Análisis de la Varianza

Prof. Susana Martín Fernández

Índice

- Análisis de la Varianza de un Factor
- Análisis de la Varianza de dos Factores
- Análisis de la Varianza de dos Factores con Interacción

Objetivo

- Estudiar la influencia de 1 o más factores en los valores de una variable aleatoria.

Procedimiento

- Descomponer la variabilidad de un experimento en componentes o factores independientes

Metodología

1. Representación gráfica de los datos.
2. Planteamiento del modelo.
3. Estimación de los parámetros.
4. Contraste de si los factores influyen o no en la variable aleatoria.
5. Comprobación de las hipótesis básicas por análisis de residuos.

Análisis de la Varianza de un Factor

1. Representación gráfica de los datos.

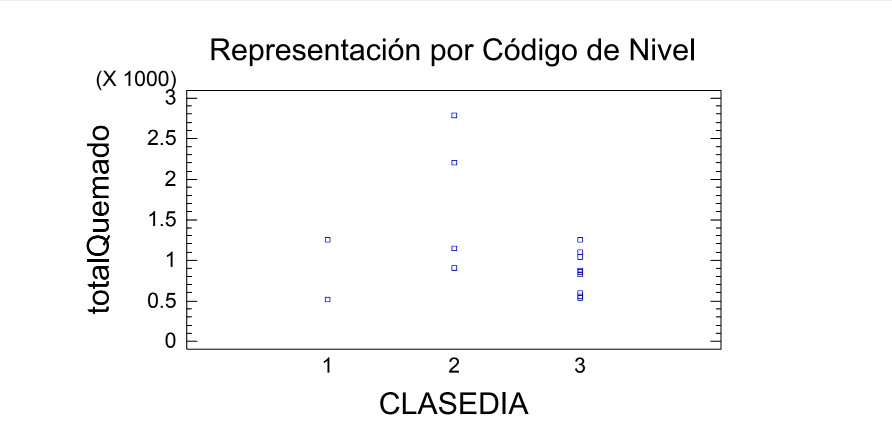
EJ. Se tienen los datos históricos de los incendios forestales en la Comunidad de Madrid. Se quiere estudiar la influencia en la superficie quemada, del tipo de día de la semana en el que se inicia el incendio.

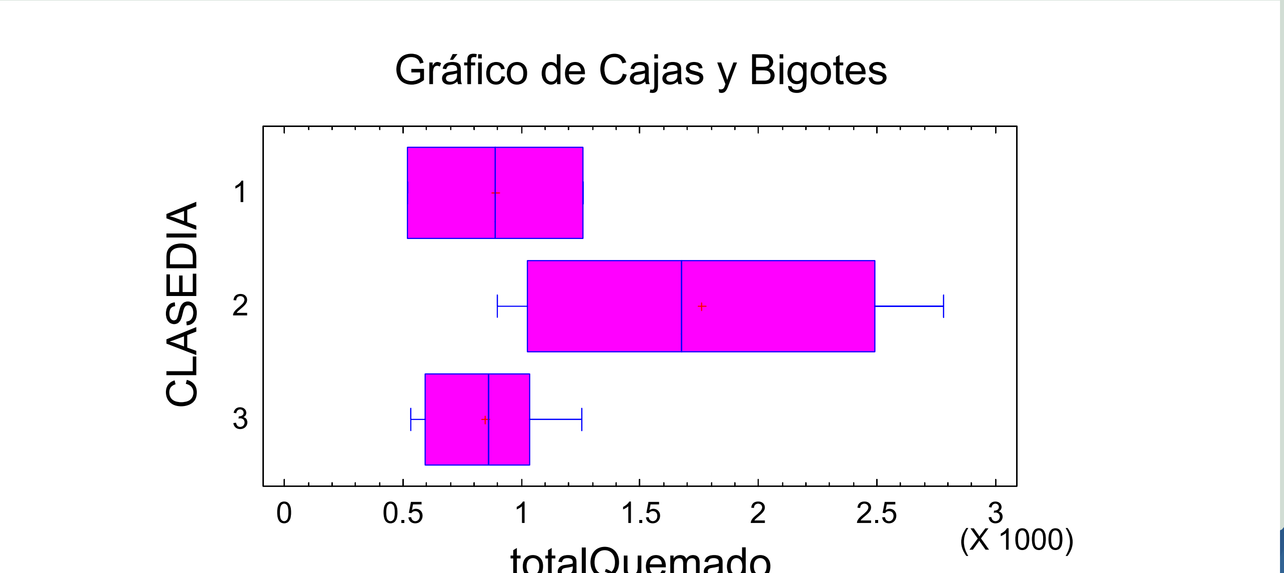
Tipo de día:

1- Festivo

2- Víspera de festivo

3-Laborable





2. Planteamiento del modelo

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1,2,\ldots,n_i;\; i = 1,2,\ldots,k$$

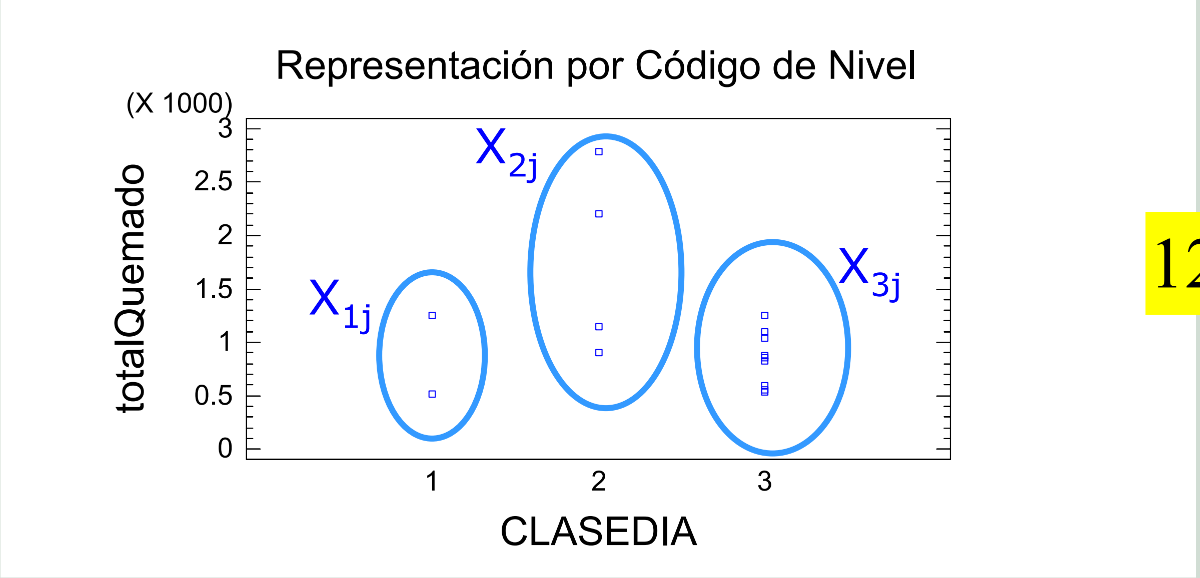
Donde

X_{ij} es el j-ésimo valor para el nivel i del factor.

μ_i es el valor medio de la variable para el nivel i del factor.

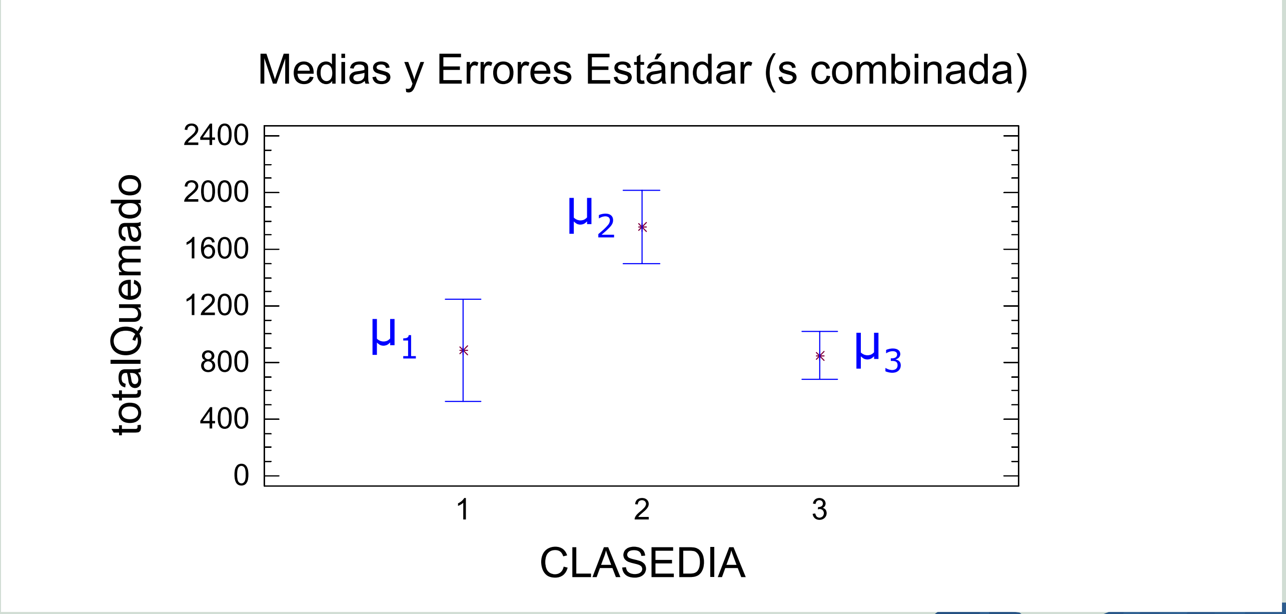
ε_{ij} es la perturbación aleatoria, variable que se supone Normal, de varianza constante, media nula e independiente.

2. Planteamiento del modelo



$$X_{11} = \mu_1 + \varepsilon_{11}$$

$$1257 = 888{,}5 + \varepsilon$$



2.Planteamiento del modelo

El modelo de forma matricial:

$$X = \beta A' + \varepsilon$$

Donde:

$$X = (X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, X_{22}, \ldots, X_{2n_2}, \ldots, X_{k1}, X_{k2}, \ldots, X_{kn_k})$$

Es un vector aleatorio de *n* componentes

$$\beta = (\mu_1, \mu_2, \ldots, \mu_k)$$

Es un vector de *k* parámetros desconocidos.

$A' = I_n$

$$\varepsilon = (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{1n_1}, \varepsilon_{21}, \varepsilon_{22}, \ldots, \varepsilon_{2n_2}, \ldots, \varepsilon_{k1}, \varepsilon_{k2}, \ldots, \varepsilon_{kn_k})$$

Es un vector aleatorio que recoge el error de medición de la variable *X.*

3.Estimación de los parámetros

Por el método de máxima verosimilitud a partir de la siguiente función de verosimilitud:

$$f(X,\mu_1,\mu_2,\dots,\mu_k) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\mu_i)^2}$$

Los estimadores de los parámetros son los siguientes:

$$\hat{\mu}_i = \overline{X}_i = \frac{\sum X_{ij}}{n_i}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2}{n}$$

3.Estimación de los parámetros

En el ejemplo, el valor de los estimadores de los parámetros son los siguientes:

Tipo de día:
1- Festivo
2- Víspera de festivo
3-Laborable

$$\overline{X}_1 = 888'5 \text{ ha.}$$

$$\overline{X}_2 = 1757'88 \text{ ha.}$$

$$\overline{X}_3 = 848'112 \text{ ha.}$$

$$\hat{\sigma} = 628'366 \text{ ha.}$$

4. Contraste para analizar la influencia del factor

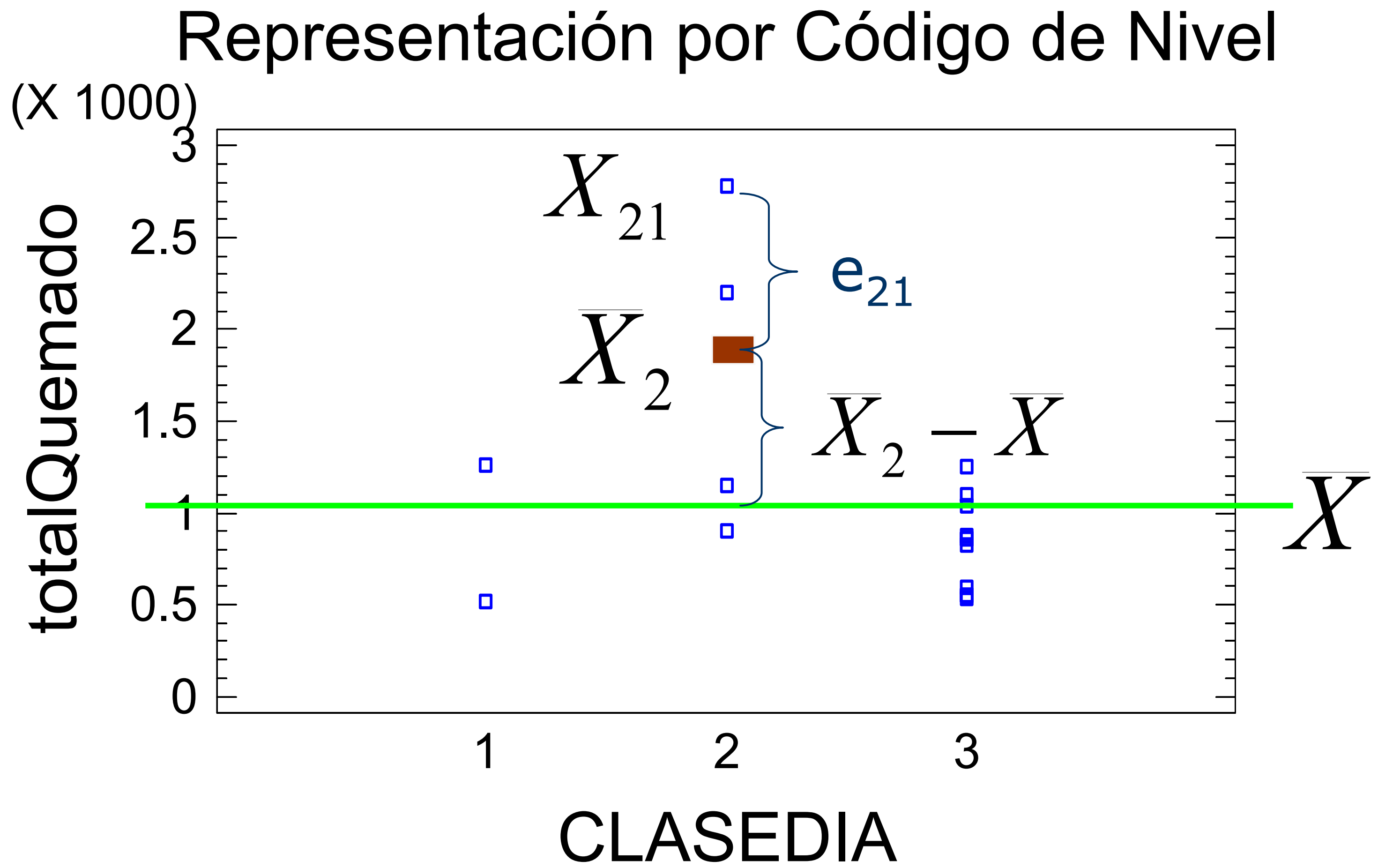
Lo que se trata de comprobar es que los factores no influyen en la variable *X*. Para ello la hipótesis más sencilla es comprobar si las medias son iguales para todos los factores:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$$

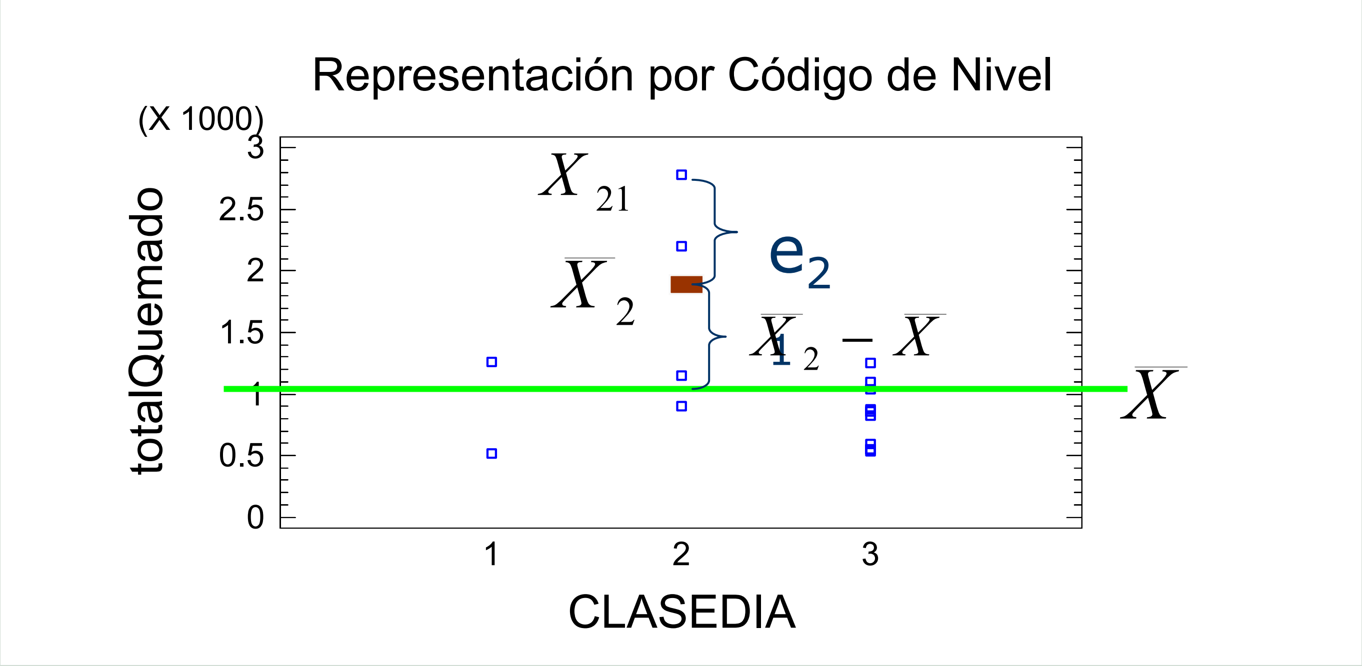
Se rechazará la hipótesis nula cuando fijado un nivel de significación el estadístico sea mayor que F_0, valor obtenido en la tabla de la F-snedecor para *(k-1,n-k)* grados de libertad.

$$\frac{(n-k)}{(k-1)} \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X})^2 - \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i.})^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i.})^2} > F_0$$

4. Contraste para analizar la influencia del factor. Descomposición de la variabilidad



4. Contraste para analizar la influencia del factor. Descomposición de la variabilidad



$$X_{21} - \overline{X} = \left(X_{21} - \overline{X}_2\right) + \left(\overline{X}_2 - \overline{X}\right)$$

4.Contraste para analizar la influencia del factor. Descomposición de la variabilidad

La variación entre los datos y la media total, se puede poner como suma de la variación de los datos y las medias parciales y la de las medias parciales y la total.

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X})^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+\sum_{i=1}^{k}n_i(\overline{X}_i-\overline{X})^2$$

VT=VNE+VE

En el contraste se comprueba:

$$F=\frac{VT-VNE}{VNE}\frac{n-k}{k-1}>F_0$$

4.Contraste para analizar la influencia del factor

La forma de trabajar es calculando la tabla de Análisis de la Varianza (ANOVA):

Fuentes de variación	**Suma de cuadrados**	**Grados de libertad**	**Varianzas**
Entre grupos (Varianza explicada)	$\sum n_i(\overline{X}_{i.}-\overline{X}_{..})^2$	**k-1**	$\hat{S}_e^2 = \frac{VE}{K-1}$
Interna o no explicada	$\sum\sum(X_{ij}-\overline{X}_{i.})^2$	**n-k**	$\hat{S}_R^2 = \frac{VNE}{n-K}$
TOTAL	$\sum\sum(X_{ij}-\overline{X}_{..})^2$	**n-1**	$\hat{S}^2_y$

4.Contraste para analizar la influencia del factor

Las fuentes de variación siguen distribuciones chi-cuadrado. Otra forma de expresar el estadístico es:

$$F_{(k-1,n-k)} = \frac{\hat{S}_e^2}{S_R^2}$$

El coeficiente de determinación R^2=VE/VT es una medida relativa de la variabilidad explicada por el modelo respecto a la total.

4.Contraste para analizar la influencia del factor

abla ANOVA para totalQuemado según CLASEDIA

Análisis de la Varianza

uente	Sumas de cuad.	Gl	Cuadrado Medio	Cociente-F	P-Valor
ntre grupos	2.39146E6	2	1.19573E6	4.57	0.0334
ntra grupos	3.13635E6	12	261362.0		
otal (Corr.)	5.52781E6	14			

Ejemplo rápido de ANOVA de un factor.

El Servicio de Parques y Jardines del Ayuntamiento de Madrid está realizando un estudio de las características morfológicas del arbolado. Se han tomado 124 datos de Ligustrum japonica. Las variables que se midieron son:

Perímetro del tronco

Diámetro de copa

Altura de la primera rama

Altura

Nivel de riesgo

Se quiere analizar si la "distancia a la fachada" influye en el "perímetro del tronco"

Distancia	
P	d< 1m
M	1-2,5 m
G	d>2,5 m

Distancia	Perímetro-CM
M	46
M	36
M	51
G	40
G	32
G	23

5. Comprobación de las hipótesis básicas por análisis de residuos

- Normalidad
- Independencia
- Homocedasticidad

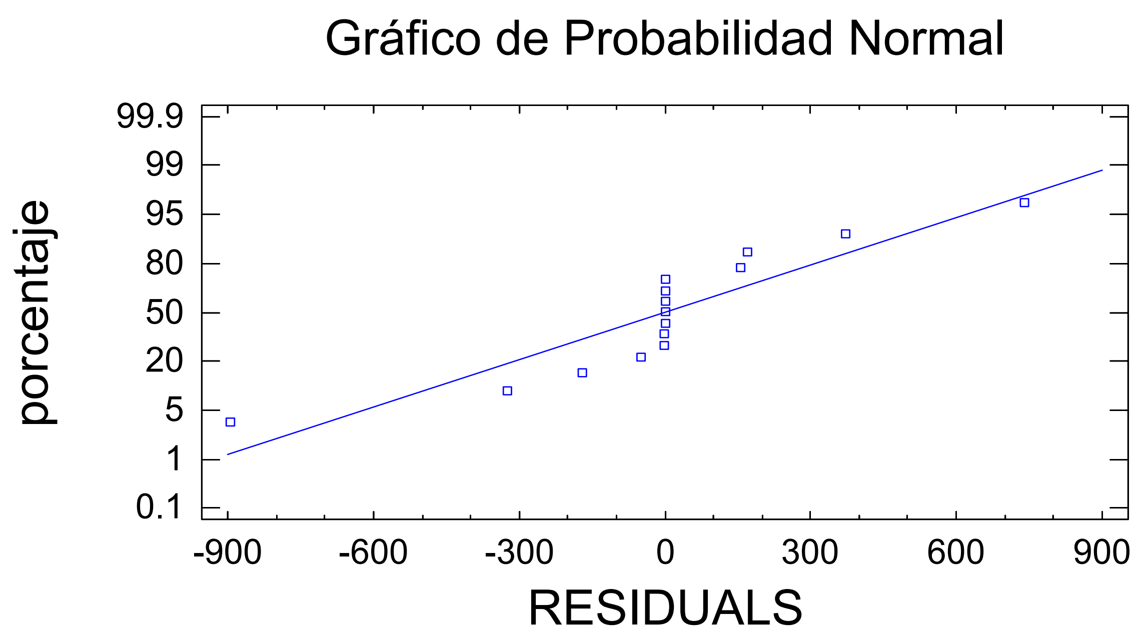
5. Comprobación de las hipótesis básicas por análisis de residuos

- Tests no paramétricos de bondad de ajuste:
  - Gráfico de Normalidad-Test de normalidad de Shapiro-Wilks
  - X^2
  - Kolmogorov-Smirnov
- Contrastes de Asimetría y Curtosis

5.Comprobación de las hipótesis básicas por análisis de residuos

Gráfico de Normalidad

En el eje X están representados los residuos o la variable a analizar. El eje Y tiene una escala de forma que la función de distribución aparezca como un recta.



5.Comprobación de las hipótesis básicas por análisis de residuos

Test de normalidad de Shapiro-Wilks

Este test, es el mismo que el de la recta, es específico para contrastar si una muestra procede de la distribución Normal, sin tener que hacer ninguna especificación de los parámetros.

Muy útil para muestras pequeñas con n<50.

El estadístico es el siguiente:

$$W = \frac{\left(\sum_{i=1}^{n/2} a_{n-1} \left(u_{n-i+1} - u_i \right) \right)^2}{\sum_{i=1}^{n} \left(u_i - \overline{u} \right)^2}$$

Donde a_i está tabulada, y u_i, es la muestra ordenada de menor a mayor.

$$u_1 \le u_2 \le \ldots \le u_n$$

La región crítica es la siguiente:

$$\mathrm{P}\left(\mathrm{W} \le \mathrm{K}/\mathrm{H}_0 \right) = \alpha$$

K está tabulada

5.Comprobación de las hipótesis básicas por análisis de residuos

Test de normalidad de Shapiro-Wilks

En el ejemplo, se aceptaría normalidad para un nivel de confianza del 99%.

```
Estadístico W de Shapiro-Wilks = 0.848611
P-valor = 0.0162002
```

5.Comprobación de las hipótesis básicas por análisis de residuos

X^2

- H_0: X F(X)
- Contraste válido para variables discretas y continuas. En el caso discreto se va a realizar una comparación punto por punto entre los datos muestrales y los de la distribución teórica, en el caso continuo, se comparan intervalos.

5. Comprobación de las hipótesis básicas por análisis de residuos

$\mathbf{X^2}$

Procedimiento para el caso continuo:

1. Enunciar la hipótesis nula.
2. Si algún parámetro de F(X) es desconocido se estima a partir de la muestra.
3. Se divide el rango de variación de X en intervalos disjuntos, $I_1, I_2, \ldots, I_k$.
4. Se calculan las frecuencias observadas en cada intervalo o clase, $f_1, f_2, \ldots, f_k$, es decir el número de observaciones en cada clase: $\sum f_i = n$

5. Comprobación de las hipótesis básicas por análisis de residuos

X^2

Procedimiento para el caso continuo:

5. Se calculan las probabilidades en estos intervalos con la función de distribución teórica; es decir $P\{X \in I_i\} = p_i$, $i=1,\ldots k$.

$$p_i = \int_{I_i} f(x)\,dx$$

6. Se calculan las frecuencias teóricas en cada intervalo, $f_{Ti} = np_i$ $i=1,\ldots k$

5. Comprobación de las hipótesis básicas por análisis de residuos

χ^2

Procedimiento para el caso continuo:

7. Se calcula el siguiente estadístico:

$$\chi = \sum_{i=1}^{k} \frac{\left(f_i - np_i\right)^2}{np_i} \rightarrow \chi^2_{k-r-1}$$

r es el número de parámetros estimados

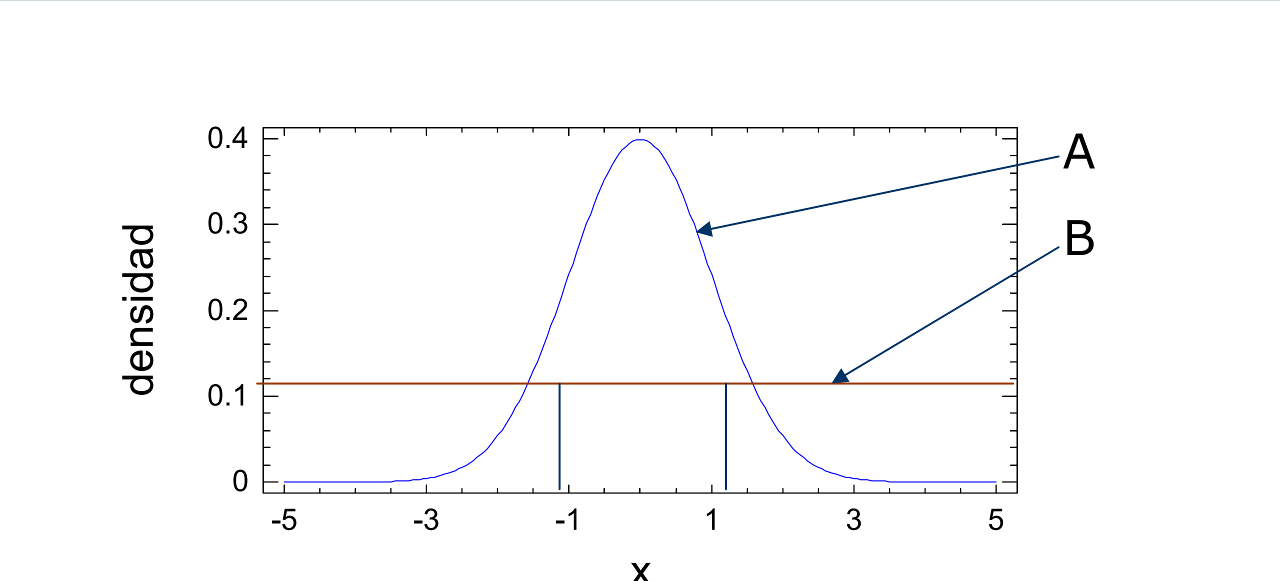
5.Comprobación de las hipótesis básicas por análisis de residuos

X^2

Condiciones de aplicación del contraste X^2:

1. Cada intervalo debe tener al menos 5 datos.

2.El número de intervalos se recomienda que sea como mínimo 5, para evitar que modelos diferentes tengan iguales frecuencias teóricas.



5.Comprobación de las hipótesis básicas por análisis de residuos

X^2

En el ejemplo:

Tests de Bondad de Ajuste para RESIDUALS

Contraste Chi-cuadrado

Límite Inferior	Límite Superior	Frecuencia Observada	Frecuencia Esperada	Chi-cuadrado
menor o igual	-291.304	2	3.00	0.33
-291.304	-87.6893	1	3.00	1.33
-87.6893	87.6895	8	3.00	8.33
87.6895	291.304	2	3.00	0.33
mayor 291.304		2	3.00	0.33

Chi-cuadrado = 10.6661 con 2 g.l.   P-Valor = 0.00482929

5.Comprobación de las hipótesis básicas por análisis de residuos

Contraste de Kolmogorov-Smirnov

- H_0: $X \equiv F(X)$
- Contraste válido para variables continuas.

Procedimiento:

1. Se ordenan los valores muestrales de menor a mayor:

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$$

5. Comprobación de las hipótesis básicas por análisis de residuos

Contraste de Kolmogorov-Smirnov

2. Se calcula la función de distribución empírica de la muestra, $F_n^*(x)$, con:

$$F_n^*(x) = \begin{cases} 0 & x < x_{(1)} \\ \frac{r}{n} & x_{(r)} \leq x < x_{(r+1)} \\ 1 & x \geq x_{(n)} \end{cases}$$

5. Comprobación de las hipótesis básicas por análisis de residuos

Contraste de Kolmogorov-Smirnov

3. Se calcula la discrepancia máxima entre las funciones de distribución observada y la teórica.

$$D_n = \max\left|F_n^*(x) - F(x)\right| \quad \text{Estadístico de 2 colas de K - S}$$

5.Comprobación de las hipótesis básicas por análisis de residuos

Contraste de Kolmogorov-Smirnov

Por tanto, para aplicar el test hay que calcular para cada punto x_h:

$$D_n(x_h) = \max\left\{\left|F_n^*(x_{h-1}) - F(x_h)\right|, \left|F_n^*(x_h) - F(x_h)\right|\right\}$$

Se acepta H_0 cuando $D_n < D_0$ tabulado

5.Comprobación de las hipótesis básicas por análisis de residuos

Contraste de Kolmogorov-Smirnov

En el ejemplo:

```
Estadístico DN global de Kolmogorov = 0.24321
P-Valor aproximado = 0.339936
```

5.Comprobación de las hipótesis básicas por análisis de residuos

Contrastes de asimetría y curtosis

◆Asimetría

El coeficiente de asimetría es :

$$CA = \frac{\sum (x_i - \bar{x})^3}{ns^3}$$

Es 0 si la hipótesis de normalidad es cierta. Si n>50, se aproxima a una normal y se puede contrastar si CA=0.

◆Curtosis o apuntamiento

$$CAP = \frac{\sum (x_i - \bar{x})^4}{ns^4}$$

Es 3 si la hipótesis de normalidad es cierta. Si n>200, se aproxima a una normal y se puede contrastar si CAP=3.

5.Comprobación de las hipótesis básicas por análisis de residuos

Contrastes de asimetría y curtosis

En el ejemplo:

```
Puntuación Z para asimetría = 0.733942
P-valor = 0.462982

Puntuación Z para curtosis = 2.30618
P-valor = 0.0211003
```

5. Comprobación de las hipótesis básicas por análisis de residuos

Análisis de dependencia de los residuos

- Coeficiente de autocorrelación.
- Contraste de Durbin-Watson

5. Comprobación de las hipótesis básicas por análisis de residuos

Análisis de dependencia de los residuos

- Coeficiente de autocorrelación.

$$r(k)=\frac{\sum_{i=k+1}^{n}\left(x_i-\bar{x}\right)\left(x_{i-k}-\bar{x}\right)}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$$

Donde k es el retardo

El coeficiente representa la correlación lineal entre las variables $X=(x_{k+1}, x_{k+2}, \ldots, x_n)$ e $Y=(x_1, x_2, \ldots, x_{n-k})$

5.Comprobación de las hipótesis básicas por análisis de residuos

Análisis de dependencia de los residuos

- Coeficiente de Durbin-Watson

$$D = \frac{\sum_{i=2}^{n} \left(e_i - e_{i-1}\right)^2}{\sum_{i=1}^{n} e_i^2}$$

Donde e_i son los residuos

Si D=0 hay dependencia positiva entre los residuos.

Si D=2 los residuos son independientes.

Si D=4 hay dependencia negativa entre los residuos.

5.Comprobación de las hipótesis básicas por análisis de residuos

Análisis de dependencia de los residuos

- Coeficiente de Durbin-Watson

Si k es el número de variables explicativas y n el tamaño de la muestra, para $\alpha=0,05$ están tabulados los valores d_L y d_U para aceptar o no independencia.

Si $0<D<d_L$ hay dependencia positiva.

Si $d_L<D<d_U$ el test no es concluyente.

Si $d_U<D<4-d_U$ los residuos son independientes.

Si $4-d_U<D<4-d_L$ el test no es concluyente.

Si $4-d_L<D<4$ dependencia negativa.

5. Comprobación de las hipótesis básicas por análisis de residuos

Análisis de la Homocedasticidad

$$H_0: \sigma_1 = \sigma_2 = \ldots = \sigma_k$$

- Contraste de Bartlett
- Contraste C de Cochran
- Contraste de Hartley
- Contraste de Levene

Análisis de la Homocedasticidad

Contraste	Población Normal	Tamaño Muestral Grupos
Bartlett	Si	cualquiera
Cochran	Cualquiera	Iguales
Hartley	Si	Iguales
Levene	Cualquiera	Cualquiera

Análisis de la Homocedasticidad

En muestras normales Bartlett es más sensible que Levene.

Los contrastes de Hartley y Cochran en general dan los mismos resultados.

Contraste de Varianza
Contraste C de Cochran: 0.764026 P-valor = 0.0
Contraste de Bartlett: 1.38176 P-valor = 0.0
Contraste de Hartley: 6.84435
Test de Levene: 2.64094 P-valor = 0.0713937

Análisis de la Varianza de dos Factores

Objetivo

Investigar los efectos de dos factores, α y β, en el resultado de un experimento.

β / α	1	2	...	b	Medias por filas
1	X_{11}	X_{12}	...	X_{1b}	$\overline{X_{1.}}$
2	X_{21}	X_{22}	...	X_{2b}	$\overline{X_{2.}}$
.	.	.	...	.	.
.	.	.	...	.	.
.	.	.	...	.	.
a	X_{a1}	X_{a2}	...	X_{ab}	$\overline{X_{a.}}$
Medias por columnas	$\overline{X_{.1}}$	$\overline{X_{.2}}$	...	$\overline{X_{.b}}$	$\overline{X}$

Metodología

1. Representación gráfica de los datos.
2. Planteamiento del modelo.
3. Estimación de los parámetros.
4. Contraste de si los factores influyen o no en la variable aleatoria.
5. Comprobación de las hipótesis básicas por análisis de residuos.

1. Representación gráfica de los datos.

Ej. Analizar si los factores *Pendiente* y *Orientación* influyen en la superficie total quemada en incendios forestales en la CAM.

Pendiente:

1- Terreno llano

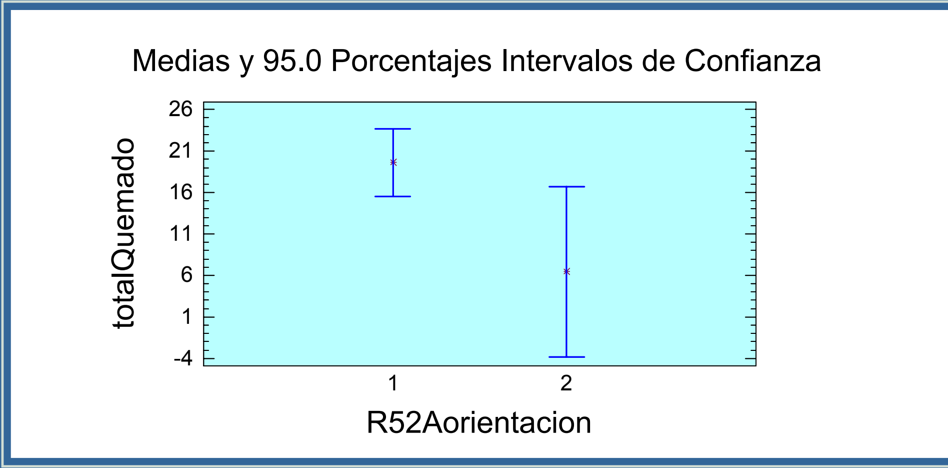
2- Ondulado

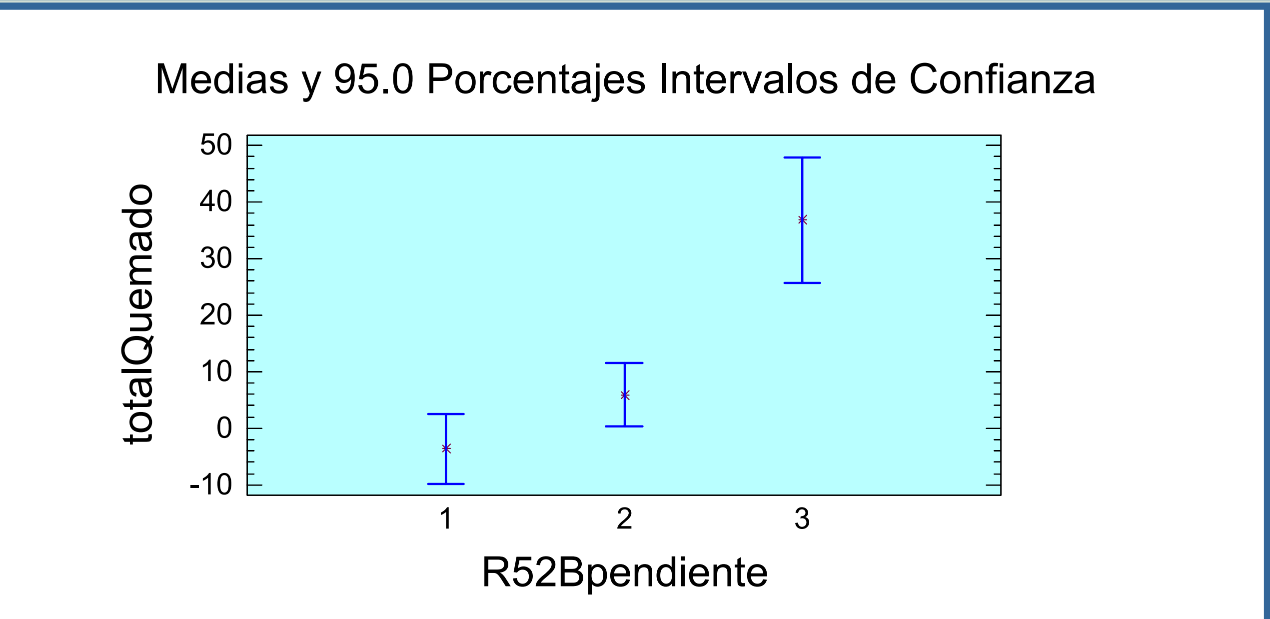
3-Abrupto

Orientación:

1- Solana

2- Umbría





2.Planteamiento del modelo

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad j = 1,2,\ldots,b;\ i = 1,2,\ldots,a$$

α_i es el efecto del i-ésimo nivel del primer factor.

β_j es el efecto del j-esimo nivel del segundo factor.

ε_{ij} representa la perturbación aleatoria, y se acepta que cumple que:

- Son independientes
- Siguen un distribución normal
- Todas tienen la misma varianza (homocedasticidad) σ^2

-$E[\varepsilon.]=0$ (media cero).

2.Planteamiento del modelo

Por tanto X_{ij} son variables aleatorias independientes que siguen una distribución normal de media $\mu + \alpha_i + \beta_j$ y varianza σ^2 .

Se asume que:

$$\sum_{i=1}^{a} \alpha_i = 0 \qquad \sum_{i=1}^{b} \beta_i = 0$$

Por tanto:

$$\beta_b = -\sum_{i=1}^{b-1} \beta_i \qquad \alpha_a = -\sum_{i=1}^{a-1} \alpha_i$$

2.Planteamiento del modelo

Otra forma de plantear el modelo es la siguiente:

$$X = \beta A' + \varepsilon$$

Donde:

$$X = (X_{11}, X_{12}, ..., X_{1b}, X_{21}, X_{22}, ..., X_{2b}, ..., X_{a1}, X_{a2}, ..., X_{ab})$$

$$\beta = (\mu, \alpha_1, \alpha_2, ..., \alpha_{a-1}, \beta_1, \beta_2, ..., \beta_{b-1})$$

$$\varepsilon = (\varepsilon_{11}, \varepsilon_{12}, ..., \varepsilon_{1b}, \varepsilon_{21}, \varepsilon_{22}, ..., \varepsilon_{2b}, ..., \varepsilon_{a1}, \varepsilon_{a2}, ..., \varepsilon_{ab})$$

3.Estimación de los parámetros

Los estimadores de los parámetros son los siguientes:

$$\hat{\mu} = \overline{X} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} X_{ij}}{ab}$$

$$\hat{\alpha}_i = \overline{X_{i.}} - \overline{X}$$

$$\overline{X_{i.}} = \frac{\sum_{j=1}^{b} X_{ij}}{b}$$

$$\hat{\beta}_j = \overline{X_{.j}} - \overline{X}$$

$$\overline{X_{.j}} = \frac{\sum_{j=1}^{a} X_{ij}}{a}$$

3.Estimación de los parámetros

Orientación Pendiente	1	2	$\overline{X}_i$
1	9 ha	2 ha	5,5 ha
2	10 ha	8 ha	9 ha
3	5 ha	1 ha	3 ha
$\overline{X}_{.j}$	8 ha	3,6 ha	$\overline{X} = 5,5\,ha$

$\alpha_1 = 0$ $\alpha_2 = 3,5$ $\alpha_3 = -2,5$ $\beta_1 = 2,5$ $\beta_2 = 1,9$

4.Contraste para analizar la influencia de los factores

Lo que se trata de comprobar es que los factores no influyen en la variable *X*. Para ello la hipótesis más sencilla es comprobar si las medias son iguales para todos los factores:

$$H_\alpha : \alpha_1 = \alpha_2 = \alpha_3 = ... = \alpha_a = 0$$

$$H_\beta : \beta_1 = \beta_2 = \beta_3 = ... = \beta_b = 0$$

4.Contraste para analizar la influencia de los factores

Tabla de Análisis de la Varianza (ANOVA):

Fuentes de variación	**Suma de cuadrados**	**Grados de libertad**	**Varianzas**	**Coefte. F**
Entre filas	$SS_1 = b\sum_{i=1}^{a}(\overline{X_{i.}} - \overline{X})^2$	**a-1**	$MS_1 = SS_1/(a-1)$	MS_1/MS
Entre columnas	$SS_2 = a\sum_{j=1}^{b}(\overline{X_{.j}} - \overline{X})^2$	**b-1**	$MS_2 = SS_2/(b-1)$	MS_2/MS
Error	$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X_{i.}} - \overline{X_{.j}} + \overline{X})^2$	**(a-1)(b-1)**	$MSE = SSE/(a-1)(b-1)$	
TOTAL	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij} - X\right)^2$	**ab-1**	$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij} - \overline{X}\right)^2 \Big/ ab - 1$	

4.Contraste para analizar la influencia de los factores

Análisis de la Varianza paratotalQuemado - Sumas de Cuadrados de Tipo III

Fuente	Suma de cuadrados	GL	Cuadrado Medio	Cociente-F	P-Valor
EFECTOS PRINCIPALES					
A:R52Aorientacion	31939.7	1	31939.7	5.81	0.0159
B:R52Bpendiente	297661.0	2	148831.0	27.07	0.0000
RESIDUOS	2.73991E7	4984	5497.41		
TOTAL (CORREGIDO)	2.77001E7	4987			

Los cocientes F están basados en el error cuadrático medio residual.

Análisis de la Varianza de Dos Factores con Interacción

Objetivo

Estudiar la influencia de dos factores en el comportamiento de una variable cuantitativa, así como la influencia conjunta de los factores en la variable explicada.

Metodología

1. Representación gráfica de los datos.
2. Planteamiento del modelo.
3. Estimación de los parámetros.
4. Contraste de si los factores influyen o no en la variable aleatoria.
5. Comprobación de las hipótesis básicas por análisis de residuos.

2.Planteamiento del modelo

$$X_{ijs} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijs}, \quad j=1,2,..b; \; i=1,2,..a; \; s=1,2,..m$$

α_i es el efecto del i-ésimo nivel del primer factor.

β_j es el efecto del j-esimo nivel del segundo factor.

γ_{ij} es el efecto de la interacción del i-esimo nivel del primer factor y del j-esimo nivel del segundo factor.

ε_{ijs} representa la perturbación aleatoria, y se acepta que cumple que:

- Son independientes

- Siguen un distribución normal

- Todas tienen la misma varianza (homocedasticidad) σ^2

-E[ε.]=0 (media cero).

2.Planteamiento del modelo

β / α	**1**	**2**	**...**	**b**	**Medias por filas**
1	X_{111} X_{112} ... X_{11m}	X_{121} X_{122} ... X_{12m}	...	X_{1b1} X_{1b2} ... X_{1bm}	$\overline{X_{1..}}$
2	X_{211} X_{212} ... X_{21m}	X_{221} X_{222} ... X_{22m}	...	X_{2b1} X_{2b2} ... X_{2bm}	$\overline{X_{2..}}$
.	.	.	...	.	.
a	X_{a1m} X_{a2m} ... X_{a1m}	X_{a2}	...	X_{ab}	$\overline{X_{a..}}$
Medias por columnas	$\overline{X_{.1.}}$	$\overline{X_{.2.}}$	...	$\overline{X_{.b.}}$	$\overline{X}$

2.Planteamiento del modelo

Por tanto X_{ij} son variables aleatorias independientes que siguen una distribución normal de media $\mu + \alpha_i + \beta_j + \gamma_{ij}$ y varianza σ^2 .

Se asume que:

$$\sum_{i=1}^{a} \alpha_i = 0 \quad \sum_{i=1}^{b} \beta_i = 0 \quad \sum_{j=1}^{b} \gamma_{ij} = 0 \ \forall j \quad \sum_{i=1}^{a} \gamma_{ij} = 0 \ \forall i$$

Por tanto:

$$\beta_b = -\sum_{i=1}^{b-1} \beta_i \qquad \alpha_a = -\sum_{i=1}^{a-1} \alpha_i$$

3.Estimación de los parámetros

Los estimadores de los parámetros son los siguientes:

$$\hat{\mu} = \overline{X} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{s=1}^{m} X_{ijs}}{abm}$$

$$\hat{\gamma}_{ij} = \overline{X}_{ij.} - \overline{X}_{i..} - \overline{X}_{.j.} + \overline{X}$$

$$\hat{\alpha}_i = \overline{X}_{i..} - \overline{X}$$

$$\overline{X}_{i..} = \frac{\sum_{j=1}^{b}\sum_{s=1}^{m} X_{ijs}}{bm}$$

$$\hat{\beta}_j = \overline{X}_{.j.} - \overline{X}$$

$$\overline{X}_{.j.} = \frac{\sum_{j=1}^{a}\sum_{s=1}^{m} X_{ijs}}{am}$$

3.Estimación de los parámetros

Los estimadores de los parámetros son los siguientes:

$$\hat{\gamma}_{ij} = \overline{X}_{ij.} - \overline{X}_{i..} - \overline{X}_{.j.} + \overline{X}$$

$$\overline{X}_{ij.} = \frac{\sum_{s=1}^{m} X_{ijs}}{m}$$

Los residuos son:

$$e_{ijs} = X_{ijs} - \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}\right) = X_{ijs} - \overline{X}_{ij.}$$

4.Contraste para analizar la influencia de los factores

Lo que se trata de comprobar es que los factores no influyen en la variable *X*.

$$H_\alpha : \alpha_1 = \alpha_2 = \alpha_3 = ... = \alpha_a = 0$$

$$H_\beta : \beta_1 = \beta_2 = \beta_3 = ... = \beta_b = 0$$

$$H_\gamma : \gamma_{11} = \gamma_{12} = \ldots = \gamma_{ab} = 0.$$

4.Contraste para analizar la influencia de los factores

Tabla de Análisis de la Varianza (ANOVA):

Fuentes de variación	**Suma de cuadrados**	**Grados de libertad**	**Varianzas**	**Cfte.F**
Factor α	$SS_1 = mb\sum_{i=1}^{a}\hat{\alpha}_i^{\,2}$	**a-1**	$MS_1 = SS_1/(a-1)$	MS_1/MSE
Factor β	$SS_2 = am\sum_{j=1}^{b}\hat{\beta}_j^{\,2}$	**b-1**	$MS_2 = SS_2/(b-1)$	MS_2/MSE
Interacción γ	$SS_3 = m\sum_{i=1}^{a}\sum_{j=1}^{b}\hat{\gamma}_{ij}^{\,2}$	**(a-1)(b-1)**	$MS_3 = SS_3/(a-1)(b-1)$	MS_3/MSE
Error	$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{s=1}^{m} e_{ijs}^2$	**ab(m-1)**	$MSE = SSE/ab(m-1)$	
TOTAL	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{s=1}^{m}\left(x_{ijs} - \bar{X}\right)^2$	**abm-1**	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{s=1}^{m}\left(x_{ijs} - \bar{X}\right)^2 \Big/ abm-1$	

4.Contraste para analizar la influencia de los factores

Tabla de Análisis de la Varianza (ANOVA):

Análisis de la Varianza paratotalQuemado - Sumas de Cuadrados de Tipo III

Fuente	Suma de cuadrados	GL	Cuadrado Medio	Cociente-F	P-Valor
EFECTOS PRINCIPALES					
A:R52Aorientacion	27327.7	1	27327.7	4.97	0.0257
B:R52Bpendiente	64876.7	2	32438.4	5.91	0.0027
INTERACCIONES					
AB	32446.4	2	16223.2	2.95	0.0523
RESIDUOS	2.73666E7	4982	5493.1		
TOTAL (CORREGIDO)	2.77001E7	4987			

Los cocientes F están basados en el error cuadrático medio residual.