

## DISEÑO DE EXPERIMENTOS CON SPSS

### 1. ANÁLISIS EXPLORATORIO DE DATOS

Definición: Conjunto de herramientas estadísticas que permiten una visualización previa al análisis definitivo de los datos en estudio. Este análisis tiene los siguientes objetivos:

- Evaluar la calidad y la consistencia de los datos
- Determinar la distribución de las variables en estudio
- Aplicar el tratamiento de datos ausentes
- Detectar datos atípicos (outliers).
- Comprobar supuestos: normalidad, linealidad, homocedasticidad.
- Resumir los datos mediante el uso conjunto de medidas de resumen y sus gráficos

El análisis exploratorio de datos puede ser:

- Univariado: Dap, altura, peso específico, temperatura, etc
- Bivariado: Peso específico y Dureza, densidad y tenacidad
- Multivariado: Se quiere medir la relación entre la densidad, dureza, tenacidad y cizallamiento

Se puede realizar:

- En forma de gráfico
- Pruebas de contraste

### ANÁLISIS EXPLORATORIOS UNIVARIADO

#### HISTOGRAMA:

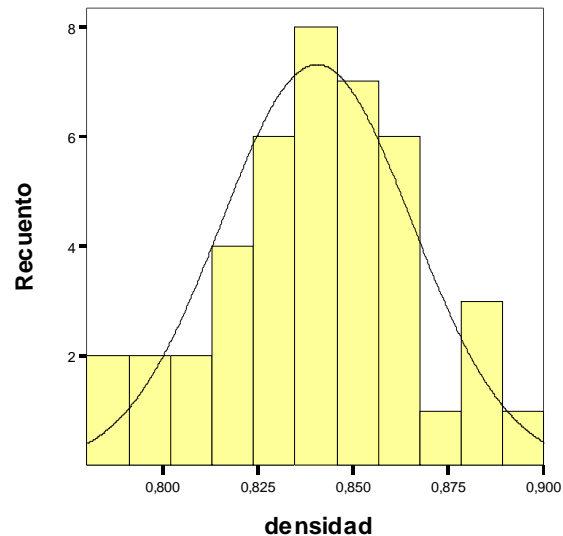
- Muestra la forma de distribución de los datos.
- Revela la presencia o no de simetría.
- Proporciona información respecto a la variabilidad de los datos.

#### Ejemplo:

##### Valores de densidad para 42 probetas de Mureillo

0,79	0,84	0,86	0,84	0,84	0,86	0,82
0,78	0,84	0,86	0,83	0,83	0,86	0,87
0,8	0,85	0,85	0,84	0,83	0,84	0,85
0,82	0,85	0,84	0,81	0,83	0,85	0,88
0,82	0,85	0,85	0,81	0,84	0,88	0,88
0,83	0,86	0,83	0,8	0,86	0,82	0,9

Histograma para valores de Densidad del Mureillo



En el histograma se observa que la distribución de los datos tienden a seguir un comportamiento normal, es decir, la mayor cantidad de valores se concentran en el centro y la menor cantidad en la cola izquierda y derecha.

**Pruebas de normalidad**

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
<b>DENSIDAD</b>	,115	42	,184	,979	42	,704

Análisis realizado con SPSS.10

Para chequear la normalidad de los datos suelen aplicarse las pruebas de contraste Kolmogorov-Smirnov y la de Shapiro-Wilk tal como se muestra en la tabla. En tal caso el contraste se define de la siguiente manera:

$H_0$  : La distribucin de los datos es normal

$H_a$  : La distribucin de los datos no es normal

Para ambas pruebas el valor  $p = 0,184$  y  $0,704$  respectivamente

De tal manera, que para un nivel de confianza del 95 % se tiene que:

$$\ddot{e} = 0,05 < \text{Valor } p = 0,184 \text{ y } \ddot{e} = 0,05 < \text{Valor } p = 0,704$$

Por lo tanto, se acepta  $H_0$ . Lo que significa que la distribución de los datos es normal

## DISEÑO DE EXPERIMENTOS CON SPSS

### 2. PRINCIPIOS BÁSICOS DEL DISEÑO DE EXPERIMENTOS

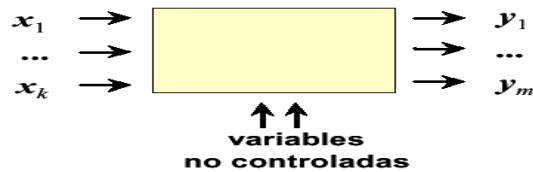
#### ¿Qué es el Diseño Estadístico de Experimentos?

Se podría definir el *Diseño Estadístico de Experimentos* (DEE), también denominado *diseño experimental*, como una metodología basada en útiles matemáticos y estadísticos cuyo objetivo es ayudar al experimentador a:

1. Seleccionar la estrategia experimental óptima que permita obtener la información buscada con el mínimo coste.
2. Evaluar los resultados experimentales obtenidos, garantizando la máxima fiabilidad en las conclusiones que se obtengan.

#### ¿En qué situaciones se aplica el Diseño Estadístico de Experimentos?

Las situaciones en las que se puede aplicar el DEE son muy numerosas. De forma general, se aplica a sistemas como el mostrado en la Figura , en los cuales se observan una o más variables experimentales *dependientes* o *respuestas* ( $y$ ) cuyo valor depende de los valores de una o más variables *independientes* ( $x$ ) controlables llamadas *factores*. Las respuestas además pueden estar influidas por otras variable que no son controladas por el experimentador. La relación entre  $x$  e  $y$  no tiene porqué ser conocida.



Ejemplos de sistemas experimentales son:

- Una reacción química, cuyo rendimiento ( $y$ ) puede ser función, entre otros, del tiempo de reacción ( $x_1$ ), la temperatura de la reacción ( $x_2$ ) y el tipo de catalizador ( $x_3$ ) utilizado. Otras variables que pueden influir son, por ejemplo, la pureza de los reactivos, la limpieza del material, la velocidad de agitación
- La resistencia que presenta un tablero de partículas ( $y$ ) al ser sometido a un esfuerzo está en función de su densidad ( $x_1$ ), el prensado ( $x_2$ ) y la temperatura ( $x_3$ ) entre otros.
- Un alimento, producido por mezcla en distintas proporciones ( $x$ ) de sus ingredientes, lo cual da lugar a diferentes olores y sabores ( $y$ )

Estos son sólo algunos ejemplos del amplio campo de aplicación del DEE, que abarca, en la industria, desde el laboratorio hasta proceso de producción. Al facilitar el desarrollo más rápido de productos, y a más bajo coste, el DEE juega un papel fundamental desde el punto de vista industrial y proporciona una ventaja competitiva importante para la empresa que lo usa.

## DISEÑO DE EXPERIMENTOS CON SPSS

### ¿Qué objetivos se suelen perseguir al estudiar sistemas como los descritos?

- Obtener un conocimiento inicial sobre un nuevo sistema en estudio. ¿En qué valores de los factores se puede centrar la investigación?
- Determinar la influencia de los factores sobre las respuestas observadas.
- De entre todos los factores que afectan al proceso, ¿cuales influyen más?, ¿cómo interactúan entre ellos?
- Optimizar respuestas. ¿Qué valores de los factores proporcionan las respuestas de mayor calidad?

Determinar la robustez del sistema. ¿Como afectan a la respuesta variaciones no controladas en el valor de los factores?

### PROCEDIMIENTO DE LA APLICACIÓN DEL DEE

La aplicación del diseño de experimentos requiere considerar las siguientes etapas:

- Comprender el problema y definir claramente el objetivo.
- Identificar los factores que potencialmente podrían influir en la función objetivo, y los valores que éstos pueden tomar. Entre estos valores se buscará información necesaria.
- Establecer una estrategia experimental, llamada *plan de experimentación*.
- Efectuar los experimentos con los valores de los factores decididos en el punto 3 para obtener los valores de las respuestas estudiadas.
- Responder las preguntas planteadas, sea directamente, sea utilizando un modelo matemático. Si es necesario, volver a la etapa 1.

Un **experimento** es el conjunto de pruebas sobre un proceso o sistema, en las que se realiza controladamente un cambio en el valor de las variables de entrada o **factores** para identificar las causas de los cambios observados en la **variable respuesta**

### GUIA PARA EL DISEÑO DE UN EXPERIMENTO

#### 1. Reconocimiento y formulación del problema

- ✓ Tener una idea clara de lo que se va a estudiar. Se recomienda trabajar en equipo
- ✓ Una aproximación secuencial utilizando pequeños experimentos es mejor estrategia

#### 2. Selección de los factores y sus niveles o rango de variación

- ✓ Factores de diseño, factores a nivel constante y factores que se permiten variar
- ✓ Factores ruido, controlables (definición de bloques) y no controlables (análisis de covarianza)

#### 3. Selección de la variable respuesta

#### 4. Selección del diseño experimental

- ✓ Número de réplicas, orden de ejecución, definir o no bloques y restricciones en la aleatorización

#### 5. Ejecución del experimento

#### 6. Análisis estadístico de los datos

#### 7. Conclusiones y recomendaciones

Planificación pre-experimental

### SUPUESTOS DEL ANALISIS DE LA VARIANZA

- Correcta relación funcional y aditividad: el modelo debe contener todos aquellos factores que tienen un gran efecto en la variabilidad del fenómeno. El modelo debe ser aditivo, es decir, debe ser la suma de los factores que intervienen .
- Normalidad: Consiste en que los errores deben tener una distribución normal. Este requisito se debe a que se usa la distribución F de Fisher. Una manera de garantizar normalidad es tomar muestras de tamaño  $> 30$  (teorema del límite central). Una consecuencia de no normalidad es que altera el nivel de significación de la prueba F y de las pruebas de comparación de medias.
- Homogeneidad de varianzas: Los datos correspondientes a los diferentes tratamientos deben tener aproximadamente la misma varianza. Algunos autores sostienen que la prueba F no es afectada por la no homogeneidad, pero que las pruebas de comparación de medias (Duncan, Tukey, Scheffe) son mucho más vulnerables y nos podrían hacer caer en conclusiones erróneas.
- Independencia de errores: Implica que la ocurrencia de un determinado error en una unidad experimental, no debe afectar en la ocurrencia de un error dado en otra unidad experimental, es decir, los errores no deben estar correlacionados. La aleatorización no garantiza la independencia de los errores, sino que atenúa su efecto. Esta suposición es quizás la más restrictiva en el análisis de la varianza, ya que al estar correlacionados los errores se está favoreciendo o perjudicando a alguno (o algunos) de los tratamientos, provocando con ello errores en la estimación de las medias.

#### Verificación de los Supuestos

Detección de poblaciones no normales: para cada tratamiento

- construya un histograma, grafique el polígono de frecuencia. Busque distribuciones con un sesgo pronunciado. La utilidad de tales gráficas para datos pequeños es limitada.
- Construya una gráfica de probabilidad normal para los datos. Si estos son aproximadamente normales, los puntos caerán aproximadamente en una línea recta.

#### Detección de varianzas desiguales:

- Para cada tratamiento, utilizar una de las diferentes pruebas estadísticas formales para probar homogeneidad de las varianzas.

Pruebas para detectar fallas en los supuestos.

1. Pruebas de normalidad
  - (a) Kolmogorov-Smirnov
  - (b) Bondad de ajuste ( $\chi^2$ )

2. Homogeneidad de varianzas.

- (a) Bartlett
- (b) Box
- (c) Levene

## DISEÑO DE EXPERIMENTOS CON SPSS

### 3. Pruebas de independencia.

- (a)  $\chi^2$
- (b) Kendall

### 4. Pruebas de aditividad.

- (a) Tukey
- (b) Rojas

### Alternativas ante las fallas de los supuestos

- 1) No Normalidad. Mediante transformaciones o haciendo uso de la teoría de la distribución específica presentada
- 1) Varianzas No Homogéneas. Por medio de: Transformaciones, análisis ponderado y dividir el error en grupos con varianzas homogéneas.
- 1) No Independencia. Uso de análisis de covarianza.
- 1) Incorrecta relación funcional y no aditividad. Transformaciones o utilizando el modelo adecuado. Para observaciones aberrantes eliminarlas y usar la técnica de la parcela perdida.

### Transformaciones

La ventaja que ofrecen es que en determinados casos corrigen simultáneamente los datos en tres de los supuestos básicos (normalidad, homogeneidad y aditividad).

Principales transformaciones en la investigación agrícola:

- (a) Raíz cuadrada (b) Angular (arcsen) (c) Logarítmica
- (b) Exponencial

La logarítmica y exponencial se usan para corregir aditividad.

Cuando x está en porcentajes, usar arcsen( )

Cuando x en proporciones, usar arcsen( )

Logarítmica  $\log(x+1)$  se recomienda cuando los valores de x son pequeños

Exponencial,  $y = xp$  para  $p = 0.5$

## DISEÑO DE EXPERIMENTOS CON SPSS

**Variable respuesta, o explicada, o dependiente:** Variable asociada al fenómeno que nos interesa estudiar

**Variabes experimentales o factores:** Conjunto de variables que influyen en la respuesta y que se fijan a ciertos niveles o tratamientos antes del experimento

**Unidades experimentales:** Cada una de las observaciones correspondientes a un nivel determinado de cada uno de los factores que intervienen en el experimento

**Tamaño del experimento:** Número total de unidades experimentales que participan en el experimento

### TRES PRINCIPIOS BÁSICOS

#### 1. Replicación

Consiste en una repetición del experimento y permite obtener:

- Una estimación del error experimental, que constituye una unidad de medida para determinar si las diferencias observadas en los datos son realmente estadísticamente distintas.
- Una estimación más precisa de los efectos
- Es diferente de la toma de medidas repetidas, en la que la variabilidad observada es un reflejo de la variabilidad inherente del instrumento de medida o de otras fuentes de variabilidad presentes en la misma ejecución

#### 2. Aleatorización

- Es la piedra de toque en el uso de los métodos estadísticos en DOE. Una hipótesis básica es que las observaciones sean v.a. independientes y aleatorizar el material experimental y el orden de ejecución ayuda a: (a) evitar la dependencia entre las observaciones y (b) compensar los efectos de las variables ruido no controlables
- Los programas informáticos ayudan en este proceso de aleatorización. Cuando no es posible una aleatorización completa hay métodos especiales

#### 3. Definición de bloques

- Mejora la precisión en las comparaciones entre los factores de interés
- Reduce la variabilidad transmitida por las variables ruido, es decir, factores que pueden influir en la respuesta y en los que no se está directamente interesado
- En general, un bloque es un conjunto de condiciones experimentales relativamente homogéneas, así cada nivel del factor ruido define un bloque

### DISEÑOS COMPLETAMENTE AL AZAR (DCA)

La variable respuesta está en función de un solo factor (unifactorial) a distintos niveles o tratamientos los cuales son asignados a las unidades experimentales, las cuales son consideradas homogéneas (requisito fundamental).

En tal sentido, se definen el número de niveles o tratamientos (t) y la cantidad de réplicas (r) de tal manera que el total de unidades experimentales viene dada por:

$$\text{Total de unid experimentales} = t \times r$$

## DISEÑO DE EXPERIMENTOS CON SPSS

### VISUALIZACION DEL MODELO

#### Diseño DCA para un Factor

E	G	H	A
C	B	H	B
F	D	G	C
A	A	H	G
B	D	F	E
D	F	C	D
C	E	H	G
F	E	A	B

Número de niveles o tratamientos = 8

Réplicas = 4 (por tratamiento)

Total de unidades experimentales =  $8 \times 4 = 32$

La asignación de los tratamientos a las unidades experimentales se realiza de forma aleatoria.

#### MODELO MATEMÁTICO DEL DCA

$$y_{ij} = \mu_i + \epsilon_{ij}; \text{ con } i = 1::t; j = 1::n$$

$y_{ij}$  es la observación  $j$  para el nivel  $i$ ,  $\mu_i$  es el valor promedio del tratamiento  $i$  y  $\epsilon_{ij}$  es un error aleatorio.

#### CONTRASTE DE HIPÓTESIS PARA LAS MEDIAS DE LOS TRATAMIENTOS

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu_F = \mu_G = \mu_H$$

$$H_a: \mu_i \neq \mu_j \text{ para al menos un par } (i; j)$$

Para inferir si se acepta la hipótesis nula  $H_0$  o se rechaza (en este caso, se acepta  $H_a$ ) se procede a realizar un análisis de la varianza ANOVA.

Si se rechaza la hipótesis nula, esto significa que hay diferencias significativas entre los tratamientos. De tal manera, que hay que aplicar una prueba estadística para determinar entre cuales tratamientos existe diferencias.



## DISEÑO DE EXPERIMENTOS CON SPSS

### **Ejemplo:**

Ensayo de 8 variedades de trigo (tratamientos) considerando 4 replicas (D.C.A).  
Los rendimientos (Kg/p.útil) son:

3	5	3,2	7	2	8	1,8	1
2,8	3	1,4	2	2,4	8	1,6	2
3,5	6	2	4	2,9	7	3	3
2,5	1	1,5	1	2,7	8	2,7	7
1,7	2	2,7	4	2,5	6	3,5	5
2,8	4	2,8	6	1,9	3	2,1	4
3	3	2,5	5	2,2	8	3,7	7
2,4	6	2	5	2	1	1,8	2

Tamaño de Parcela útil 10 m<sup>2</sup>

Es recomendable construir una tabla por tratamiento y réplicas (facilita la introducción de los datos), tal como se muestra a continuación:

<b>1</b>	1,8	2,5	1,5	2
<b>2</b>	1,4	1,6	2	1,7
<b>3</b>	2,8	3	1,9	3
<b>4</b>	2	2,7	2,8	2,1
<b>5</b>	3	3,5	2,5	2
<b>6</b>	3,5	2,5	2,8	2,4
<b>7</b>	3,2	2,9	2,7	3,7
<b>8</b>	2	2,4	2,7	2,2

El paquete estadístico Statgraph muestra la tabla ANOVA que va a permitir establecer si se rechaza H<sub>0</sub> o si se acepta.

Variable dependiente: RENDIM

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
TRAT	199,730	8	24,966	125,616	,000
Error	4,770	24	,199		
Total	204,500	32			

Por tanto, si valor  $p < \alpha$  entonces se rechaza H<sub>0</sub> ; si valor  $p > \alpha$  se acepta H<sub>0</sub>

Valor p viene dada por la significación, es decir, valor  $p = 0$

Para el 95 % de confianza se tiene que  $\alpha = 0,05$ .

Por tanto, valor  $p = 0 < 0,05$ . Por lo que se rechaza H<sub>0</sub>.

## DISEÑO DE EXPERIMENTOS CON SPSS

Por lo tanto existe diferencias significativas entre el rendimiento promedio para las 8 especies de trigo. Para hallar las especies que tienen diferencias estadísticamente significativas se procede a realizar la prueba de Tukey.

RENDIM  
DHS de Tukey

TRAT	N	Subconjunto		
		1	2	3
B	4	1,6250		
A	4	1,9500	1,9500	
H	4	2,3250	2,3250	2,3250
D	4	2,4000	2,4000	2,4000
C	4		2,6750	2,6750
E	4		2,7500	2,7500
F	4		2,8000	2,8000
G	4			3,1250
Significación		,260	,172	,227

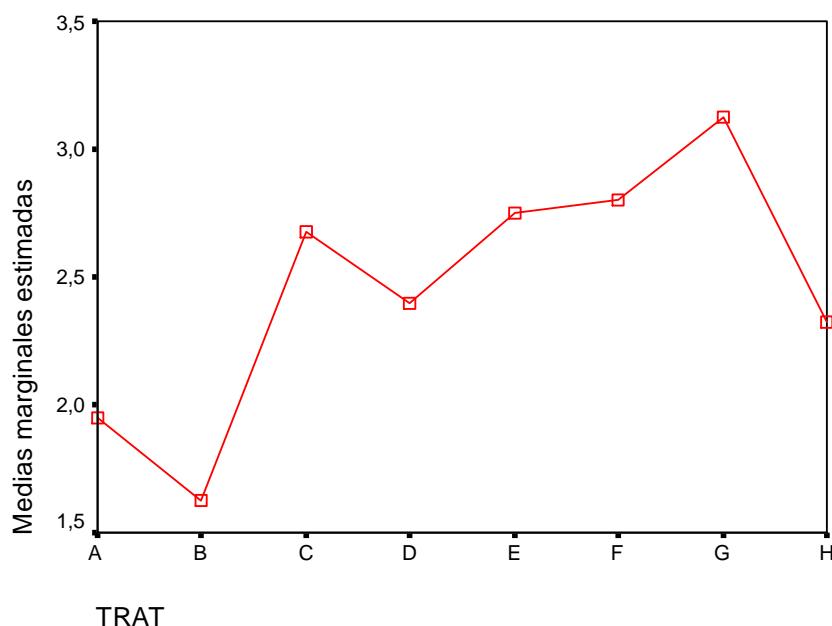
Se muestran las medias para los grupos en subconjuntos homogéneos.

Tal como se observa en la tabla se tienen tres subconjuntos homogéneos.

Con respecto al primer subconjunto para los tratamientos BAHD no existe diferencia estadísticamente significativa para el rendimiento promedio. De manera análoga, para los tratamientos AHDCEF (subconjunto 2) no hay diferencias significativas. Para los tratamientos HDCEFG (subconjunto 3) no hay diferencias significativas. Obsérvese que hay diferencia significativa entre el tratamiento B y los tratamientos CEFG; entre el tratamiento A y el tratamiento G.

En el siguiente gráfico se pueden visualizar los subconjuntos homogéneos

Medias marginales estimadas de RENDIM



## DISEÑO DE EXPERIMENTOS CON SPSS

### DISEÑO COMPLETAMENTE ALEATORIZADO POR BLOQUES (DCAB)

Si se sospecha que en el diseño hay un factor que genera variabilidad es importante considerarlo. De lo contrario el análisis de la varianza puede arrojar resultados no deseados, como es el caso que no se detecten diferencias significativas cuando en realidad si las hay. De ahí surge la necesidad de considerar un factor bloque.

Los **diseños aleatorizados** previenen de la influencia de factores ruido existentes, desconocidos y no controlables. Cuando se tiene en cuenta la presencia de un **factor ruido conocido y controlable**, cuya influencia no interesa directamente, **la definición de bloques** permite eliminar sistemáticamente su efecto y realizar comparaciones homogéneas entre los tratamientos.

- La **técnica de bloquear** el experimento se usa para reducir en lo posible el error experimental. Se desea controlar estadísticamente los efectos de la variable que define los bloques para **estimar con mayor precisión los efectos de otro factor sobre la variable respuesta**
- En cada bloque, los experimentos se realizan de forma aleatoria. Sin embargo, los bloques representan **una restricción sobre la aleatorización**
- El diseño se denomina **completo** porque cada bloque contiene todos los tratamientos
- La técnica de bloquear puede ser útil en situaciones en las que interesa comprobar **la robustez de un proceso** frente a condiciones ambientales o escenarios variadas difícilmente controlables

De tal manera, que un DCAB se tiene un factor (tratamiento) y un factor ruido (bloque), donde el número de tratamientos (t) y el número de bloques (b) definen el total de unidades experimentales, esto es:  $t \times b$

**Ejemplo:** Se quiere comprobar si hay diferencia significativas entre cuatro tipo de fertilizantes y la altura de las plántulas de pino entonces se pueden considerar dos situaciones: si las unidades experimentales están bajo condiciones más o menos homogéneas entonces se establece un DCA. En el caso contrario si hay variabilidad con respecto a la disposición de las unidades experimentales hay que introducir el factor bloque.

#### VISUALIZACIÓN DEL DISEÑO

BLOQUES			
B1	B2	B3	B4
C	A	D	B
A	D	C	D
B	C	A	A
D	B	B	C

Número de tratamientos = 4

Número de bloques = 4

Total de unidades experimentales =  $4 \times 4 = 16$

#### MODELO MATEMÁTICO DEL DCAB

$$y_{ij} = \bar{\mu} + \text{trat}_i + \text{Bloq}_j + \bar{\epsilon}_{ij} \text{ con } i = 1::t \text{ y } j = 1::b$$

## DISEÑO DE EXPERIMENTOS CON SPSS

### CONTRASTE DE HIPOTESIS

Para este tipo de diseño hay que realizar dos contrastes de hipótesis:

#### Para Tratamientos

$$H_0: \bar{\mu}_A = \bar{\mu}_B = \bar{\mu}_C = \bar{\mu}_D$$

$$H_a: \bar{\mu}_i \neq \bar{\mu}_j \text{ para al menos un par } (i; j)$$

#### Para Bloques

$$H_0: \bar{\mu}_{B1} = \bar{\mu}_{B2} = \bar{\mu}_{B3} = \bar{\mu}_{B4}$$

$$H_a: \text{al menos una media del bloque es diferente}$$

#### Ejemplo:

Ensayo de 6 especies de maderas. Realizar un análisis de la varianza para un diseño

D.B.C.A, considerando 6 bloques. Se midió la altura de las plántulas en cm:

Especie	1	2	3	4	5	6
<b>Bloques</b>						
1	13,85	11,85	12,95	14,5	12,5	10,5
2	12,2	13,55	11,3	15,3	8,9	10,65
3	13,05	11,5	10,65	10,85	8,85	8,5
4	10,75	10,15	13,15	10,6	9,48	11,36
5	12,85	13,05	8,95	12,45	9,6	9,6
6	8,05	6,85	10,5	11,15	7,23	4,7

Si no se considerará el efecto bloque, la tabla ANOVA mostraría lo siguiente:

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	,532	3	,177	,042	,988
Intra-grupos	50,527	12	4,211		
Total	51,059	15			

Lo que significa que el valor  $p = 0,988 > \alpha = 0,05$  (95 %). Por lo tanto, se aceptaría  $H_0$ . Es decir, no habría diferencia estadísticamente significativa entre las especies. El hecho de no considerar el factor bloque (cuando hay variabilidad o heterogeneidad) enmascara la información, lo que implica que se realice una mala inferencia estadística.

Veamos la tabla ANOVA considerando el factor Bloque.

Pruebas de los efectos inter-sujetos

Variable dependiente: ALTURA

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
<b>Intersección</b>	<b>4266,702</b>	<b>1</b>	<b>4266,702</b>	<b>1984,041</b>	<b>,000</b>
<b>ESPECIES</b>	<b>50,775</b>	<b>5</b>	<b>10,155</b>	<b>4,722</b>	<b>,004</b>
<b>BLOQUE</b>	<b>74,880</b>	<b>5</b>	<b>14,976</b>	<b>6,964</b>	<b>,000</b>
Error	53,763	25	2,151		
Total	4446,120	36			

## DISEÑO DE EXPERIMENTOS CON SPSS

Tal como se observa en la tabla ANOVA, el valor p para especies y bloques es menor que  $\alpha = 0,05$ . Por lo que se rechaza la hipótesis nula. Por tanto, hay diferencias significativas tanto por especies como por bloques. Para detectar entre cuales tratamientos y bloques hay diferencias se realiza una prueba de rangos múltiples de Tukey. A continuación se muestran los subconjuntos homogéneos.

ALTURA  
DHS de Tukey

	N	Subconjunto	
ESPECIES		1	2
6,00	6	9,2183	
5,00	6	9,4267	
2,00	6	11,1583	11,1583
3,00	6	11,2500	11,2500
1,00	6	11,7917	11,7917
4,00	6		12,4750
Significación		,055	,634

ALTURA  
DHS de Tukey

	N	Subconjunto	
BLOQUE		1	2
6,00	6	8,0800	
3,00	6	10,5667	10,5667
4,00	6		10,9150
5,00	6		11,0833
2,00	6		11,9833
1,00	6		12,6917
Significación		,068	,159

Datos Procesados en SPSS.10

### DISEÑO CON TRES FACTORES

**Ejercicio:** Determinar si hay diferencia estadísticamente significativa para los tres factores.

**Ensayo de 5 variedades de papa. DCL (para controlar fertilidad y pendiente).**

**Resultados en Kg por parcela de 6.4 m<sup>2</sup>.**

Hil	Col	I	II	III	IV	V
I		A(1.7)	B(3)	C(2.5)	D(3)	E(0.6)
II		E(2)	C(3.6)	D(3.2)	B(4)	A(3)
III		C(3)	D(3)	A(3)	E(2)	B(2.4)
IV		B(3)	A(2.7)	E(1.1)	C(2.2)	D(3.1)
V		D(3.2)	E(1.7)	B(5.4)	A(2.6)	C(3.3)

## DISEÑO DE CUADRO LATINO

Estos diseños también utilizan el principio de **definición de bloques para eliminar dos fuentes de variabilidad**

Se bloquea en dos direcciones, lo que supone **dos restricciones sobre la aleatorización**. El diseño se utiliza para eliminar dos fuentes de variabilidad, bloqueando en dos direcciones, de forma que ambos factores son ortogonales a los tratamientos.

Supongamos que deseamos determinar el efecto de cinco fertilizantes diferentes A, B, C, D y E sobre la cosecha de patatas y que poseemos una finca que podemos dividir en  $5 \times 5 = 25$  parcelas (ni la finca ni las distintas parcelas tienen que ser cuadrados perfectos). La disposición del cuadrado latino se muestra en la figura siguiente y posee la propiedad de que cada fila y cada columna recibe cada tratamiento (fertilizante) exactamente una sola vez. Supongamos, por ejemplo, que la humedad del suelo en la finca varía a lo largo de una dirección y que el gradiente de fertilidad lo hace perpendicular a la anterior. Estas influencias del entorno pueden eliminarse utilizando este diseño.

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>Fila I</b>
<b>C</b>	<b>D</b>	<b>E</b>	<b>A</b>	<b>B</b>	<b>Fila II</b>
<b>E</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>Fila III</b>
<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>A</b>	<b>Fila IV</b>
<b>D</b>	<b>E</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>Fila V</b>
<b>Col.1</b>	<b>Col.2</b>	<b>Col.3</b>	<b>Col.4</b>	<b>Col.5</b>	

**Ejemplo:** Interesa comparar los efectos de 5 formulaciones diferentes de un propulsor sobre la fuerza explosiva. Un lote de materia prima no es suficiente para comprobar el efecto de las formulaciones, además éstas se preparan por diferentes operarios, con diferentes habilidades y experiencia.

Lotes de materia prima	Operadores				
	1	2	3	4	5
1	A=24	B=20	C=19	D=24	E=24
2	B=17	C=24	D=30	E=27	A=36
3	C=18	D=38	E=26	A=27	B=21
4	D=26	E=31	A=26	B=23	C=22
5	E=22	A=30	B=20	C=29	D=31

El modelo estadístico para un cuadrado latino es:

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{ijk}, \quad \begin{cases} i = 1, \dots, p \\ j = 1, \dots, p \\ k = 1, \dots, p \end{cases}$$

$y_{ijk}$  es la observación de la fila  $i$  y columna  $k$  para el tratamiento  $j$ ,  $\mu$  es la media global,  $\alpha_i$  es el efecto de la fila  $i$ ,  $\tau_j$  es el efecto del tratamiento  $j$ ,  $\beta_k$  es el efecto de la columna  $k$  y  $\epsilon_{ijk}$  es el término del error.

El modelo de efectos del cuadrado latino es **aditivo**, es decir, no hay interacciones entre las filas, las columnas y los tratamientos.

Hay una sola observación por celda. Cada tratamiento aparece exactamente una vez en cada fila y en cada columna.

### **ANÁLISIS DE LA VARIANZA**

#### **TGreantaemraile Lnitnoear Model: Explosion versus Filas; Columnas;**

Factor Type Levels Values

Filas-lotes fixed 5 1 2 3 4 5

Columnas fixed 5 1 2 3 4 5

Tratamie fixed 5 1 2 3 4 5

#### **Analysis of Variance for Explosio, using Adjusted SS for Tests**

<b>FUENTE</b>	<b>GL</b>	<b>Seq SS</b>	<b>AdjSS</b>	<b>AdjMS</b>	<b>F</b>	<b>P</b>
Filas	4	68,00	68,00	17,00	1,59	0,239
Columnas	4	150,00	150,00	37,50	3,52	0,040
Tratamie	4	330,00	330,00	82,50	7,73	0,003
Error	12	128,00	128,00	10,67		
Total	24	676,00				

## ANEXOS

### Diagnos y adecuación del modelo

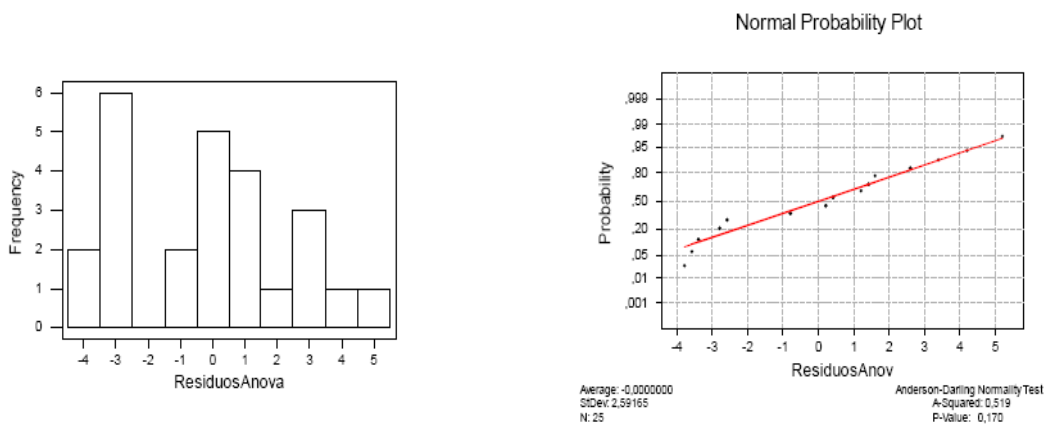
Comprobar que

El **análisis de los residuos** constituye una herramienta fundamental en el estudio de la adecuación del modelo. Los residuos contienen información sobre la variabilidad no explicada.

#### HIPOTESIS DE NORMALIDAD

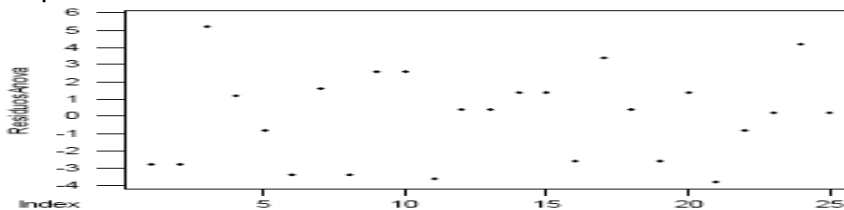
- Un histograma de los residuos, un gráfico de probabilidad normal de los residuos y un contraste de normalidad
- El análisis de varianza en el modelo de efectos fijos es robusto a la hipótesis de normalidad. La falta de normalidad puede afectar a la precisión de la estimación de la varianza.
- Los gráficos de los residuos son de gran ayuda en la detección de **outliers**.

### Ejemplo:



#### HIPOTESIS DE INDEPENDENCIA

- La representación de los residuos en el orden de recogida de datos ayuda a detectar correlación entre los residuos.
- Si se identifica dependencia**, los cálculos sobre la precisión de los estimadores son erróneos
- El no cumplimiento de esta hipótesis** es un problema serio y difícil de corregir, por ello es importante prevenir el problema en la fase de recogida de datos. Una **aleatorización apropiada** es un paso importante para obtener independencia.





## HIPOTESIS DE VARIANZA CONSTANTE

- Gráficos de los residuos frente al tiempo.
- Los gráficos de los residuos frente a los valores ajustados detectan, en ocasiones, que la varianza crece al aumentar el nivel del factor.
- La **heterocedasticidad** también surge cuando los datos no siguen un comportamiento normal, dado que para las distribuciones asimétricas la varianza tiende a ser una función de la media.
- Contrastes estadísticos sobre la igualdad de varianzas:**
- Contraste de Bartlett:** Basado en las cuasivarianzas muestrales de cada tratamiento asume la normalidad de las poblaciones, siendo muy sensible a esta hipótesis
- Contraste de Levene modificado:** Basado en las diferencias absolutas de las observaciones y su mediana, es un contraste robusto a la falta de normalidad

## HIPOTESIS DE VARIANZA CONSTANTE

- El no cumplimiento de la hipótesis** afecta poco al contraste F si el diseño es de efectos fijos y balanceado.
- La aproximación más común en el tratamiento de varianza no constante es aplicar una **transformación sobre los datos** que la estabilice.
- Selección empírica** de la transformación:

**Comparación de las medias con un control:** Se comparan los  $a-1$  tratamientos con el control.

- ✓ **Procedimiento de Dunnett:** Basado en la diferencia de las medias de cada tratamiento con el control y con un nivel de significación global fijado  $\alpha$

$$\begin{aligned} H_0 &: \mu_i = \mu_a & i = 1, \dots, a - 1 \\ H_1 &: \mu_i \neq \mu_a \end{aligned}$$

El análisis de varianza se ha centrado en determinar qué niveles del factor provocan diferencias en las medias de los tratamientos **efectos de localización**  
Cuando el interés se centra en descubrir los niveles del factor que afectan a la variabilidad, se trata de descubrir los **efectos de dispersión**

Sin título - Editor de datos SPSS

Archivo Edición Ver Datos Transformar **Analizar** Gráficos Utilidades Ventana ?

15:

	tiempo	maquina	agua
1	24,00	1	1
2	26,00	1	1
3	29,00	1	1
4	27,00	2	1
5	30,00	2	1
6	32,00	2	1
7	26,00	3	1
8	27,00	3	1
9	30,00	3	1
10	25,00	4	1
11	28,00	4	1

- Informes
- Estadísticos descriptivos
- Tablas personalizadas
- Comparar medias
- Modelo lineal general**
  - Univariante...**
  - Multivariante...
  - Medidas repetidas...
  - Componentes de la varianza...
- Modelos mixtos
- Correlaciones
- Regresión
- Loglineal
- Clasificar
- Reducción de datos
- Escalas
- Pruebas no paramétricas
- Series temporales
- Supervivencia
- Respuestas múltiples

Univariante

Dependiente: tiempo

Factores fijos: Máquina (factor) [m], Calidad del agua [b]

Factores aleatorios:

Covariables:

Ponderación MCP:

Modelo...  
Contrastes...  
Gráficos...  
Post hoc...  
Guardar...  
Opciones...

Aceptar Pegar Restablecer Cancelar Ayuda

Especificar modelo

Factorial completo  Personalizado

Factores y covariables: maquina(F), agua(F)

Modelo: maquina, agua

Construir términos: Efectos princip.

Suma de cuadrados: Tipo III  Incluir la intersección en el modelo

Continuar Cancelar Ayuda

**Univariante: Opciones** [X]

Medias marginales estimadas

Factores e interacciones de los factores: (GLOBAL)  
maquina  
agua

Mostrar las medias para: (GLOBAL)  
maquina  
agua

Comparar los efectos principales

Ajuste del intervalo de confianza:  
Bonferroni

Mostrar

<input type="checkbox"/> Estadísticos descriptivos	<input type="checkbox"/> Pruebas de homogeneidad
<input type="checkbox"/> Estimaciones del tamaño del efecto	<input checked="" type="checkbox"/> Diagramas de dispersión x nivel
<input type="checkbox"/> Potencia observada	<input checked="" type="checkbox"/> Gráfico de los residuos
<input type="checkbox"/> Estimaciones de los parámetros	<input type="checkbox"/> Falta de ajuste
<input type="checkbox"/> Matriz de coeficientes de contraste	<input type="checkbox"/> Función estimable general

Nivel de significación: .05    Los intervalos de confianza son 95%

[Continuar] [Cancelar] [Ayuda]