

# Diseño y Análisis de Experimentos

## Diseño con Estructura Factorial

Johan Steven Aparicio

Escuela de Estadística  
Facultad de Ingeniería - Universidad del Valle  
[johan.aparicio@correounivalle.edu.co](mailto:johan.aparicio@correounivalle.edu.co)

Julio, 2022

# Tabla de Contenido

- 1 Introducción
  - Diseño con Estructura Factorial
- 2 Modelo de Análisis
- 3 Factores Aleatorios

El aspecto esencial de estos diseños es que se estudia la interacción entre dos o más factores de tratamiento.

Los experimentos con estructura factorial pueden disponerse en diseños completamente aleatorizados, en bloques al azar o en otro tipo de diseños.

# Generalidades

- El objetivo del experimento factorial consiste en valorar el efecto de mas de un factor sobre la variable respuesta.

# Generalidades

- El objetivo del experimento factorial consiste en valorar el efecto de mas de un factor sobre la variable respuesta.
- Se realizan todas las combinaciones posibles entre los niveles de los factores.

# Generalidades

- El objetivo del experimento factorial consiste en valorar el efecto de mas de un factor sobre la variable respuesta.
- Se realizan todas las combinaciones posibles entre los niveles de los factores.
- En cada ensayo o réplica completa del experimento se investigan todas las combinaciones posibles de los niveles de los factores.

# Generalidades

- El objetivo del experimento factorial consiste en valorar el efecto de mas de un factor sobre la variable respuesta.
- Se realizan todas las combinaciones posibles entre los niveles de los factores.
- En cada ensayo o réplica completa del experimento se investigan todas las combinaciones posibles de los niveles de los factores.
- Se plantea la posibilidad de que los factores presenten un efecto de interacción. Es decir, la diferencia en la respuesta entre niveles de un factor no es la misma en todos los niveles de los otros factores.

# Ventajas

- Se amplía la base de la inferencia en relación a un factor ya que se estudia en las diferentes condiciones representadas por los niveles de otros factores.



# Ventajas

- Se amplia la base de la inferencia en relación a un factor ya que se estudia en las diferentes condiciones representadas por los niveles de otros factores.
- Se puede obtener una estimación de la interacción de los efectos, es decir, se determina el grado y la forma en la cual se modifica el efecto de un factor en presencia de los niveles de los otros factores.

# Desventajas

- El gran número de combinaciones de tratamientos, cuando se estudian muchos factores o muchos niveles, aumenta el costo del experimento al tener muchas unidades experimentales.

# Desventajas

- El gran número de combinaciones de tratamientos, cuando se estudian muchos factores o muchos niveles, aumenta el costo del experimento al tener muchas unidades experimentales.
- Difícil interpretación principalmente de las interacciones de orden superior (interacciones de más de tres factores).

# Efecto Interacción

- Cuando la conclusión de un factor depende de otro factor, se dice que hay interacción, la presencia de interacción indica que no se puede estudiar el efecto de un factor sobre la respuesta independiente del otro factor.

# Efecto Interacción

- Cuando la conclusión de un factor depende de otro factor, se dice que hay interacción, la presencia de interacción indica que no se puede estudiar el efecto de un factor sobre la respuesta independiente del otro factor.

**Nota:** Uno de los objetivos más importantes de un experimento factorial es detectar si existe interacción de los factores.

# Ejemplo

		B: velocidad			Total $Y_{i..}$
		0.20	0.25	0.30	
A: profundidad	0.15	74 64 <span>198</span> 60	92 86 <span>266</span> 88	99 98 <span>299</span> 102	763
	0.18	79 68 <span>220</span> 73	98 104 <span>290</span> 88	104 99 <span>298</span> 95	808
	0.21	82 88 <span>262</span> 92	99 108 <span>302</span> 95	108 110 <span>317</span> 99	881
	0.24	99 104 <span>299</span> 96	104 110 <span>313</span> 99	114 111 <span>332</span> 107	944
	Total $Y_{.j.}$	979	1 171	1 246	$Y_{...} = 3\ 396$

El modelo, en este caso con dos factores, debe considerar la existencia de un nuevo tipo de parámetro que permita hacer pruebas estadísticas sobre ellos, para definir la existencia o no de interacción entre los factores.

$$y_{ijk} = \mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \varepsilon_{ijk}$$

$$i = 1, \dots, t, \quad j = 1, \dots, b, \quad k = 1, \dots, r$$

El modelo, en este caso con dos factores, debe considerar la existencia de un nuevo tipo de parámetro que permita hacer pruebas estadísticas sobre ellos, para definir la existencia o no de interacción entre los factores.

$$y_{ijk} = \mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \varepsilon_{ijk}$$

$$i = 1, \dots, t, \quad j = 1, \dots, b, \quad k = 1, \dots, r$$

$$\text{Var}(\text{Total}) = \text{Var}(\tau) + \text{Var}(\alpha) + \text{Var}(\tau\alpha) + \text{Var}(\text{Error})$$

Cuales serán las matrices  $X$ ,  $\beta$ ,  $X'X$  y  $X'Y$ ?



- $y_{ijk}$  = Valor aleatorio que toma la variable respuesta en la k-ésima replica que ha sido expuesta al j-ésimo nivel del factor B y al i-ésimo nivel del factor A.

- $y_{ijk}$  = Valor aleatorio que toma la variable respuesta en la k-ésima replica que ha sido expuesta al j-ésimo nivel del factor B y al i-ésimo nivel del factor A.
- $\mu$  = Parámetro de centralidad o promedio global.

- $y_{ijk}$  = Valor aleatorio que toma la variable respuesta en la k-ésima replica que ha sido expuesta al j-ésimo nivel del factor B y al i-ésimo nivel del factor A.
- $\mu$  = Parámetro de centralidad o promedio global.
- $\tau_i$  = Efecto debido al i-ésimo nivel del factor A.

- $y_{ijk}$  = Valor aleatorio que toma la variable respuesta en la k-ésima replica que ha sido expuesta al j-ésimo nivel del factor B y al i-ésimo nivel del factor A.
- $\mu$  = Parámetro de centralidad o promedio global.
- $\tau_i$  = Efecto debido al i-ésimo nivel del factor A.
- $\alpha_j$  = Efecto debido al j-ésimo nivel del factor B.

- $y_{ijk}$  = Valor aleatorio que toma la variable respuesta en la k-ésima replica que ha sido expuesta al j-ésimo nivel del factor B y al i-ésimo nivel del factor A.
- $\mu$  = Parámetro de centralidad o promedio global.
- $\tau_i$  = Efecto debido al i-ésimo nivel del factor A.
- $\alpha_j$  = Efecto debido al j-ésimo nivel del factor B.
- $(\tau\alpha)_{ij}$  = Efecto debido a la interacción de primer orden entre los factores A y B.

- $y_{ijk}$  = Valor aleatorio que toma la variable respuesta en la k-ésima replica que ha sido expuesta al j-ésimo nivel del factor B y al i-ésimo nivel del factor A.
- $\mu$  = Parámetro de centralidad o promedio global.
- $\tau_i$  = Efecto debido al i-ésimo nivel del factor A.
- $\alpha_j$  = Efecto debido al j-ésimo nivel del factor B.
- $(\tau\alpha)_{ij}$  = Efecto debido a la interacción de primer orden entre los factores A y B.
- $\varepsilon_{ijk}$  = Efecto debido al error experimental.

Sea  $X$  la matriz de diseño:

$$\begin{bmatrix} X_{0(n \times 1)} & X_{1(n \times t)} & X_{2(n \times b)} & X_{3(n \times tb)} \end{bmatrix}$$

con  $n = tbr$ .

Donde ya son conocidas las submatrices  $X_0$ ,  $X_1$  y  $X_2$  y la submatriz  $X_3$  será una matriz diagonal en bloques de  $r$  unos.

El número de parámetros a estimar es  $p = 1 + t + b + tb$ , donde

$$\beta = \begin{bmatrix} \mu \\ \tau_{t \times 1} \\ \alpha_{b \times 1} \\ \tau\alpha_{tb \times 1} \end{bmatrix}$$



El número de parámetros a estimar es  $p = 1 + t + b + tb$ , donde

$$\beta = \begin{bmatrix} \mu \\ \tau_{t \times 1} \\ \alpha_{b \times 1} \\ \tau\alpha_{tb \times 1} \end{bmatrix} \quad X'Y = \begin{bmatrix} Y_{...} \\ Y_{1..} \\ \vdots \\ Y_{t..} \\ Y_{,1.} \\ \vdots \\ Y_{.b.} \\ Y_{11.} \\ \vdots \\ Y_{tb.} \end{bmatrix}$$

Las ecuaciones normales, dado por  $X'X\beta = X'Y$ , serán:

$$\begin{aligned} E[Y_{...}] &= E \left[ \sum_{i=1}^t \sum_{j=1}^b \sum_{k=1}^r (\mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \varepsilon_{ijk}) \right] \\ &= tbr\mu + br \sum_{i=1}^t \tau_i + tr \sum_{j=1}^b \alpha_j + r \sum_{i=1}^t \sum_{j=1}^b (\tau\alpha)_{ij} \end{aligned}$$

Las ecuaciones normales, dado por  $X'X\beta = X'Y$ , serán:

$$\begin{aligned} E[Y_{...}] &= E \left[ \sum_{i=1}^t \sum_{j=1}^b \sum_{k=1}^r (\mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \varepsilon_{ijk}) \right] \\ &= tbr\mu + br \sum_{i=1}^t \tau_i + tr \sum_{j=1}^b \alpha_j + r \sum_{i=1}^t \sum_{j=1}^b (\tau\alpha)_{ij} \end{aligned}$$

$$\begin{aligned} E[Y_{i..}] &= E \left[ \sum_{j=1}^b \sum_{k=1}^r (\mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \varepsilon_{ijk}) \right] \\ &= br\mu + br\tau_i + r \sum_{j=1}^b \alpha_j + r \sum_{j=1}^b (\tau\alpha)_{ij} \end{aligned}$$

$$\begin{aligned} E[Y_{.j}] &= E \left[ \sum_{i=1}^t \sum_{k=1}^r (\mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \varepsilon_{ijk}) \right] \\ &= tr\mu + r \sum_{i=1}^t \tau_i + tr\alpha_j + r \sum_{i=1}^t (\tau\alpha)_{ij} \end{aligned}$$

$$\begin{aligned} E[Y_{.j.}] &= E \left[ \sum_{i=1}^t \sum_{k=1}^r (\mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \varepsilon_{ijk}) \right] \\ &= tr\mu + r \sum_{i=1}^t \tau_i + tr\alpha_j + r \sum_{i=1}^t (\tau\alpha)_{ij} \end{aligned}$$

$$\begin{aligned} E[Y_{ij.}] &= E \left[ \sum_{k=1}^r (\mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \varepsilon_{ijk}) \right] \\ &= r\mu + r\tau_i + r\alpha_j + r(\tau\alpha)_{ij} \end{aligned}$$

Para solucionar el SEN se utilizan las siguientes restricciones:

$$\sum_{i=1}^t \tau_i = 0, \quad \sum_{j=1}^t \alpha_j = 0, \quad \sum_{i=1}^t \sum_{j=1}^b (\tau\alpha)_{ij} = 0$$

$$\hat{\mu} = \frac{Y_{...}}{tbr} = \bar{Y}_{...}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\mu} \\ \hat{\tau} \\ \hat{\alpha} \\ \hat{\tau\alpha} \end{bmatrix};$$

Para solucionar el SEN se utilizan las siguientes restricciones:

$$\sum_{i=1}^t \tau_i = 0, \quad \sum_{j=1}^t \alpha_j = 0, \quad \sum_{i=1}^t \sum_{j=1}^b (\tau\alpha)_{ij} = 0$$

$$\hat{\mu} = \frac{Y_{...}}{tbr} = \bar{Y}_{...}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\mu} \\ \hat{\tau} \\ \hat{\alpha} \\ \hat{\tau\alpha} \end{bmatrix}; \quad \hat{\tau}_i = \frac{Y_{i..}}{br} - \bar{Y}_{...} = \bar{Y}_{i..} - \bar{Y}_{...}$$

Para solucionar el SEN se utilizan las siguientes restricciones:

$$\sum_{i=1}^t \tau_i = 0, \quad \sum_{j=1}^t \alpha_j = 0, \quad \sum_{i=1}^t \sum_{j=1}^b (\tau\alpha)_{ij} = 0$$

$$\hat{\mu} = \frac{Y_{...}}{tbr} = \bar{Y}_{...}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\mu} \\ \hat{\tau} \\ \hat{\alpha} \\ \tau\hat{\alpha} \end{bmatrix}; \quad \begin{aligned} \hat{\tau}_i &= \frac{Y_{i..}}{br} - \bar{Y}_{...} = \bar{Y}_{i..} - \bar{Y}_{...} \\ \hat{\alpha}_j &= \frac{Y_{.j.}}{tr} - \bar{Y}_{...} = \bar{Y}_{.j.} - \bar{Y}_{...} \end{aligned}$$



Para solucionar el SEN se utilizan las siguientes restricciones:

$$\sum_{i=1}^t \tau_i = 0, \quad \sum_{j=1}^t \alpha_j = 0, \quad \sum_{i=1}^t \sum_{j=1}^b (\tau\alpha)_{ij} = 0$$

$$\hat{\mu} = \frac{Y_{...}}{tbr} = \bar{Y}_{...}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\mu} \\ \hat{\tau} \\ \hat{\alpha} \\ \hat{\tau\alpha} \end{bmatrix};$$

$$\hat{\tau}_i = \frac{Y_{i..}}{br} - \bar{Y}_{...} = \bar{Y}_{i..} - \bar{Y}_{...}$$

$$\hat{\alpha}_j = \frac{Y_{.j.}}{tr} - \bar{Y}_{...} = \bar{Y}_{.j.} - \bar{Y}_{...}$$

$$(\hat{\tau\alpha})_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$$

La reducción de  $\beta$  estaría dada por:

$$R(\beta) = R(\mu, \tau, \alpha, \tau\alpha) = \hat{\beta}'X'Y$$

Encontrar la anterior reducción para hallar las sumas de cuadrados correspondientes al ANOVA.

# Hipótesis

Hipótesis de Interacción:

$$H_0 : (\tau\alpha)_{ij} = 0 \forall ij \quad \text{vs} \quad H_1 : (\tau\alpha)_{ij} \neq 0$$

Hipótesis de efectos principales:

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_t = 0 \quad \text{vs} \quad H_1 : \tau_i \neq 0 \text{ para algún } i$$

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_b = 0 \quad \text{vs} \quad H_1 : \alpha_j \neq 0 \text{ para algún } j$$

## Tabla ANOVA

$F.V$	$g.l$	$SC$	$CM$	$F_0$	$E(CM)$
$\tau$	$t - 1$	$\sum_{i=1}^t \frac{Y_{i..}^2}{br} - \frac{Y_{...}^2}{tbr}$	$\frac{SC(\tau)}{t-1}$	$\frac{CM\tau}{CME}$	$br \sum_i \frac{(\tau_i - \bar{\tau})^2}{t-1} + \sigma^2$
$\alpha$	$b - 1$	$\sum_{j=1}^b \frac{Y_{.j.}^2}{tr} - \frac{Y_{...}^2}{tbr}$	$\frac{SC(\alpha)}{b-1}$	$\frac{CM\alpha}{CME}$	$tr \sum_j \frac{(\alpha_j - \bar{\alpha})^2}{b-1} + \sigma^2$
$\tau\alpha$	$(t-1)(b-1)$	$SC(\tau\alpha)$	$\frac{SC(\tau\alpha)}{(t-1)(b-1)}$	$\frac{CM\tau\alpha}{CME}$	$r \sum_i \sum_j \frac{(\tau\alpha_{ij} - \tau\alpha)^2}{(t-1)(b-1)} + \sigma^2$
$Error$	$tb(r-1)$	$SC(Error)$	$\frac{SC(Error)}{tb(r-1)}$		$\sigma^2$
$Total$	$tbr - 1$	$Y'Y - FC$			

Con:

$$SC(\tau\alpha) = \sum_{i=1}^t \sum_{j=1}^b \frac{Y_{ij.}^2}{r} - \sum_{i=1}^t \frac{Y_{i..}^2}{br} - \sum_{j=1}^b \frac{Y_{.j.}^2}{tr} + \frac{Y_{...}^2}{tbr}$$

## Diseño con tres factores

La inclusión de más factores al diseño experimental aumenta la complejidad de los patrones de interacción entre los factores de tratamiento.

El número de combinaciones de tratamientos aumenta tanto como se agregan factores al diseño, es decir, un diseño de tres factores con  $a$  niveles del Factor 1,  $b$  niveles del Factor 2 y  $c$  niveles del Factor 3, tiene  $abc$  tratamientos y  $abcr$  unidades muestrales.

El diseño con dos factores permite la investigación de la interacción de primer orden o doble (AB). En el diseño con tres factores se consideran las interacciones dobles adicionales (AC y BC) además de la interacción de segundo orden o triple (ABC).

$$y_{ijkl} = \mu + \tau_i + \alpha_j + \gamma_k + (\tau\alpha)_{ij} + (\tau\gamma)_{ik} + (\alpha\gamma)_{jk} + (\tau\alpha\gamma)_{ijk} + \varepsilon_{ijkl}$$

$$i = 1, \dots, t, \quad j = 1, \dots, b, \quad k = 1, \dots, c, \quad l = 1, \dots, r$$

# Experimentos con Factores Aleatorios

Es común que un experimentador esté interesado en un factor que tiene un gran número de posibles niveles. Cuando el experimentador selecciona aleatoriamente a de estos niveles de la población de los niveles del factor, entonces se dice que el factor es aleatorio.

Debido a que los niveles del factor utilizados en el experimento se eligieron al azar, se hacen inferencias acerca de la población completa de los niveles del factor.



El modelo estadístico lineal es:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

$$i = 1, \dots, t, \quad j = 1, \dots, r$$

donde tanto  $\tau_i$  como  $\varepsilon_{ij}$  son variables aleatorias.

Si  $\tau_i$  tiene varianza  $\sigma_\tau^2$ , la varianza de cualquier observación es:

$$V(y_{ij}) = \sigma_\tau^2 + \sigma^2$$

A las varianzas se les llama los componentes de varianza y al modelo se le llama modelo de efectos aleatorios o de los componentes de varianza.



Para probar hipótesis en este modelo se requiere que:

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\tau_i \sim N(0, \sigma_\tau^2)$$

$\tau_i$  y  $\varepsilon_{ij}$  sean independientes

La suma de cuadrados identidad sigue siendo valida:

$$SC_{Total} = SC_{Tratamientos} + SC_{Error}$$

En este contexto, no tiene sentido probar hipótesis acerca de los efectos de tratamientos individuales, por lo que en su lugar se prueban hipótesis acerca del componente de varianza  $\sigma_\tau^2$  :

$$H_0 : \sigma_\tau^2 = 0 \text{ vs } H_1 : \sigma_\tau^2 > 0$$

En este contexto, no tiene sentido probar hipótesis acerca de los efectos de tratamientos individuales, por lo que en su lugar se prueban hipótesis acerca del componente de varianza  $\sigma_\tau^2$  :

$$H_0 : \sigma_\tau^2 = 0 \text{ vs } H_1 : \sigma_\tau^2 > 0$$

Bajo  $H_0$  todos los tratamientos son idénticos, pero bajo  $H_1$  existe variabilidad entre los tratamientos.

El procedimiento de cálculo y el análisis de la tabla de varianza del modelo de efectos aleatorios son idénticos a los que se utilizaron en el caso de efectos fijos. Sin embargo, las conclusiones son muy diferentes, ya que se aplican a la población completa de los tratamientos.

Como

$$\frac{SC_{Error}}{\sigma^2} \sim \chi_{(n-t)} \text{ y } \frac{SC_{Trat}}{\sigma^2} \sim \chi_{(t-1)}$$

entonces bajo  $H_0 : \sigma_\tau^2 = 0$ , el cociente

$$F_0 = \frac{\frac{SC_{Trat}}{t-1}}{\frac{SC_{Error}}{n-t}} = \frac{CM_{Trat}}{CM_{Error}} \sim F_{(t-1, n-t)}$$

Como

$$E[CM_{Trat}] = \sigma^2 + r\sigma_\tau^2$$

$$E[CM_{Error}] = \sigma^2$$

Entonces los estimadores de los componentes de varianza son:

$$\hat{\sigma}^2 = CM_{Error}$$

$$\hat{\sigma}_\tau^2 = \frac{CM_{Trat} - CM_{Error}}{r}$$