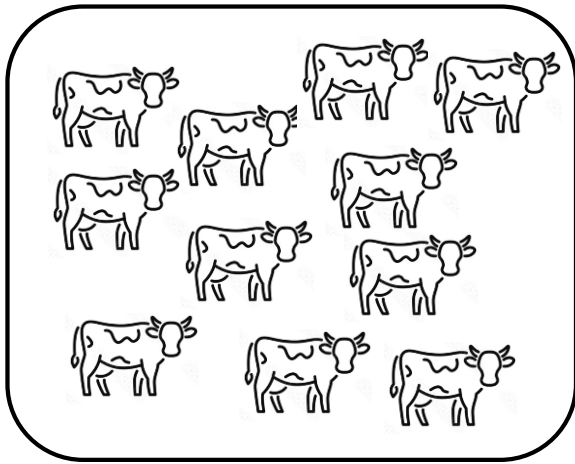
Características del diseño

Se definen los t tratamientos que se van a aplicar a las n unidades experimentales, de tal forma que a r unidades experimentales les va a corresponder un tipo de tratamiento.

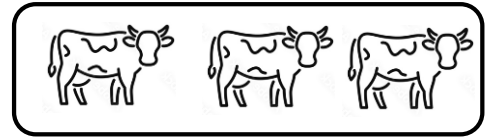
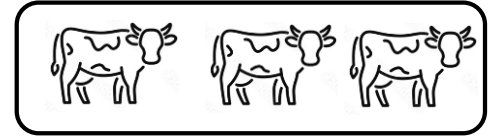
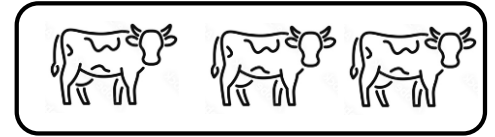
Las unidades experimentales se sortean para la asignación a cada tratamiento.

Se define la variable a medir.

Se tienen N unidades experimentales



Se selecciona cada unidad experimental completamente al azar y también al azar se asigna a un tratamiento



M

Modelo estadístico y análisis de varianza

Acción

Ejemplo

1

Se definen los tratamientos y se sortean las unidades experimentales. Se realiza el experimento y se recopilan los datos. Suponiendo que son tres tratamientos y cuatro repeticiones, y que se midió el crecimiento de ciertas plantas, los resultados se acomodan en una tabla.

	Rep 1	Rep 2	Rep 3	Rep 4
Trat 1	13	12	13	11
Trat 2	11	14	13	12
Trat 3	9	8	11	9
			Total	136

y_{32} valor del tratamiento 3 repetición 2

2

Se suman todos los valores de las unidades experimentales. A ese valor se le llamará $y_{..}$. Se obtiene el cuadrado de todos los valores de la unidades experimentales y luego se suman, a ese valor se le llamará $\sum y_{ij}^2$

	Rep 1	Rep 2	Rep 3	Rep 4
Trat 1	13	12	13	11
Trat 2	11	14	13	12
Trat 3	9	8	11	9
			Total	136

$y_{..}$

	Rep 1	Rep 2	Rep 3	Rep 4
Trat 1	169	144	169	121
Trat 2	121	196	169	144
Trat 3	81	64	121	81
			Total	1580

3 Se calcula la suma de cuadrados del total con la fórmula:
 Suma Cuadrados total = $\sum y_{ij}^2 - (y_{..})^2 / n$
 Donde n es el total de los datos

Suma de cuadrados total =
 $1580 - (136)^2 / 12 = 38.6$

4 Es necesario encontrar la varianza entre los tratamientos. Primero se obtiene la suma de cada uno de los tratamientos (que se llamarán $y_{i.}$). Cada suma de tratamientos se eleva al cuadrado y se suman los cuadrados.

	Rep 1	Rep 2	Rep 3	Rep 4	suma	
Trat 1	13	12	13	11	49	$y_{1.}$
Trat 2	11	14	13	12	50	$y_{2.}$
Trat 3	9	8	11	9	37	$y_{3.}$
			suma	136	136	

	2401
	2500
	1369
Suma	6270

cuadrados $\sum y_{i.}^2$ →

5 Se calcula la suma de cuadrados de los tratamientos con la fórmula:
 Suma Cuadrados de tratamientos
 $= (\sum y_{i.}^2) / r - (y_{..})^2 / n$
 Donde r es el número de repeticiones. Nótese que el segundo término ya está calculado.

Suma de cuadrados de tratamientos
 $= 6270 / 4 - (136)^2 / 12 = 26.16$

JM

6 Se calcula los grados de libertad de los tratamientos que serán: $t - 1$
 Donde t es el número de tratamientos

Grados de libertad de tratamientos:
 $= 3 - 1 = 2$

7 Se calcula los grados de libertad del total:
 $n - 1$

Grados de libertad del total
 $= 12 - 1 = 11$

8 Los datos hasta ahora calculados se llenan en la tabla de análisis de varianza. GL son los grados de libertad, SC es la suma de cuadrados y CM son los cuadrados medios.

	GL	SC	CM	F
Trat	2	26.16		
Error				
Total	11	38.6		

9

Se calcula los grados de libertad del error:

grados de libertad del error : $t(r - 1)$

Donde t es el número de tratamientos, r el número de repeticiones.

También se puede calcula GL del error como :

GL error = GL Total – GL tratamientos

	GL	SC	CM	F
Trat	2	26.16		
Error	9			
Total	11	38.6		

10

Se calcula la suma de cuadrados del error, la fórmula es:

$$SC \text{ error} = \sum y_{ij}^2 - \sum y_i.^2 / r$$

El primer término se puede tomar de la fórmula de la SC total, el segundo término de la SC trat.

Otra forma de calcular la SC del error es:

SC error = SC total – SC tratamiento

	GL	SC	CM	F
Trat	2	26.16		
Error	9	12.44		
Total	11	38.6		

11

Se calculan los cuadrados medios de los tratamientos con la siguiente ecuación:

$$CM \text{ trat} = SC \text{ trat} / GL \text{ trat}$$

	GL	SC	CM	F
Trat	2	26.16	13.08	
Error	9	12.44		
Total	11	38.6		

12

Se calculan los cuadrados medios del error con la siguiente fórmula:

$$CM \text{ error} = SC \text{ error} / GL \text{ error}$$

	GL	SC	CM	F
Trat	2	26.16	13.08	
Error	9	12.44	1.382	
Total	11	38.6		

13

Se calcula el valor F con el siguiente ecuación

$$F = CM \text{ trat} / CM \text{ error}$$

	GL	SC	CM	F
Trat	2	26.16	13.08	9.463
Error	9	12.44	1.382	
Total	11	38.6		

14

Se busca en las tablas de la distribución F el valor al 0.05% de significancia. Los grados de libertad de los tratamientos serán los grados de libertad del numerador y los grados de libertad del error serán los grados de libertad de denominador.

$$F_{0.05, 2, 9} = 4.26$$

JM

15

Si la F calculada es mayor que la F de las tablas, se concluye que sí hay diferencia entre tratamientos, de lo contrario se concluye que no hay diferencias entre tratamientos.

Como $9.46 > 4.26$, se concluye que sí hay diferencias entre tratamientos

Hipótesis de un diseño completamente al azar

En un diseño completamente al azar, la hipótesis nula es que los efectos de tratamientos (β) son todos iguales, lo que se expresa por:

$$H_0: \beta_1 = \beta_2 = \beta_3 \dots$$

La hipótesis alterna es que hay al menos un efecto de tratamiento que es diferente a los demás. Para probar la hipótesis, en la tabla ANVA se comparan los cuadrados medios de tratamientos respecto a los cuadrados medios del error, los primeros deben ser suficiente mayores que los segundos..

Debido a que los cuadrados medios se distribuye ji-cuadrada, al dividir dos variables con distribución ji-cuadrada se obtiene una variable con una distribución F (Fisher). Es por esto que la división de los cuadrados medios de tratamientos entre cuadrados medios del error se le llama «F calculada», y se puede ubicar en esta distribución

El nivel de significancia es de 0.05 (95% de seguridad), es decir, que se tiene que identificar el punto (F tabular con nivel de significancia de 0.05), cuya área a la derecha sea de 0.05; si F_c logra ser mayor que F_t entonces es cuando se rechaza la H_0 .

En el caso de los paquetes computacionales, no se usan las columnas F_c y F_t , se cambia por la probabilidad de que ocurra F_c , cuando esto ocurre, se rechaza H_0 cuando la probabilidad es menor a 0.05.

