DISEÑO Y ANÁLISIS DE EXPERIMENTOS
Introducción

Problema: Las prestaciones a menudo dependen de más de un factor, tales como el sistema o la carga de trabajo. ¿Cuál es el más importante?, ¿Cuál es la influencia de cada factor?

Objetivo: Obtener la máxima información con el mínimo número de experimentos, de forma que nos permita aislar los efectos de cada factor y considerar únicamente los más importantes. Con estas técnicas se podrá:

- Diseñar un conjunto apropiado de experimentos para medición o simulación.
- Desarrollar el modelo que mejor se adapte a los datos obtenidos.
- Estimar la contribución de cada alternativa a las prestaciones del sistema.
- Aislar los errores de medición.
- Comprobar si las alternativas son significativamente diferentes.
- Chequear si el modelo es adecuado.

DISEÑO Y ANÁLISIS DE EXPERIMENTOS
Introducción

Terminología:

Se explicarán los términos empleados en el diseño de experimentos con un ejemplo.

Ejemplo: Diseño de un computador personal con varias opciones:

Factores de Sistema			Factores de Carga	
Procesador:	Memoria:	Discos:	Carga:	Nivel de formación:
Pentium II 300MHz	32 MB	1	Ofimática	Bachiller
AMD K6-2 300MHz	64 MB	2	Administrativa	Universitario
Pentium II 350MHz	128 MB	3	Científica	Postgraduado
		4		

Se denomina:

Variable respuesta: Al resultado del experimento. Normalmente es una medida de prestaciones del sistema (Productividad, Tiempo de respuesta...).

Factores: Cada variable que afecte a la variable respuesta y tenga varias alternativas.

Niveles: Son los valores que puede tomar un factor, es decir, cada nivel constituye una alternativa para un factor.

Factores Primarios: Son aquellos factores cuyos efectos necesitan cuantificarse. Por ejemplo el tipo de procesador y el tamaño de la memoria.

Factores Secundarios: Son factores que afectan a las prestaciones, pero cuyo efecto no estamos interesados en cuantificar. Por ejemplo, el nivel de formación o la carga.

Replicación: La repetición de todos o algunos experimentos se denomina replicación. Por ejemplo, si todos los experimentos se repiten tres veces, tres replicaciones.

Diseño: Consiste en indicar el número de experimentos, las combinaciones de factores y el número de replicaciones. En el ejemplo, realizando experimentos para todas las combinaciones de factores: 3×3×4×3×3=324 experimentos. Con 5 replicaciones: 324×5=1620 experimentos.

Interacción: Dos factores A y B se dice que interactúan si el efecto de uno depende del nivel del otro.

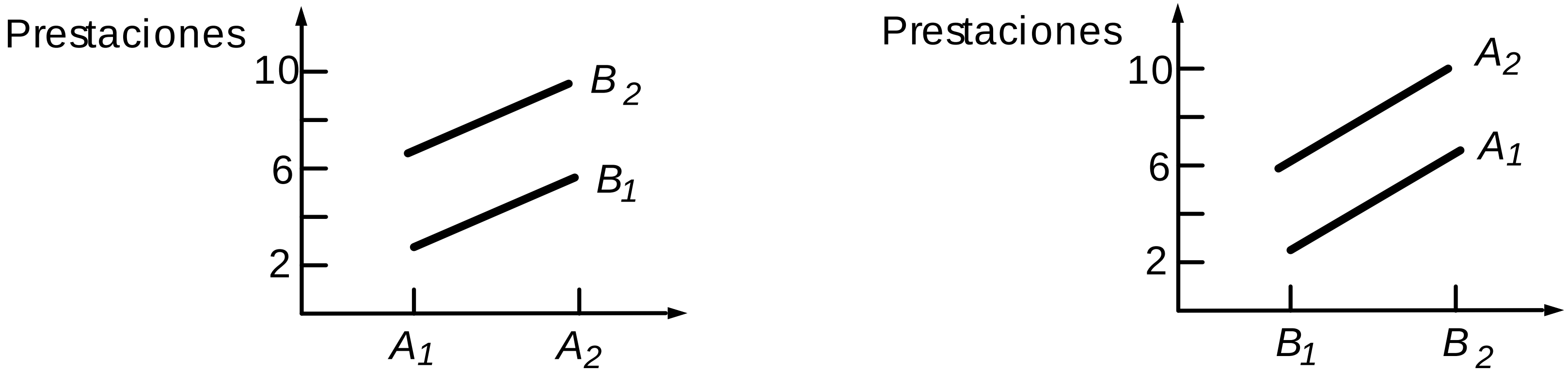
Factores no interactivos

	A_1	A_2
B_1	3	5
B_2	6	8

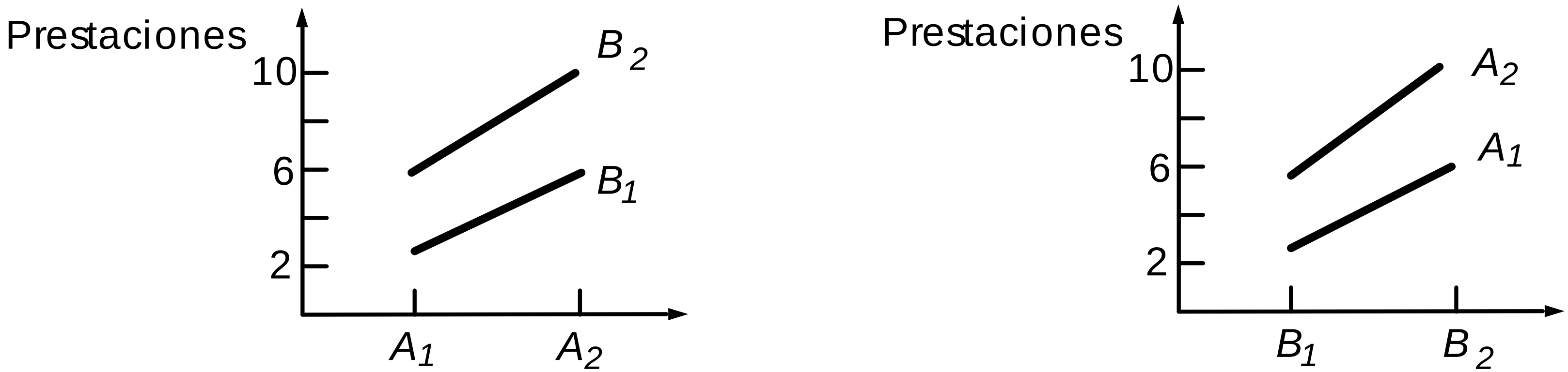
Factores interactivos

	A_1	A_2
B_1	3	5
B_2	6	9

DISEÑO Y ANÁLISIS DE EXPERIMENTOS
Introducción



(a) No interacción



(b) Interacción

Errores frecuentes:

Para el correcto desarrollo de un análisis experimental, ha de prestarse atención a los siguientes puntos:

- *Variación debida al error experimental*. Todos los valores medidos son valores aleatorios. Cada vez que se repite el experimento los valores medidos serán ligeramente diferentes, incluso si todos los factores controlables se mantienen en el mismo valor.
- *Controlar los parámetros importantes*. Elegir aquellos que afecten más al comportamiento del sistema.
- *Aislar los efectos de los diferentes factores*. Tratar de cuantificar la influencia de cada factor.
- *Ignorar las interacciones*. En ocasiones los efectos debidos a la interacción entre factores son importantes.
- *Realizar demasiados experimentos*. Establecer una estrategia de pruebas adecuada para obtener la influencia de los distintos factores realizando el menor número de experimentos.

Tipos de "Diseños Experimentales":

1) Diseño Simple

Se comienza con una configuración típica y se va variando un factor de cada vez para obtener los efectos en las prestaciones producidos por ese factor.

Ejemplo: Podríamos tomar una configuración típica de sistema, carga y usuario. Analizaremos en primer lugar las variaciones de respuesta variando únicamente la CPU, elegiríamos la de mejores prestaciones y pasaríamos a la selección del segundo factor y así sucesivamente.

Dados *k* factores, con el *i*-esimo factor con n_i niveles, y un diseño simple se necesitan *n* experimentos, donde:

$$n = 1 + \sum_{i=1}^{k} (n_i - 1)$$

Ventajas:
Número reducido de experimentos

Inconvenientes:
No es estadísticamente eficiente.
No considera los efectos de las interacciones.

2) Diseño Factorial Completo

Un diseño factorial completo utiliza todas las posibles combinaciones en todos los niveles de todos los factores. Un estudio de prestaciones con *k* factores, con el *i*-esimo factor con n_i niveles, requiere *n* experimentos, donde:

$$n = \prod_{i=1}^{k} n_i$$

Ventajas:
Examina todas las posibles combinaciones de configuración y carga.

Inconvenientes:
Coste elevado por el gran número de experimentos a realizar.

Alternativas para reducir el coste del estudio:

- Reducir el número de niveles para cada factor*
- Reducir el número de factores
- Usar el diseño factorial fraccional

3) Diseño Factorial Fraccional

Se usa este tipo de diseño cuando el número de experimentos requerido por un diseño factorial completo es demasiado grande.

Ejemplo: Considerando cuatro de los cinco factores de estudio del ejemplo, (ignoramos el número de dispositivos de disco), el número de experimentos necesarios será:

$$n = (3\,\mathrm{CPUs})(3\,\mathrm{Memoria})(3\,\mathrm{Carga})(3\,\mathrm{Formacion}) = 81$$

Se podría realizar un análisis factorial fraccional: 3^{4-2} con sólo 9 experimentos de la forma:

Número de experimento	CPU	Nivel de memoria	Tipo de carga	Nivel de formación
1	PII 300	32M	Administrativa	Bachiller
2	PII 300	64M	Científica	Postgraduado
3	PII 300	128M	Ofimática	Universitario
4	K6 300	32M	Científica	Universitario
5	K6 300	64M	Ofimática	Bachiller
6	K6 300	128M	Administrativa	Postgraduado
7	PII 35 0	32M	Ofimática	Postgraduado
8	PII 350	64M	Administrativa	Universitario
9	PII 350	128M	Científica	Bachiller

Inconveniente: Se obtiene menos información, faltarían interacciones entre factores.

Diseño Factorial 2^k

Se utiliza un diseño experimental de tipo 2^k para determinar los efectos de k factores, cada uno de los cuales tiene **dos niveles**. Este tipo de diseño factorial es fácil de analizar y ayuda a ordenar los factores por la importancia de su impacto.

Ejemplo con k = 2: Considérese el problema de estudiar el impacto del tamaño de memoria y tamaño de cache en las prestaciones de una estación a diseñar. Se tomarán dos niveles de cada factor para la simulación inicial. Las prestaciones de la estación en millones de instrucciones por segundo (MIPS) son:

	Tamaño de memoria	
Tamaño Cache (kbytes)	4 Mbytes	16 Mbytes
1	15	45
2	25	75

Se definen dos variables x_A y x_B:

$$x_A = \begin{cases} -1 & \text{Si 4 Mbytes de memoria} \\ 1 & \text{Si 16 Mbytes de memoria} \end{cases}$$

$$x_B = \begin{cases} -1 & \text{Si 1 Kbyte de cache} \\ 1 & \text{Si 2 Kbytes de cache} \end{cases}$$

Las prestaciones, y, pueden expresarse en función de x_A y x_B con un modelo de regresión no lineal

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B$$

Se resolvería el sistema de ecuaciones y se obtienen los efectos de cada factor e interacción.

Método de la tabla de signos:

En lugar de resolver el sistema de ecuaciones propuesto por el modelo de regresión no lineal visto, se utiliza el método denominado de la tabla de signos.

El tamaño de la tabla es de 2^k, donde k es el número de factores; 2^2 en el ejemplo.

Combinaciones de los factores

Producto de valores de factores

I	A	B	AB	y
1	-1	-1	1	15
1	1	-1	-1	45
1	-1	1	-1	25
1	1	1	1	75
160	80	40	20	Total
40	20	10	5	Total/4

Columna de unos

Valores de la variable respuesta en cada experimento

Suma de columna (Factor×V.Resp.)

Promedio por el número de experimentos

q_0, Valor promedio

q_A, Efecto del factor A

q_B, Efecto del factor B

q_{Ab}, Efecto de la interacción AB

Asignación de la Variación:

Sirve para determinar el porcentaje de influencia de cada factor, se basa en:

$$\text{Varianza muestral de } y = s_y^2 = \frac{\sum_{i=1}^{2^2}(y_i - y)^2}{2^2 - 1}$$

Al numerador se llama variación total de *y* o Suma total de cuadrados (SST):

$$\text{Variacion total de } y = SST = \sum_{i=1}^{2^2}(y_i - y)^2$$

Esta expresión puede dividirse en:

$$SST = 2^2 q_A^2 + 2^2 q_B^2 + 2^2 q_{AB}^2 \Longrightarrow SST = SSA + SSB + SSAB$$

En forma de fracción:

$$\text{Fracción de la variación explicada por } A = \frac{SSA}{SST}$$

En el ejemplo:

$$\bar{y} = \tfrac{1}{4}(15 + 45 + 25 + 75) = 40$$

$$\text{Variación Total } = \sum_{i=1}^{4}(y_i - \bar{y})^2 = (25^2 + 5^2 + 15^2 + 35^2)$$

$$= 2100 = 4 \times 20^2 + 4 \times 10^2 + 4 \times 5^2$$

La variación total es de 2100, de la cual 1600 (76%) puede atribuirse a la memoria, 400 (19%) puede atribuirse a la cache y sólo 100 (5%) puede atribuirse a la interacción.

Diseño Factorial 2^k General:

Ejemplo: En el diseño de una máquina LISP, los tres factores que necesitan estudiarse son: tamaño de caché, tamaño de memoria y si se usarán uno o dos procesadores. Los tres factores y su asignación de niveles son:

Factor	Nivel - 1	Nivel 1
Tamaño de memoria, *A*	4 Mbytes	16 Mbytes
Tamaño de caché, *B*	1 kbyte	2 kbytes
Número de procesadores, *C*	1	2

Los valores obtenidos del diseño experimental son:

	4 Mbytes		16 Mbytes	
T. caché (kb)	1 Procesador	2 Procesadores	1 Procesador	2 Procesadores
1	14	46	22	58
2	10	50	34	86

El análisis utilizando la tabla de signos:

I	A	B	C	AB	AC	BC	ABC	y
1	-1	-1	-1	1	1	1	-1	14
1	1	-1	-1	-1	-1	1	1	22
1	-1	1	-1	-1	1	-1	1	10
1	1	1	-1	1	-1	-1	-1	34
1	-1	-1	1	1	-1	-1	1	46
1	1	-1	1	-1	1	-1	-1	58
1	-1	1	1	-1	-1	1	-1	50
1	1	1	1	1	1	1	1	86
320	80	40	160	40	16	24	9	Total
40	10	5	20	5	2	3	1	Total/8

Y el efecto de cada factor:

$$SST = 2^3\left(q_A^2 + q_B^2 + q_C^2 + q_{AB}^2 + q_{AC}^2 + q_{BC}^2 + q_{ABC}^2\right)$$
$$= 8\left(10^2 + 5^2 + 20^2 + 5^2 + 2^2 + 3^2 + 1^2\right)$$
$$= 800 + 200 + 3200 + 200 + 32 + 72 + 8 = 4512$$
$$(18\%)(4\%)(71\%)(4\%)(1\%)(2\%)(0\%)$$

Diseño Factorial con Replicaciones $2^k r$

Un problema con el diseño factorial 2^k es que no es posible estimar los errores experimentales, puesto que no se repite el experimento. Los errores experimentales pueden cuantificarse repitiendo las medidas con las mismas combinaciones de valores de factores. Si cada uno de los 2^k experimentos se repite *r* veces, tendremos $2^k r$ *v*alores.

El modelo de regresión utilizado es: $y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e$

Supongamos ahora que en el ejemplo del estudio memoria-cache, los experimentos se repitieron tres veces, el análisis quedará:

I	A	B	AB	y	Media $\bar{y}$
1	-1	-1	1	(15, 18, 12)	15
1	1	-1	-1	(45, 48, 51)	48
1	-1	1	-1	(25, 28, 19)	24
1	1	1	1	(75, 75, 81)	77
164	86	38	20		Total
41	21.5	9.5	5		Total/4

Estimación de los Errores Experimentales:

Por definición, la suma de errores en los experimentos de este tipo es nula; para estimar la variación del error se utiliza la suma de cuadrados de los errores (SSE). $\Longrightarrow$ $SSE = \sum_{i=1}^{2^2} \sum_{j=1}^{r} e_{ij}^2$

El error se define como la diferencia entre la respuesta real y la estimada:

$$e_{ij} = y_{ij} - \hat{y}_i = y_{ij} - q_0 - q_A x_{Ai} - q_B x_{Bi} - q_{AB} x_{Ai} x_{Bi}$$

Para el ejemplo anterior:

	Efectos				Respuesta	Respuesta					
	I	A	B	AB	estimada	medida			Errores		
i	41	21.5	9.5	5	$\hat{y}_i$	y_{i1}	y_{i2}	y_{i3}	e_{i1}	e_{i2}	e_{i3}
1	1	-1	-1	1	15	15	18	12	0	3	-3
2	1	1	-1	-1	48	45	48	51	-3	0	3
3	1	-1	1	-1	24	25	28	19	1	4	-5
4	1	1	1	1	77	75	75	81	-2	-2	4

$$SSE = 0^2 + 3^2 + (-3)^2 + (-3)^2 + 0^2 + 3^2 + 1^2 + 4^2 + (-5)^2 + (-2)^2 + (-2)^2 + 4^2 = 102$$

Asignación de la Variación:

En este caso será necesario cuantificar el porcentaje de variación explicado por los errores experimentales. La variación total se reparte ahora:

$$SST = SSA + SSB + SSAB + SSE$$

O también:

$$SST = SSY - SS0 = SSA + SSB + SSAB + SSE$$

Donde SS0 representa la suma de cuadrados de la media, es decir, $2^k r\ q_0^2$ y SSY la suma de todos los valores de la variable respuesta elevada al cuadrado.

En ejemplo de la memoria-cache:

$$SSY = 15^2 + 18^2 + 12^2 + 45^2 + \cdots + 75^2 + 75^2 + 81^2 = 27204$$

$$SS0 = 2^2 r\, q_0^2 = 12 \times 41^2 = 20172$$

$$SSA = 2^2 r\, q_A^2 = 12 \times (21.5)^2 = 5547$$

$$SSB = 2^2 r\, q_B^2 = 12 \times (9.5)^2 = 1083$$

$$SSAB = 2^2 r\, q_{AB}^2 = 12 \times 5^2 = 300$$

$$SSE = 27204 - 2^2 \times 3\left(41^2 + 21.5^2 + 9.5^2 + 5^2\right) = 102$$

$$SST = SSY - SS0 = 27204 - 20172 = 7032$$

La variación total de 7032 puede dividirse en cuatro partes: el factor *A* explica 5547/7032 ó 78.88% de la variación. De la misma forma, el factor *B* explica el 15.40% y la interacción entre factores un 4.27%. El resto de variación, 1.45%, sin explicar, puede atribuirse a los errores.

Diseño Factorial Fraccional 2^{k-p}

Si el número de factores es grande, el análisis factorial completo requiere un gran número de experimentos, que puede resultar demasiado caro. En estos casos se puede usar un diseño factorial fraccional que requiere un número considerablemente menor de experimentos. Por ejemplo, un diseño 2^{k-1} requería sólo la mitad de experimentos, uno del tipo 2^{k-2} la cuarta parte, …

Este caso también se puede analizar utilizando el método de la tabla de signos, pero con algunas variaciones para que la tabla de signos continúe cumpliendo las propiedades de ortogonalidad, los pasos a dar son:

- Elección de k-p factores y preparación de una tabla de signos completa para un análisis factorial completo con k-p factores. Resultará una tabla con 2^{k-p} filas y 2^{k-p} columnas. La primera columna se marcará como *I* y consistirá en todo unos. Las siguientes k-p columnas serán los k-p factores elegidos. Las restantes columnas son productos simples de estos factores.
- De las 2^{k-p} - $(k\text{-}p) - 1$ columnas de la derecha, se eligen *p* columnas y se marcan con los *p* factores no elegidos en el paso 1.

Ejemplos:

Para análisis 2^{7-4} ⟹ *A* *B* *C* *D* *E* *F* *G*

Para análisis 2^{5-2} ⟹ *A* *B* *C* *AB* *AC* *D* *E*

Para análisis 2^{4-1} ⟹ *A* *B* *C* *AB* *AC* *BC* *D*

Nº Experimento	*A*	*B*	*C*	*AB*	*AC*	*BC*	*ABC*
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Tabla de signos básica para un análisis 2^3

Indeterminaciones:

El problema del análisis factorial fraccional es que algunos efectos no pueden calcularse. Este problema se conoce como **indeterminación** o **confusión**, y los efectos cuya influencia no puede separarse se dice que están **indeterminados** o **confundidos**.

En el caso del análisis 2^{4-1} el efecto debido al factor D está confundido con el efecto debido a la interacción ABC.

I	**A**	**B**	**C**	**AB**	**AC**	**BC**	**D**
ABCD	**BCD**	**ACD**	**ABD**	**CD**	**BD**	**AD**	**ABC**

En este mismo análisis, se podría denominar D a otra columna, en ese caso el efecto de D se confunde con otra interacción de dos factores (AB, AC ó BC). Como es más probable que los efectos de la interacción de los términos con muchos factores sean pequeños, es por eso que se prefiere la tabla de signos en la que se sustituye el nuevo factor por la fila de más a la derecha.

Este tipo de análisis es más adecuado cuando sabemos que las interacciones entre factores son pequeñas.

Experimentos con un Factor

Este tipo de experimentos se usan para comparar varias alternativas de una única variable categórica. Por ejemplo, varios procesadores, varios sistemas computadores, o varios esquemas de cache.

No existe un límite al número de niveles que puede tomar el factor.

Modelo:

El modelo matemático a partir del cual se realiza el análisis es:

r réplicas. i=1, …, r

a alternativas j=1, …, a

Valor promedio

Respuesta i al nivel j $\rightarrow$ $y_{ij} = \mu + \alpha_j + e_{ij}$ $\leftarrow$ Término de error

Efecto de la alternativa j $\rightarrow$ $\sum \alpha_j = 0$

Cálculo de los Efectos:

Tendremos $r \times a$ experimentos que se organizan en una matriz de dimensión $r \times a$. Donde:

$y_{..}$ Promedio global

$y_{.j}$ Promedio para la columna j

$$\mu = y_{..} \qquad \alpha_j = y_{.j} - \mu = y_{.j} - y_{..}$$

El análisis se puede realizar de forma tabular, un ejemplo de cálculo de efectos:

Ejemplo: Se realiza un estudio para determinar el número de bytes requerido para codificar una carga sobre tres procesadores diferentes, R, V y Z. El tamaño del código realizado se midió 5 veces para cada procesador (la carga fue codificada cada vez por un programador diferente). Los datos medidos fueron:

R	V	Z
144	101	130
120	144	180
176	211	141
288	288	374
144	72	302

Cada columna representa la carga para un procesador. Los valores de cada experimento para los procesadores (filas) se consideran no relacionados, en caso contrario este análisis no sería correcto.

Niveles 3 (R, V, Z)
Réplicas 5

		R	V	Z	Total
		144	101	130	
		120	144	180	
		176	211	141	
		288	288	374	
		144	72	302	
$\sum y_{.j}$	Suma de Columnas	872	816	1127	2815
$y_{.j}$	Media de Columnas	174.4	163.2	225.4	μ = 187.7
$\alpha_j = y_{.j} - \mu$	Efecto de la Columna	-13.3	-24.5	37.7	

Comportamiento respecto al promedio

Estimación de los Errores Experimentales:

El error en cada caso se calcula como diferencia entre la respuesta indicada por el modelo y la respuesta obtenida de la medida real.

Por construcción del modelo, la suma de errores es nula. Se obtiene por tanto la suma de los errores al cuadrado (SSE).

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{a} e_{ij}^2$$

En el Ejemplo, los errores serían

$$\begin{bmatrix} 144 & 101 & 130 \\ 120 & 144 & 180 \\ 176 & 211 & 141 \\ 288 & 288 & 374 \\ 144 & 72 & 302 \end{bmatrix} = \begin{bmatrix} 187.7 & 187.7 & 187.7 \\ 187.7 & 187.7 & 187.7 \\ 187.7 & 187.7 & 187.7 \\ 187.7 & 187.7 & 187.7 \\ 187.7 & 187.7 & 187.7 \end{bmatrix} + \begin{bmatrix} -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \end{bmatrix} + \begin{bmatrix} -30.4 & -62.2 & -95.4 \\ -54.4 & -19.2 & -45.4 \\ 1.6 & 47.8 & -84.4 \\ 113.6 & 124.8 & 148.6 \\ -30.4 & -91.2 & 76.6 \end{bmatrix}$$

$$y_{ij} = \mu + \alpha_j + e_{ij}$$

La suma de errores al cuadrado:

$$SSE = (-30.4)^2 + (-54.4)^2 + \cdots + (76.6)^2 = 94365.20$$

Asignación de la Variación:

En este caso la variación total se debe tanto al efecto del factor a estudiar como a los errores experimentales. La variación total:

$$SST = \sum_{i,j}\left(y_{i,j} - y_{..}\right)^2 = \sum_{i,j} y_{ij}^2 - ary_{..}^2 = SSY - SS0 = SSA + SSE$$

Donde:

$$SSA = \sum_{i=1}^{r}\sum_{j=1}^{a}\alpha_j^2 = r\sum_{j=1}^{a}\alpha_j^2 \qquad SS0 = \sum_{i=1}^{r}\sum_{j=1}^{a}\mu^2 = ar\mu^2 \qquad SSY = \text{Suma de cuadrados de y}$$

La variación total, se divide en SSA y SSE que representan la parte explicada y la inexplicada por la variación del factor. Interesa un porcentaje de explicación alto.

$$SSY = 144^2 + 120^2 + \cdots + 302^2 = 633639$$

$$SS0 = ar\mu^2 = 3\times 5\times (187.7)^2 = 528281.7$$

$$SSA = r\sum_j \alpha_j^2 = 5\left[(-13.3)^2 + (-24.5)^2 + (37.6)^2\right] = 10992.1$$

$$SST = SSY - SS0 = 633639.0 - 528281.7 = 105357.3$$

$$SSE = SST - SSA = 105357.3 - 10992.1 = 94365.2$$

<u>En el ejemplo</u>

Porcentaje de variación explicado SSA/SST = 10.4%

SSE/SST=89.6% Porcentaje debido a errores

Análisis de Varianza (ANOVA):

Procedimiento para comprobar si es estadísticamente significativa la contribución de un factor a la varianza de la respuesta en relación con los errores

Cada suma de cuadrados tiene asociados unos grados de libertad

SSY	**=**	**SS0**	**+**	**SSA**	**+**	**SSE**
a·r		**1**	**+**	**(a-1)**	**+**	**a(r-1)**

Se supone que

- Los errores siguen una distribución normal
- SSE y SSA siguen distribuciones chi-cuadrado

La relación $\dfrac{SSA/\nu_A}{SSE/\nu_B}$ Sigue una distribución $F_{\nu_A;\nu_B}$

$$MSA = \frac{SSA}{\nu_A}$$ Promedio de cuadrados del factor A

$$MSE = \frac{SSE}{\nu_E}$$ Promedio de cuadrados de los errores

Para un nivel de significación α:

SI $\dfrac{MSA}{MSE} > F_{[1-\alpha;\nu_A,\nu_E]}$

ENTONCES SSA se considera significativo en relación a SSE

DISEÑO Y ANÁLISIS DE EXPERIMENTOS
Experimentos con un Factor

Tabla ANOVA para un experimento de un solo factor

(resume todos los cálculos de un análisis de varianza)

Elemento	Suma de cuadrados	% variación	Grados de libertad	Promedio de cuadrados	F muestral	F teórica
y	$SSY = \sum y_{ij}^2$		ar			
$\bar{y}_{..}$	$SS0 = ar\mu^2$		1			
$y - \bar{y}_{..}$	$SST = SSY - SS0$	100	$ar - 1$			
A	$SSA = r\sum \alpha_j^2$	$100\left(\frac{SSA}{SST}\right)$	$a - 1$	$MSA = \frac{SSA}{a-1}$	$\frac{MSA}{MSE}$	$F_{[1-\alpha;a-1,a(r-1)]}$
E	$SSE = SST - SSA$	$100\left(\frac{SSE}{SST}\right)$	$a(r-1)$	$MSE = \frac{SSE}{a(r-1)}$		

(valores para el ejemplo)

Elemento	Suma de cuadrados	% variación	Grados de libertad	Promedio de cuadrados	F muestral	F teórica
y	633639.00					
$\bar{y}_{..}$	528281.69					
$y - \bar{y}_{..}$	105357.31	100.0	14			
A	10992.13	10.4	2	5496.1	0.7	2.8
E	94365.20	89.6	12	7863.8		

Conclusión: Como Fmuestral=0.7<Fteórica=2.8, la influencia del factor no es significativa