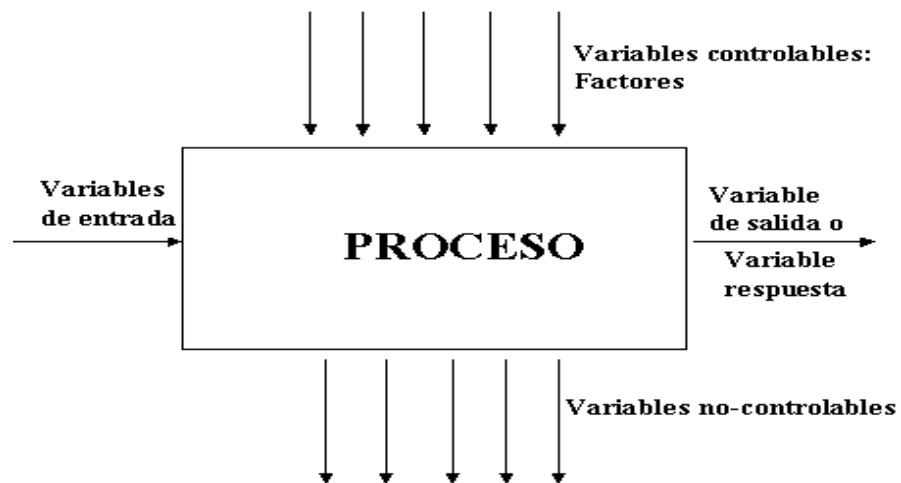


Capítulo 3

Diseño Estadístico de Experimentos

Una prueba o serie de pruebas en las cuales se introducen cambios deliberados en las variables de entrada que forman el proceso, de manera que sea posible observar e identificar las causas de los cambios en la variable de salida



REALIZAR UN EXPERIMENTO

Aplicar los distintos niveles, o combinaciones de niveles cuando hay presentes más de un factor, a distintas unidades experimentales y se observa el valor de la variable respuesta.

- Unidades experimentales: (personas, elementos físicos, ...)
- Factor: Variable controlable por el experimentador (Niveles del factor o tratamientos)
- Variable de interés: Variable Respuesta
- Error experimental o perturbación: Variables no controlables por el experimentador
- Tamaño del experimento: número total de observaciones.

OBJETIVO

Estudiar el efecto que sobre la Variable Respuesta tiene un conjunto de otras variables que reciben el nombre de Factores

ETAPAS

- 1) Diseñar un experimento con una estructura lo más adecuada posible a la situación que se desea estudiar y a los medios disponibles.
 - a) Planteamiento general del problema y de los objetivos que se persiguen.
 - b) Selección y definición de la variable respuesta.
 - c) Elección de los factores y niveles que han de intervenir en el experimento.
 - d) Determinación del conjunto de unidades experimentales incluidas en el estudio.
 - e) Determinación de los procedimientos por los cuales los tratamientos se asignan a las unidades experimentales.
- 2) Realizar la experimentación de acuerdo con el plan previamente establecido en el diseño.
- 3) Analizar estadísticamente los resultados obtenidos y comprobar si las hipótesis establecidas y el modelo de diseño elegido se adecuan a la situación estudiada.
- 4) Realizar las modificaciones oportunas para ampliar o modificar el diseño.
- 5) Obtener las conclusiones apropiadas.

PRINCIPIOS BÁSICOS DEL DISEÑO DE EXPERIMENTOS

- **Aleatorización**: La asignación de las unidades experimentales a los distintos tratamientos y el orden en el que se realizan los ensayos se determinan al azar.
- **Replicación**.
- **Homogeneidad del material experimental**.

DISEÑO COMPLETAMENTE ALEATORIZADO

- Una compañía algodonera que emplea diversos fertilizantes desea comprobar si éstos tienen efectos diferentes sobre el rendimiento de la semilla de algodón.
 - Una profesora de estadística que imparte en grupos experimentales de alumnos, en los que explica la misma materia pero siguiendo distintos métodos de enseñanza, desea comprobar si el método de enseñanza utilizado influye en las calificaciones de los alumnos.
 - Una industria química, que obtiene un determinado producto, está interesada en comprobar si los cambios de temperatura influyen en la cantidad de producto obtenido.
-
- ★ **INTERÉS**: Un solo factor con varios niveles o tratamientos
 - ★ **TÉCNICA ESTADÍSTICA**: Análisis de la Varianza de un factor o una vía
 - ★ **OBJETIVO**: Comparar entre sí varios grupos o tratamientos
 - ★ **MÉTODO**: Descomposición de la variabilidad total de un experimento en componentes independientes

OTROS FACTORES QUE INFLUYEN

- Pequeñas variaciones en la cantidad de riego, en la pureza de los insecticidas suministrados, etc.
- El nivel cultural del alumno, el grado de atención y de interés del alumno, etc.
- La pureza de la materia prima, la habilidad de los operarios, etc.

Teóricamente es posible dividir esta variabilidad en dos partes, la originada por el factor de interés y la producida por los restantes factores que entran en juego, conocidos o no, controlables o no, que recibe el nombre de perturbación o error experimental.

MODELO ESTADÍSTICO

$$y_{ij} = \mu + \tau_i + u_{ij} \quad , \quad i = 1, \dots, I; \quad j = 1, \dots, n_i$$

- y_{ij} : Variable aleatoria que representa la observación j -ésima del i -ésimo tratamiento (nivel i -ésimo del factor).
- μ : Efecto constante, común a todos los niveles. Media global.
- τ_i : Efecto del tratamiento i -ésimo. Es la parte de y_{ij} debida a la acción del nivel i -ésimo, que será común a todos los elementos sometidos a ese nivel del factor.
- u_{ij} : Variables aleatorias que engloban un conjunto de factores, cada uno de los cuales influye en la respuesta sólo en pequeña magnitud pero que de forma conjunta debe tenerse en cuenta. Deben verificar las siguientes condiciones:

★ La media sea cero: $E[u_{ij}] = 0 \quad \forall i, j$.

★ La varianza sea constante: $\text{Var}[u_{ij}] = \sigma^2 \quad \forall i, j$

- ★ Independientes entre sí: $E[u_{ij}u_{rk}] = 0 \quad i \neq r \quad \text{ó} \quad j \neq k.$
- ★ Distribución sea normal.

OBJETIVO
Estimar lo efectos de los tratamientos y contrastar las hipótesis

- 1) Todos los tratamientos producen el mismo efecto.
 $H_0 : \tau_i = 0, \quad \forall i$
- 2) Frente a la alternativa: Al menos dos difieren significativamente entre sí:
 $H_1 : \tau_i \neq 0$ por lo menos para algún i

o equivalentemente

- 1') Todos los tratamientos tienen la misma media:
 $H_0 : \mu_1 = \dots = \mu_I = \mu$
- 2') $H_1 : \mu_i \neq \mu_j$ por lo menos para algún par (i, j)

SITUACIONES (EFECTOS)

- Modelo de efectos fijos: $\sum_i n_i \tau_i = 0$
- Modelo de efectos aleatorios

SITUACIONES (TAMAÑOS MUESTRALES)

- **Modelo equilibrado o balanceado:** Todas las muestras del mismo tamaño ($n_i = n$)
- **Modelo no-equilibrado o no-balanceado:** Los tamaños, n_i , de las muestras son distintos.

TABLA ANOVA

Fuentes de Variación	Sumas de Cuadrados	Grados de libertad	Cuadrados Medios	F_{exp}
Entre grupos	SCT_r	$I - 1$	CMT_r	$\frac{CMT_r}{CMR}$
Dentro de grupos	SCR	$n - I$	CMR	
TOTAL	SCT	$n - 1$	CMT	

Aceptar H_0 si $F_{exp} \leq F_{\alpha; I-1, N-I}$; Rechazar H_0 si $F_{exp} > F_{\alpha; I-1, N-I}$

$$SCT = SCT_r + SCR$$

- 1) SCT : Suma de cuadrados total
 - 2) SCT_r : Suma de cuadrados entre tratamientos
 - 3) SCR : Suma de cuadrados dentro de los tratamientos o residual.
- 1') CMT : Cuadrado medio total: $CMT = SCT / (N - 1)$
 - 2') CMT_r : Cuadrado medio entre tratamientos: $CMT_r = SCT_r / (I - 1)$
 - 3') CMR : Cuadrado medio residual: $CMR = SCR / (N - I)$

Nota: Las expresiones de estas sumas de cuadrados están dadas en el Apéndice.

COEFICIENTE DE DETERMINACIÓN

$$R^2 = \frac{SCT_r}{SCT}$$

R^2 : Proporción de la variabilidad total presente en los datos que es explicada por el modelo de análisis de la varianza.

EJEMPLOS

1. Una compañía textil utiliza diversos telares para la producción de telas. Aunque se desea que los telares sean homogéneos con el objeto de producir tela de resistencia uniforme, se supone que puede existir una variación significativa en la resistencia de la tela debida a la utilización de distintos telares. A su disposición tiene 5 tipos de telares con los que realiza determinaciones de la resistencia de la tela. Este experimento se realiza en orden aleatorio y los resultados se muestran en la tabla siguiente

Telares	Resistencia					
1	51	49	50	49	51	50
2	56	60	56	56	57	
3	48	50	53	44	45	
4	47	48	49	44		
5	43	43	46	47	45	46

- . En este experimento, se han considerado 5 tipos de telares y se han realizado 6, 5, 5, 4 y 6 determinaciones de la resistencia de tela manufacturada con cada uno, respectivamente.
- La variable de interés o variable respuesta es la resistencia de la tela.
- El factor: Los telares
- Niveles del factor: 5
- Modelo unifactorial de efectos fijos, no-equilibrado

2. En una determinada fábrica de galletas se desea saber si las harinas de sus cuatro proveedores producen la misma viscosidad en la masa. Para ello, produce durante un día 16 masas, 4 de cada tipo de harina, y mide su viscosidad. Los resultados obtenidos son:

Proveedor A	Proveedor B	Proveedor C	Proveedor D
98	97	99	96
91	90	93	92
96	95	97	95
95	96	99	98

- Variable respuesta: viscosidad
- Factor: Proveedor
- Tratamientos: 4
- Modelo unifactorial de efectos fijos equilibrado

3. Una fábrica de textiles dispone de un gran número de telares. En principio, se supone que cada uno de ellos debe producir la misma cantidad de tela por unidad de tiempo. Para investigar esta suposición se seleccionan al azar cinco telares, y se mide la cantidad de tela producida en cinco ocasiones diferentes. Se obtienen los datos de la tabla adjunta. ¿Del estudio se concluye que todos los telares tienen el mismo rendimiento?

Telares	Producción				
1	14.0	14.1	14.2	14.0	14.1
2	13.9	13.8	13.9	14.0	14.0
3	14.1	14.2	14.1	14.0	13.9
4	13.6	13.8	14.0	13.9	13.7
5	13.8	13.6	13.9	13.8	14.0

- Variable respuesta: cantidad de tela
- Factor: Telares
- Tratamientos: 5
- Modelo unifactorial de efectos aleatorios equilibrado

DIAGNOSIS Y VALIDACIÓN DEL MODELO

Hipótesis básicas del modelo están o no en contradicción con los datos observados

HIPÓTESIS DEL MODELO

- ★ La media sea cero: $E[u_{ij}] = 0 \quad \forall i, j$.
- ★ La varianza sea constante: $\text{Var}[u_{ij}] = \sigma^2 \quad ; \quad \forall i, j$
- ★ Independientes entre sí: $E[u_{ij}u_{rk}] = 0 \quad ; \quad i \neq r \quad \text{ó} \quad j \neq k$.
- ★ Distribución sea normal.

VERIFICACIÓN

ESTIMADORES DE LAS PERTURBACIONES: RESIDUOS

$$e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \hat{\mu} - \hat{\tau}_i = y_{ij} - \bar{y}_i \quad .$$

1) Independencia de los residuos

- ♣ Gráfico de los residuos en función del tiempo

2) Normalidad de los residuos

- ♣ Histograma: Apariencia de una distribución Normal centrada en cero
- ♣ Gráfico probabilístico normal (Q-Q-Plot)

3) Homocedasticidad (Varianza constante)

- ♣ Residuos frente a los valores ajustados
- ♣ Residuos frente a ciertas variables de interés
Ambas gráficas también se utilizan para comprobar la hipótesis de independencia
- ♣ Contrastes: Barlett, Cochran, Hartley y Levene

$$H_0 : H_0 : \sigma_1^2 = \dots = \sigma_I^2 \quad \text{vs} \quad H_1 : \sigma_i^2 \neq \sigma_j^2 \quad \text{para algún par } (i, j)$$

COMPARACIONES MÚLTIPLES

Técnicas cuyo objeto es identificar:

- ★ Qué tratamientos son diferentes (estadísticamente) y
- ★ en cuánto oscila el valor de esas diferencias.

MODELO DE
EFECTOS
FIJOS

OBJETIVO FUNDAMENTAL

Comparar entre sí medias de tratamientos o grupos de ellas

PROCEDIMIENTOS ANALÍTICOS

Comparar por parejas los efectos de I tratamientos

$$H_0: \mu_i = \mu_j \quad \text{vs} \quad ; \quad H_1: \mu_i \neq \mu_j$$

- Método LSD
- Método de Bonferroni
- Método de Tukey o método HSD
- Método de rango múltiple de Duncan
- Test de Newman-Keuls
- Método Scheffé
- Método de Dunnett

DISEÑOS EN BLOQUES COMPLETOS ALEATORIZADOS

- **HOMOGENEIDAD ENTRE LAS UNIDADES EXPERIMENTALES:** En la industria algodonera: las parcelas de terreno son de la misma calidad e igual superficie.
- El error experimental reflejará esta variabilidad entre las parcelas de terreno.
- El error experimental sea lo más pequeño posible. Se debe sustraer del error experimental la variabilidad producida por las parcelas de terreno. Para ello, el experimentador puede:
 - 1) Considerar parcelas de terreno muy homogéneas.
 - 2) O bien, formar bloques de terreno de manera que el terreno de cada bloque sea lo más homogéneo posible y los bloques entre sí sean heterogéneos.

RECORDEMOS

- 1) En el diseño completamente aleatorizado asignábamos los tratamientos al azar a las parcelas sin restricción alguna.
- 2) En el diseño en bloques aleatorizados primero agrupamos las parcelas en bloques y a continuación asignamos los tratamientos a las parcelas en cada bloque.

SUPONGAMOS

- Se realiza una observación por tratamiento en cada bloque: $N = IJ$ observaciones.
- La asignación de los tratamientos a las unidades experimentales en cada bloque se determina aleatoriamente.
- Los tratamientos y los bloques son factores de efectos fijos.
- No hay interacción entre los tratamientos y los bloques: (El efecto de un factor no depende del nivel del otro factor): Efectos de los factores son aditivos.

Diseño en bloques aleatorizado

	Bloques					
Tratamientos	1	2	...	j	...	J
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1J}
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2J}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{iJ}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
I	y_{I1}	y_{I2}	...	y_{Ij}	...	y_{IJ}

MODELO ESTADÍSTICO

$$y_{ij} = \mu + \tau_i + \beta_j + u_{ij} \quad i = 1, 2, \dots, I ; j = 1, 2, \dots, J$$

- y_{ij} : La variable aleatoria que representa la observación (i)-ésima del bloque (j)-ésimo.
- μ es un efecto constante. Media global.
- τ_i : El efecto producido por el nivel i -ésimo del factor principal. $\sum_i \tau_i = 0$.
- β_j : El efecto producido por el nivel j -ésimo del factor secundario o factor de bloque. Se supone que $\sum_j \beta_j = 0$.
- u_{ij} : Variables aleatorias independientes con distribución $N(0, \sigma)$.

DOS FACTORES

- 1) Factor tratamiento \longrightarrow factor principal
- 2) Factor bloque \longrightarrow factor secundario

Interés fundamentalmente está centrado en el primero y el factor bloque se introduce en el modelo para eliminar su influencia en la variable respuesta.

OBJETIVO

- Estimar los efectos de los tratamientos y de los bloques y contrastar la hipótesis:

$$\star H_0 : \tau_i = 0 \quad \forall i \quad \text{vs} \quad H_1 : \tau_i \neq 0 \quad \text{por lo menos para algún } i$$

$$\star H_0 : \beta_j = 0 \quad \forall j \quad \text{vs} \quad H_1 : \beta_j \neq 0 \quad \text{por lo menos para algún } j$$

TABLA ANOVA

Tabla ANOVA. Modelo de Bloques Aleatorizados

F. V.	S.C. de	G. L.	C. M.	F_{exp}
Entre tratami.	$SCTr$	$I - 1$	CMT_r	CMT_r/CMR
Entre bloques	$SCBl$	$J - 1$	$CMBl$	$CMBl/CMR$
Residual	SCR	$(I - 1)(J - 1)$	CMR	
TOTAL	SCT	$IJ - 1$	CMT	

$$SCT = SCT_r + SCBl + SCR$$

- 1) SCT : Suma total de cuadrados.
 - 2) SCT_r : Suma de cuadrados entre tratamientos.
 - 3) $SCBl$: Suma de cuadrados entre bloques
 - 4) SCR : Suma de cuadrados del error o residual.
- 1') CMT : Cuadrado medio total : $CMT = SCT/(N - 1)$
 - 2') CMT_r : Cuadrado medio entre tratamientos: $CMT_r = SCT_r/(I - 1)$
 - 3') $CMBl$: Cuadrado medio entre bloques: $CMBl = SCBl/(J - 1)$
 - 4') CMR : Cuadrado medio residual: $CMR = SCR/(I - 1)(J - 1)$

Nota: Las expresiones de estas sumas de cuadrados están dadas en el Apéndice.

ANÁLISIS ESTADÍSTICO

- Contraste de interés: $H_{0\tau} \equiv \tau_1 = \dots = \tau_I = 0$

$$F_\tau = \frac{CMT_r}{CMR} \rightsquigarrow F_{(I-1), (I-1)(J-1)}$$

Rechazar H_0 a nivel α si $F_{\tau(exp)} > F_{\alpha; I-1, (I-1)(J-1)}$

- También es interesante contrastar: $H_{0\beta} \equiv \beta_1 = \dots = \beta_J = 0$

$$F_\beta = \frac{CMB_l}{CMR} \rightsquigarrow F_{(J-1), (I-1)(J-1)}$$

Rechazar H_0 a nivel α si $F_{\beta(exp)} > F_{\alpha; J-1, (I-1)(J-1)}$

EJEMPLO

Una industria desea comprobar el efecto que tienen cinco productos químicos sobre la resistencia de un tipo particular de fibra. Como también puede influir la máquina empleada en la fabricación, decide utilizar un diseño en bloques aleatorizados, considerando las distintas máquinas como bloques. La industria dispone de 4 máquinas a las que asigna los 5 productos químicos en orden aleatorio. Los resultados obtenidos se muestran en la tabla adjunta.

	Tipos de máquinas			
Producto químico	A	B	C	D
1	87	86	88	83
2	85	87	95	85
3	90	92	95	90
4	89	97	98	88
5	99	96	91	90

Variable respuesta: Resistencia de la fibra

Factor principal: Producto químico. (Niveles: 5)

Factor secundario o factor bloque: Máquinas. (Niveles: 4)

Diseño en bloques completos al azar

APÉNDICE

DISEÑO COMPLETAMENTE ALEATORIZADO

$$SCT = \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$SCT_r = \sum_{i=1}^I \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$$

$$SCR = SCT - SCT_r$$

DISEÑO EN BLOQUES COMPLETOS ALEATORIZADOS

$$SCT_r = \sum_{i=1}^I \frac{y_{i.}^2}{J} - \frac{y_{..}^2}{IJ}$$

$$SCBl = \sum_{j=1}^J \frac{y_{.j}^2}{I} - \frac{y_{..}^2}{IJ}$$

$$SCT = \sum_{i=1}^I \sum_{j=1}^J y_{ij}^2 - \frac{y_{..}^2}{IJ}$$

$$SCR = SCT - SCT_r - SCBl$$

Bibliografía utilizada:

★ **Lara Porras A.M. (2001).** “Diseño estadístico de experimentos, análisis de la varianza y temas relacionados: tratamiento informático mediante SPSS”. Ed.: Proyecto Sur.

◆ **Temporalización:** Dos horas