Tema 8

INTRODUCCIÓN A TEORÍA GENERAL DE MODELOS LINEALES: REGRESIÓN Y ANÁLISIS DE LA VARIANZA

8.1. Descripción del modelo lineal general. Modelo de Gauss-Markov.

La teoría de los modelos lineales proporciona las bases para tratar una gran cantidad de problemas reales en los que se pretende estudiar el comportamiento de una variable aleatoria en términos de otras variables (aleatorias o no).

Estadísticamente, se trata de hacer inferencia sobre los parámetros que definen las relaciones entre las variables con el fin de precisar dichas relaciones.

Este planteamiento se resuelve adecuadamente en el contexto del llamado Modelo Lineal General (GLM), bajo especificaciones y criterios de estimación apropiados.

Definición: Un *modelo lineal* es una expresión de la forma $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, donde $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ es un vector (columna) aleatorio $n-$dimensional, observable, $\mathbf{X}$ es una matriz de orden $n \times k$, con $k < n$, a que se denota por matriz de diseño, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$ es un vector desconocido, no observable, al que se conoce como vector de efectos y $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ es un vector aleatorio, no observable, que se denomina vector de errores:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \ldots & x_{1k} \\ \vdots & \vdots & \vdots \\ x_{n1} & \ldots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Si se denota por x_j a la columna $j-$ésima de la matriz $\mathbf{X}$, se puede escribir que $\mathbf{Y} = \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$, donde se puede comprobar que cada β_j pondera el efecto (influencia) de la columna x_j en $\mathbf{Y}$, de ahí que se denomine a $\boldsymbol{\beta}$ vector de efectos.

Los modelos lineales se pueden clasificar, atendiendo a que el vector $\boldsymbol{\beta}$ sea no aleatorio o aleatorio, en modelos de efectos fijos y modelos de efectos aleatorios. En el caso de los modelos fijos, los cuales centrarán nuestro estudio, al ser $\mathbf{X}$ conocida y $\boldsymbol{\beta}$ no aleatorio, $\boldsymbol{\varepsilon}$ es una perturbación aleatoria en la que se engloban todos los factores aleatorios que intervienen en el comportamiento de $\mathbf{Y}$.

La teoría de modelos se puede generalizar al caso en que $\mathbf{X}$ sea aleatoria, y atendiendo al rango de $\mathbf{X}$, los modelos lineales se pueden clasificar en modelos de rango máximo o completo, si $rg(\mathbf{X}) = k$, y modelos lineales de rango no máximo, si $rg(\mathbf{X}) < k$.

Dentro de los modelos lineales, este tema se va a centrar en el estudio del modelo de Gauss-Markov, es decir, un modelo lineal de efectos fijos y tal que las componentes del vector $\boldsymbol{\varepsilon}$ sean variables aleatorias de segundo orden, centradas, incorreladas y homocedásticas, es decir, $\forall i = 1, \ldots, n$ se verifica: $E[\varepsilon_i] = 0$, $E[\varepsilon_i^2] = \sigma^2$ y $E[\varepsilon_i \varepsilon_j] = 0$ $i \neq j$.

Las tres condiciones impuestas al vector $\boldsymbol{\varepsilon}$ se pueden resumir en las dos siguientes: $E[\boldsymbol{\varepsilon}] = 0$ y $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 I_n$.

El objetivo del estudio de estos modelos será hacer inferencia sobre las componentes indeterminadas del modelo: el vector de efectos $\boldsymbol{\beta}$ y σ^2 a partir de una observación del vector $\mathbf{Y}$.

8.2. Estimación de un modelo de Gauss-Markov

Sea $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ un modelo lineal de Gauss-Markov, es decir verificando $E[\boldsymbol{\varepsilon}] = 0$ y $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 I_n$. Cada componente del vector $\mathbf{Y}$ verifica

$$Y_i = \sum_{j=1}^{k} x_{ij}\beta_j + \varepsilon_i,\ i = 1, \ldots, n \text{ con } E[\varepsilon_i] = 0,\ Var[\varepsilon_i] = \sigma^2 \text{ y } Cov[\varepsilon_i, \varepsilon_j] = 0,\ i \neq j.$$

Para estimar el modelo hay que estimar el vector de efectos del mismo y la varianza.

8.2.1. Estimación mínimo cuadrática del vector β

El método de mínimos cuadrados consiste en minimizar la suma de los cuadrados de los errores cometidos al aproximar $\mathbf{Y}$ por $\mathbf{X}\boldsymbol{\beta}$, esto es, se trata de minimizar Ia función

$$S^2(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2 = ||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2$$

Para ello se deriva la función $S^2(\boldsymbol{\beta})$ con respecto a cada una de las componentes del vector $\boldsymbol{\beta}$

$$\frac{\partial \sum_{i=1}^n \varepsilon_i^2}{\partial \beta_h} = -2\sum_{i=1}^n \left(Y_i - \sum_{j=1}^k x_{ij}\beta_j \right) x_{ih},\ h = 1, \ldots, k.$$

Igualando a cero estas derivadas se obtienen las denominadas ecuaciones normales:

$$\sum_{i=1}^n Y_i x_{ih} = \sum_{i=1}^n \sum_{j=1}^k x_{ij}x_{ih}\beta_j,\ h = 1, \ldots, k,$$

que matricialmente se puede expresar como $\mathbf{X}^T\mathbf{Y} = (\mathbf{X}^T\mathbf{X})\boldsymbol{\beta}$.

Cualquier solución de las ecuaciones normales que de un mínimo absoluto de $S^2(\boldsymbol{\beta})$ se denomina un estimador de mínimos cuadrados de $\boldsymbol{\beta}$ y se denota $\widehat{\boldsymbol{\beta}}(\mathbf{Y}) = \left(\widehat{\beta}_1(\mathbf{Y}), \ldots, \widehat{\beta}_k(\mathbf{Y})\right)$ o para simplificar $\widehat{\boldsymbol{\beta}}$.

Propiedades:

- Existencia: existe, al menos un estimador de mínimos cuadrados de $\boldsymbol{\beta}$.
- Unicidad: no está garantizada, salvo si el modelo es de rango máximo en cuyo caso el único estimador mínimo cuadrático viene dado por:

  $$\widehat{\boldsymbol{\beta}} := \widehat{\boldsymbol{\beta}}(\mathbf{Y}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

  el cual es una función lineal de $\mathbf{Y}$ con

  $$E[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta} \qquad Cov[\widehat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.$$

8.2.2. Estimación de funciones lineales de β. Teorema de Gauss-Markov

Definición: Una función escalar de las componentes de $\boldsymbol{\beta}$, $\psi(\boldsymbol{\beta})$, se dice que es *estimable* si admite un estimador insesgado función lineal de las componentes de $\mathbf{Y}$:

$$\psi(\boldsymbol{\beta}) \text{ estimable} \Leftrightarrow \exists \mathbf{c} \in \mathbb{R}^n \ / \ E[\mathbf{c}^T\mathbf{Y}] = \psi(\boldsymbol{\beta}),\ \forall \boldsymbol{\beta}.$$

Caracterización de funciones estimables:

- $\psi(\boldsymbol{\beta})$ es estimable $\Leftrightarrow \psi(\boldsymbol{\beta}) = \mathbf{c}^T\mathbf{X}\boldsymbol{\beta}$, $\mathbf{c} \in \mathbb{R}^n$.
- Si el modelo es de rango máximo, $\psi(\boldsymbol{\beta})$ es estimable $\Leftrightarrow \psi(\boldsymbol{\beta}) = \mathbf{a}^T\boldsymbol{\beta}$, $\mathbf{a} \in \mathbb{R}^k$.

En particular, cualquier componente del vector $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$ es estimable ya que

$$Y_i = (\mathbf{X}\boldsymbol{\beta})_i = (0, \ldots, 1, \ldots, 0)\mathbf{X}\boldsymbol{\beta}.$$

Teorema de Gauss-Markov: Toda función estimable, $\mathbf{a}^T\boldsymbol{\beta}$, admite un único estimador insesgado uniformemente de mínima varianza en la clase de estimadores lineales insesgados. Dicho estimador es $\mathbf{a}^T\widehat{\boldsymbol{\beta}}(\mathbf{Y})$, y se denomina estimador de mínimos cuadrados de $\mathbf{a}^T\boldsymbol{\beta}$.

Ya que cada componente de $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$ es estimable, por el Teorema de Gauss-Markov, existe su estimador de mínimos cuadrados y es la correspondiente componente del vector $\mathbf{X}\widehat{\boldsymbol{\beta}}$.

$$Y_i = (\mathbf{X}\boldsymbol{\beta})_i = (0,\ldots,1,\ldots,0)\mathbf{X}\boldsymbol{\beta} \Rightarrow \widehat{Y}_i = \widehat{(\mathbf{X}\boldsymbol{\beta})_i} = (0,\ldots,1,\ldots,0)\mathbf{X}\widehat{\boldsymbol{\beta}} = (\mathbf{X}\widehat{\boldsymbol{\beta}})_i.$$

Puesto que estos estimadores son insesgados se tiene $E[\widehat{\mathbf{Y}}] = E[\mathbf{X}\widehat{\boldsymbol{\beta}}] = \mathbf{X}\boldsymbol{\beta}$.

8.2.3. Modelo estimado, residuos y estimación de la varianza

Según acabamos de ver, si denotamos por $\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_1,\ldots,\widehat{\beta}_k\right)^T$ a $\widehat{\boldsymbol{\beta}}(\mathbf{Y})$, el estimador de mínimos cuadrados de $\boldsymbol{\beta}$, la estimación del modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$ viene dada por

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \Rightarrow \widehat{Y}_i = \sum_{j=1}^{k} x_{ij}\widehat{\beta}_j,\ i = 1,\ldots,n.$$

Además, $\widehat{Y}_i$, $i = 1,\ldots,n$, es el estimador lineal insesgado de mínima varianza de $E[Y_i] = \sum_{j=1}^{k} x_{ij}\beta_j$ al ser el estimador de mínimos cuadrados y ser insesgado en dicha función.

Definición: Se define el *vector de residuos mínimo cuadráticos* como el que resulta de hacer la diferencia entre el verdadero valor del modelo y el aproximado según el modelo estimado:

$$\mathbf{R} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$$

cuyas componentes son $\mathbf{R}_i = \mathbf{Y}_i - \widehat{\mathbf{Y}_i}$, $i = 1,\ldots,n$.

Propiedades:

- Los residuos son variables aleatorias con media nula, $E[R_i] = 0$, $\forall i = 1,\ldots,n$, ya que
$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta} \text{ y } E[\widehat{\mathbf{Y}}] = \mathbf{X}\boldsymbol{\beta}.$$
- El vector de residuos es ortogonal a los vectores columna de $\mathbf{X}$, $\mathbf{X}^T\mathbf{R} = 0$, ya que al ser $\widehat{\boldsymbol{\beta}}$ un estimador de mínimos cuadrados, verifica el sistema de ecuaciones normales $\mathbf{X}^T\mathbf{Y} = (\mathbf{X}^T\mathbf{X})\boldsymbol{\beta}$.
- El vector de residuos es ortogonal al vector estimado, $\widehat{\mathbf{Y}}^T\mathbf{R} = 0$. Esto se deduce de la propiedad anterior y de que $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$.

Definición: Se define la *varianza residual* como el cociente entre la suma de cuadrados de los residuos y el número de residuos linealmente independientes.

$$S_R^2 = \frac{\sum_{i=1}^n R_i^2}{n-r} = \frac{\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2}{n-r} = \frac{(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})^T(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})}{n-r} = \frac{||\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}||^2}{n-r} = \frac{||\mathbf{R}||^2}{n-r}$$

El número de residuos linealmente independientes viene determinado por el rango de la matriz $\mathbf{X}$, $r = Rango(\mathbf{X})$, ya que al verificarse que $\mathbf{X}^T\mathbf{R} = 0$, las k columnas de $\mathbf{X}$ determinan k relaciones lineales entre los residuos igualadas a cero. Por lo tanto, el rango de $\mathbf{X}$ indica cuantas de dichas relaciones no son proporcionales y, por tanto, cuantas son linealmente dependientes. De ahí que el número de residuos linealmente independientes es $n - r$.

Además, puede probarse que la varianza residual, S_r^2, es un estimador insesgado de σ^2.

8.3. Inferencia bajo hipótesis de normalidad

8.3.1. Distribución normal N-dimensional

Sean

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}$$

Se dice que el vector aleatorio $X = (X_1, \dots, X_n)^T$ tiene una distribución normal n-dimensional de media μ y matriz de varianzas-covarianzas Σ, definida positiva, y se notará $\mathcal{N}_n(\mu, \Sigma)$, si y sólo si su función de densidad es

$$f(x) = \frac{1}{(2\pi)^{n/2}[\det(\Sigma)]^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}, \quad x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$$

Propiedades

Si $X \rightsquigarrow \mathcal{N}_n(\mu, \Sigma)$, $X = (X_1, \dots, X_n)^T$

1. $E[X] = \mu$ y $Cov[X] = \Sigma$.
2. Su función generatriz de momentos será: $M(t) = e^{t^T\mu + \frac{t^T\Sigma t}{2}}$.
3. Las distribuciones marginales de cualquier dimensión son normales y, en particular, $X_i \rightsquigarrow \mathcal{N}(\mu_i, \sigma_{ii})$.

4. Si $\gamma_{n\times 1}$ es un vector constante $\Rightarrow X+\gamma \sim N_n(\mu+\gamma,\Sigma)$.

5. $X_1,\ldots,X_n$ son independientes $\Leftrightarrow \Sigma$ es diagonal, es decir, $\sigma_{ij}=0$, $\forall i\neq j$,

8.3.2. Inferencia

Si en un modelo de Gauss-Markov añadimos la condición de que los residuos sean normales, además de centrados, independientes y homocedásticos, es decir si consideramos el modelo:

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \text{ con } \boldsymbol{\varepsilon} \rightsquigarrow \mathcal{N}_n(0,\sigma^2\, I_{n\times n})$$

estamos considerando de forma equivalente que cada componente del vector $\mathbf{Y}$ sigue una distribución normal unidimensional

$$Y_i=\sum_{j=1}^{k} x_{ij}\beta_j+\varepsilon_i \rightsquigarrow \mathcal{N}\left(\sum_{j=1}^{k} x_{ij}\beta_j,\sigma^2\right)$$

y que las componentes del vector $\mathbf{Y}=(Y_1,\ldots,Y_n)^T$ son independientes. Por lo tanto

$$\mathbf{Y} \rightsquigarrow \mathcal{N}_n\left(\mathbf{X}\beta,\sigma^2 I_{n\times n}\right)$$

Estimadores máximo verosímiles

Para determinar los estimadores máximo verosímiles del modelo, es decir del vector de efectos, $\boldsymbol{\beta}$, y del parámetro σ^2, hay que determinar primero la función de verosimilitud:

$$\mathbf{y}\in\mathbb{R}^n \rightarrow L_{\mathbf{y}}(\boldsymbol{\beta},\sigma^2)=\frac{1}{(2\pi)^{n/2}\sigma^n}\,\exp\left\{-\frac{||\mathbf{y}-\mathbf{X}\boldsymbol{\beta}||^2}{2\sigma^2}\right\}.$$

- Puesto que maximizar $L_{\mathbf{y}}(\boldsymbol{\beta},\sigma^2)$ en $\boldsymbol{\beta}$ es equivalente a mín $||\mathbf{y}-\mathbf{X}\boldsymbol{\beta}||^2$ en ese mismo parámetro, el estimador máximo verosímil de $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, será el estimador mínimo cuadrático.

- Para obtener el estimador máximo verosímil de σ^2 hay que, tras tomar logaritmo de la verosimilitud, resolver:

$$\frac{\partial \ln L_{\mathbf{y}}(\boldsymbol{\beta},\sigma^2)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\,||\mathbf{y}-\mathbf{X}\boldsymbol{\beta}||^2=0,$$

  de donde se deduce que:

$$\widehat{\sigma}^2(\mathbf{y})=\frac{||\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}||^2}{n}=\frac{\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2}{n}=\frac{\sum_{i=1}^{n}R_i^2}{n}=\frac{(n-r)S_R^2}{n}.$$

Test de razón de verosimilitud para la hipótesis lineal general

Definición: Se denomina *hipótesis lineal general* a la hipótesis formulada por $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$ siendo $\mathbf{C}_{q\times k}$ una matriz conocida de rango $q(\leq k)$.

$$\left\{\begin{array}{cccc} c_{11}\beta_1 & +\cdots+ & c_{1k}\beta_k = & 0 \\ \vdots & \vdots & \vdots & \vdots \\ c_{q1}\beta_1 & +\cdots+ & c_{qk}\beta_k = & 0 \end{array}\right.$$

Si todas las componentes del vector $\mathbf{C}\boldsymbol{\beta}$ son estimables, entonces el test de razón de verosimilitud, de tamaño α, que resuelve el contraste

$$\left\{\begin{array}{l} H_0 : \mathbf{C}\boldsymbol{\beta} = 0 \\ H_1 : \mathbf{C}\boldsymbol{\beta} \neq 0 \end{array}\right.$$

viene dado por

$$\varphi(\mathbf{Y}) = \left\{\begin{array}{ll} 1 & F(\mathbf{Y}) > F_{q,n-r;\,\alpha} \\ 0 & F(\mathbf{Y}) \leq F_{q,n-r;\,\alpha} \end{array}\right.$$

donde

$$F(\mathbf{Y}) = \frac{n-r}{q}\left(\frac{||\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}^0||^2 - ||\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}||^2}{||\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}||^2}\right)$$

siendo $\widehat{\boldsymbol{\beta}}^0$ el estimador máximo verosímil de $\boldsymbol{\beta}$ bajo H_0.

8.4. Modelo de regresión lineal simple

El problema de regresión consiste en determinar una función que aplicada a una serie de variables aleatorias $X_1, \ldots, X_k$ permita predecir, con la mayor precisión posible, los valores de otra variable Y según los valores de $X_1, \ldots, X_k$: $Y = \varphi(X_1, \ldots, X_k) + E$ (E debe ser una variable aleatoria que expresa el error cometido al predecir Y por $\varphi(X_1, \ldots, X_k)$).

Las variables $X_1, \ldots, X_k$ son variables explicativas, independientes o regresoras que fijamos al principio. La variable Y es la variable explicada, dependiente o de respuesta. Además, como se trata de un modelo univariante, se considera que la variable Y es unidimensional. Finalmente, al considerarse un modelo de regresión lineal, se asume que φ es lineal, es decir, $X_1, \ldots, X_k$ influyen en la respuesta de forma lineal. El resto de los factores que pueden influir en Y se engloban en E.

Al tratarse de un modelo simple, el problema se centra en el caso $k = 1$, es decir hay una única variable explicativa X y una variable explicada Y, ambas unidimensionales, y se asume el modelo de regresión lineal simple:

$$Y = \beta_0 + \beta_1 X + E$$

y se resolverá el problema de regresión desde el punto de vista empírico, basándose en observaciones de la variable, $(x_1, \ldots, x_n)$. El objetivo va a ser estimar el modelo a partir de las observaciones de la variable Y correspondiente a cada una de las x_i.

8.4.1. Planteamiento del modelo

Sean X e Y dos variables aleatorias tales que $E[Y^2] < +\infty$ y, fijado un valor arbitrario, $X = x$ se cumple:

$$E[Y/X = x] = \beta_0 + \beta_1 x, \quad Var[Y/X = x] = \sigma^2.$$

Sean $x_1, \ldots, x_n$ un conjunto de valores de X (se exigen al menos dos valores distintos). Supongamos que para cada $i = 1, \ldots, n$, Y_i es la observación aleatoria de la variable Y supuesto que $X = x_i$. Entonces $\forall i = 1, \ldots, n$

$$E[Y_i] = E[Y/X = x_i] = \beta_0 + \beta_1 x_i, \quad Var[Y_i] = Var[Y/X = x_i] = \sigma^2$$

Por tanto, $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ con $E[\varepsilon_i] = 0$ y $Var[\varepsilon_i] = \sigma^2$, $i = 1, \ldots, n$. Es decir, se tiene

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

que es un modelo lineal de rango máximo, al exigir que $\exists x_i \neq x_j$. Si las observaciones se toman de forma independiente, entonces las variables $\varepsilon_1, \ldots, \varepsilon_n$ son independientes, y centradas, por lo que el modelo es de Gauss-Markov de rango máximo.

8.4.2. Estimación de los parámetros del modelo

Como el modelo es de Gauss-Markov de rango máximo, sabemos que es estimador mínimo cuadrático de $\boldsymbol{\beta}$ es $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, de donde se puede ver que la estimación de los parámetros del modelo, β_0 y β_1, es:

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{x} \qquad \widehat{\beta}_1 = \frac{\sigma_{xY}}{\sigma_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Propiedades:

1. Los estimadores mínimos cuadráticos son único (por ser el modelo de rango máximo) y son funciones lineales de las componentes de $\mathbf{Y}$.
2. Los estimadores mínimo cuadráticos son insesgados: $E[\widehat{\beta}_0] = \beta_0$ y $E[\widehat{\beta}_1] = \beta_1$.
3. $Var[\widehat{\beta}_0] = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{n\sigma_x^2} \right)$ y $Var[\widehat{\beta}_1] = \sigma^2 \frac{1}{n\sigma_x^2}$.
4. $Cov[\widehat{\beta}_0, \widehat{\beta}_1] = -\sigma^2 \frac{\bar{x}}{n\sigma_x^2}$.
5. La estimación del modelo es $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$, por lo tanto, la estimación de cada componente $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ es
$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i = \bar{Y} + \frac{\sigma_{xY}}{\sigma_x^2}(x_i - \bar{x}),\ i = 1, \ldots, n.$$
6. El vector de residuos, $\mathbf{R} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$, en este caso está compuesto por
$$R_i = Y_i - \widehat{Y}_i = Y_i - \bar{Y} - \frac{\sigma_{xY}}{\sigma_x^2}(x_i - \bar{x}),\ i = 1, \ldots, n.$$
Además, sabemos $\mathbf{X}^T\mathbf{R} = 0$ y $\widehat{\mathbf{Y}}^T\mathbf{R} = 0$, por lo tanto:
   - $\mathbf{X}^T\mathbf{R} = 0 \Rightarrow \sum_{i=1}^n R_i = 0 \Rightarrow \sum_{i=1}^n R_i/n = 0 \Rightarrow \sum_{i=1}^n \widehat{Y}_i/n = \bar{Y}$
   - $\widehat{\mathbf{Y}}^T\mathbf{R} = \sum_{i=1}^n \widehat{Y}_i R_i = 0$
7. El estimador de σ^2 es la varianza residual:
$$\widehat{\sigma}^2 = S_R^2 = \frac{\sum_{i=1}^n R_i^2}{n-2} = \frac{\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2}{n-2} = \frac{\sum_{i=1}^n (Y_i - \bar{Y} - \frac{\sigma_{xY}}{\sigma_x^2}(x_i - \bar{x}))^2}{n-2}.$$

8.4.3. Análisis de la bondad

La bondad del modelo ajustado depende en magnitud de la componente residual. Ésta es una variable aleatoria, por tanto, una forma de medirla es a través de la varianza residual S_R^2, que no debe olvidarse que es la estimación de σ^2.

El inconveniente que presenta esta medida de la bondad es que tiene dimensión y, por tanto, no es válida para hacer comparaciones en general. Esto motiva la búsqueda de otra medida de la bondad que sea adimensional. Para ello se estudia la descomposición de la variabilidad.

Descomposición de la variabilidad

La variabilidad total (VT) de las observaciones aleatorias $Y_1, \ldots, Y_n$ viene determinada por $\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = n\sigma_Y^2$. Esta VT se puede descomponer en dos términos:

- La variabilidad explicada (VE), que es la medida de la variabilidad de las estimaciones $\widehat{Y}_1, \ldots, \widehat{Y}_n$: $\sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})^2 = \dfrac{n\sigma_{xY}^2}{\sigma_x^2}$.
- La variabilidad no explicada (VNE), que es la variabilidad de la parte residual: $\sum_{i=1}^{n} R_i^2 = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 = (n-2)S_R^2$.

La descomposición de la variabilidad, por tanto, viene dada por: VT=VE+VNE.

Coeficiente de determinación

El coeficiente de determinación se define como la proporción de la variabilidad total explicada por el modelo de regresión:

$$R^2 = \frac{\text{VE}}{\text{VT}} = \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = \frac{n\sigma_{xY}^2}{\sigma_x^2 n\sigma_Y^2} = \frac{\sigma_{xY}^2}{\sigma_x^2\sigma_Y^2}$$

Propiedades:

1. R^2 es adimensional, por lo tanto es una medida de la bondad adecuada para hacer comparaciones.
2. $0 \leq R^2 \leq 1$.
3. Cuanto mayor sea R^2, menores serán los residuos y, por tanto, mejor será el modelo.

8.4.4. Predicción con el modelo estimado

Utilizando el modelo estimado puede obtenerse una predicción para Y basada en el valor que toma X. Para ello, teniendo en cuenta que si $X = x_p$ entonces

$$Y_p = \beta_0 + \beta_1 x_p + \varepsilon_p;\ E[\varepsilon_p] = 0,\ Var[\varepsilon_p] = \sigma^2$$

se tiene que la predicción de Y para $X = x_p$ según el modelo estimado viene dada por:

$$\widehat{Y}_p = \widehat{\beta}_0 + \widehat{\beta}_1 x_p = \bar{Y} + \frac{\sigma_{xY}}{\sigma_x^2}(x_p - \bar{x}).$$

Propiedades:

- $E[\widehat{Y}_p] = E[Y_p]$.
- $Var[\widehat{Y}_p] = \sigma^2 \left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2}\right]$

Para medir la bondad de las predicciones se va a utilizar el error cuadrático medio.

$$ECM = E[\widehat{Y}_p - Y_p]^2 = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2}\right]$$

A saber, el ECM será menor cuanto más se aproxime x_p a $\bar{x}$.

8.4.5. Aplicación práctica: recta de regresión estimada

La recta de regresión estimada a partir de $(x_1, y_1), \ldots, (x_n, y_n)$, observaciones independientes del vector aleatorio (X, Y), se obtiene a partir del modelo de regresión teórico, obtenido anteriormente, con los valores $x_1, \ldots, x_n$:

$$\widehat{Y}_i = \bar{Y} + \frac{\sigma_{xY}}{\sigma_x^2}(x_i - \bar{x})$$

La estimación concreta del modelo, para $Y_1 = y_1, \ldots, Y_n = y_n$, viene dada por

$$\widehat{y}_i = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x_i - \bar{x}),\ i = 1 \ldots, n.$$

Por tanto, la recta de regresión estimada a partir de las observaciones (x_i, y_i) será

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x}).$$

La recta de regresión teórica sería

$$y = E[Y] + \frac{Cov[X, Y]}{Var[X]}(x - E[X]).$$

Coeficiente de determinación

La estimación del coeficiente de determinación lineal R^2 es

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2}$$

y su correspondiente poblacional viene dado por

$$\rho_{XY}^2 = \frac{Cov^2[X, Y]}{Var[X]Var[Y]}.$$

Predicción

La recta de regresión estimada permite predecir valores de la variable Y a partir de la variable X con valores que no han intervenido en la predicción. Para ello basta con considerar un valor arbitrario x_p de X, y la recta de regresión estimada a partir de las observaciones $(x_1, y_1), \ldots, (x_n, y_n)$ proporcionará una predicción de Y:

$$\widehat{y}_p = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x_p - \bar{x}).$$

Puesto que para medir la bondad de las predicciones se usa el ECM que depende de σ^2, si lo desconocemos se puede estimar el ECM estimando σ^2 por la varianza residual s_R^2:

$$\widehat{ECM}(\widehat{Y}_p) = s_R^2\left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2}\right].$$

8.4.6. Contraste de regresión bajo hipótesis de normalidad

El contraste de regresión trata de resolver el problema de ver si la variable X ejerce o no influencia lineal sobre la variable Y, es decir, se quiere contrastar que la variable X no tiene inferencia lineal sobre la variable Y. Para ello se plantea el contraste de hipótesis nula

$$H_0 : \beta_1 = 0$$

en el modelo $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, \ldots, n$, modelo lineal con hipótesis de normalidad ($\varepsilon_i \rightsquigarrow \mathcal{N}(0, \sigma^2)$) de rango máximo.

Para resolver este contraste se aplica el test de la razón de verosimilitud (TRV) para la hipótesis lineal general, ya que el modelo de regresión lineal es un MLG de rango máximo.

En este caso se tiene $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$ con $\mathbf{C} = (0, 1)$ y $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$. Luego $q = rang(\mathbf{C}) = 1$, $r = rang(\mathbf{X}) = 2$ y, según se estudió anteriormente, el TRV sería:

$$\varphi(\mathbf{Y}) = \begin{cases} 1 & F(\mathbf{Y}) > F_{q,n-r;\alpha} = F_{1,n-2;\alpha} \\ 0 & F(\mathbf{Y}) \leq F_{q,n-r;\alpha} = F_{1,n-2;\alpha} \end{cases}$$

con

$$F(\mathbf{Y}) = \frac{n-r}{q}\left(\frac{||\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^0||^2 - ||\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}||^2}{||\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}||^2}\right) = (n-2)\frac{VT - VNE}{VNE} = \frac{VE}{VNE/(n-2)} = \frac{VE}{S_R^2}$$

Puesto que $R^2 = VE/VT$, $F(\mathbf{Y}) = (n-2)\dfrac{R^2}{1-R^2}$, creciente en R^2. Por lo tanto, grandes valores de R^2 conducen al rechazo de H_0, lo que concuerda con la definición de R^2 como medida de la bondad del modelo estimado.

8.5. Análisis de la varianza de una vía

Si se desea medir el efecto de un supuesto factor de variación sobre el comportamiento de una variable aleatoria se consideran muestras de la variable aleatoria bajo distintos niveles del factor de variación y se descompone la variabilidad total de los datos en dos sumandos: uno que exprese la variabilidad entre las muestras y otro que exprese la variabilidad de los datos por ser observaciones de una variable aleatoria.

Ejemplo: Si se desea estudiar como afecta el tipo de alimentación en el peso de ciertos individuos, el factor de variabilidad sería el tipo de alimentación y la variable aleatoria el peso. Para dicho estudio se mediría el peso de los individuos, tras dividirlos en k grupos, y dar a cada grupo un tipo de alimentación diferente. La variabilidad que se apreciaría en los datos obtenidos puede venir dada por ser observaciones de una variable aleatoria o por ser individuos de distintos grupos.

El análisis de la varianza de una vía estudia esta variabilidad. Para ello se asumen dos supuestos, dados por Fisher:

i) La variable de interés no está afectada por factores distintos del que es objeto de estudio.

ii) En cada uno de los k niveles del factor de variación, la variable aleatoria se distribuye normalmente y con varianza común (homocedasticidad).

El problema de analizar la variabilidad se puede estudiar mediante el contraste del efecto de variación en los k grupos lo que se puede reducir a comparar las medias de k poblaciones normales con varianza común. Para ello se considera $(Y_{i1}, \ldots, Y_{in_i})$ m.a.s. de $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \ldots, k$, donde cada Y_i es la variable aleatoria que esta influida por el $i-$ésimo factor de variación. Las muestras de cada variable Y_i deben ser independientes entre si.

Cada una de las variables de la muestra se puede descomponer en dos sumandos $Y_{ij} = \mu_i + \varepsilon_{ij}$, siendo ε_{ij} una variable aleatoria con distribución $\mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, k$; $j = 1, \ldots, n_i$; $n = \sum_{i=1}^{k} n_i$. De esta forma, considerando el vector $\mathbf{Y}$, la matriz $\mathbf{X}$ de rango k, el vector $\boldsymbol{\mu}$ y el vector $\boldsymbol{\varepsilon}$ siguientes, el problema de análisis de la varianza de una vía se puede ver como un modelo lineal general. Es más, el modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$ es un modelo de Gauss-Markov de rango máximo, k.

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \hline Y_{21} \\ \vdots \\ Y_{2n_2} \\ \hline \vdots \\ \hline Y_{k1} \\ \vdots \\ Y_{kn_k} \end{pmatrix}_{n\times 1} \qquad \mathbf{X} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ \hline & & \vdots & \\ \hline 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{n\times k} \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}_{k\times 1} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \hline \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \hline \vdots \\ \hline \varepsilon_{k1} \\ \vdots \\ \varepsilon_{kn_k} \end{pmatrix}_{n\times 1}$$

8.5.1. Problema de contraste

Como se ha indicado, el problema de análisis de una vía se puede reducir a un contraste de igualdad de medias del tipo: $H_0 : \mu_1 = \ldots = \mu_k$, el cual equivale a contrastar $H_0 : \mu_1 - \mu_2 = 0, \ldots, \mu_1 - \mu_k = 0$ (donde se establecen $k-1$ relaciones lineales entre las componentes de μ). Es decir, el contraste que se quiere estudiar es una hipótesis lineal general del tipo:

$$H_0 : \mathbf{C}\boldsymbol{\mu} = 0, \quad \mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ \vdots & 0 & -1 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & -1 \end{pmatrix}_{(k-1)\times k} \qquad q = rang(\mathbf{C}) = k-1.$$

El test de razón de verosimilitud que resuelve este contraste es, de nuevo, el TRV para el MLG, es decir:

$$\varphi(\mathbf{Y}) = \begin{cases} 1 & F(\mathbf{Y}) > F_{q,n-r;\alpha} = F_{k-1,n-k;\alpha} \\ 0 & F(\mathbf{Y}) \leq F_{q,n-r;\alpha} = F_{k-1,n-k;\alpha} \end{cases}$$

con

$$F(\mathbf{Y}) = \frac{n-k}{k-1}\left(\frac{||\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\mu}}^0||^2 - ||\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\mu}}||^2}{||\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\mu}}||^2}\right)$$

donde $\widehat{\boldsymbol{\mu}}$ es el EMV de $\boldsymbol{\mu}$ y $\widehat{\boldsymbol{\mu}}^0$ es el EMV de $\boldsymbol{\mu}$ bajo H_0.

Dichos estimadores son fácilmente deducibles teniendo en cuenta que se esta trabajando con distribuciones normales.

Estimador máximo verosímil (mínimos cuadrados) de $\boldsymbol{\mu}$

El parámetro $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^T$ el la media de la distribución del vector $\mathbf{Y} \rightsquigarrow \mathcal{N}_k(\boldsymbol{\mu}, \sigma^2\mathbf{I}_k)$, es decir, es el parámetro media de una normal multivariante. Pero, además, $\mathbf{Y}$ es un vector formado por k variables aleatorias independientes con distribuciones $\mathcal{N}(\mu_i, \sigma^2)$, donde cada componente del vector $\boldsymbol{\mu}$ es la media de una variable aleatoria normal unidimensional.

Según se ha estudiado anteriormente, el estimador máximo verosímil (EMV) del parámetro media de una distribución normal es la media muestral y, por tanto, como en este caso se tiene $Y_{i1}, \ldots, Y_{in_i}$ m.a.s. de Y_i:

$$\widehat{\mu}_i = \bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}.$$

Luego el EMV de $\boldsymbol{\mu}$ es: $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_k)^T = (\bar{Y}_1, \ldots, \bar{Y}_k)^T$.

Estimador máximo verosímil (mínimos cuadrados) de $\boldsymbol{\mu}$ bajo H_0

Bajo $H_0 : \mu_1 = \ldots = \mu_k$, el parámetro $\boldsymbol{\mu}$ se puede ver como un vector formado por k componentes iguales, a las que también se las puede llamar μ, es decir, $\boldsymbol{\mu} = (\mu, \ldots, \mu)^T$. Por lo tanto, se tiene que μ es un parámetro unidimensional que determina la media de la distribución de $Y \rightsquigarrow \mathcal{N}_k(\mu 1_k, \sigma^2 I_k)$. Pero, en este caso, Y es un vector formado por k variables aleatorias independientes e idénticamente distribuidas $\mathcal{N}(\mu, \sigma^2)$, por lo tanto μ es la media de una variable aleatoria normal unidimensional.

Al igual que antes, se sabe que el EMV del parámetro media de una distribución normal es la media muestral y, por tanto, como en este caso se tiene $Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{k1}, \ldots, Y_{kn_k}$ m.a.s. de $\mathcal{N}(\mu, \sigma^2)$:

$$\widehat{\mu} = \bar{Y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}}{n} \quad \text{donde } n = \sum_{i=1}^{k} n_i$$

Luego el EMV de $\boldsymbol{\mu}$ bajo H_0 es: $\widehat{\boldsymbol{\mu}}^0 = (\widehat{\mu}, \ldots, \widehat{\mu})^T = (\bar{Y}, \ldots, \bar{Y})^T$.

El estadístico del contraste se puede reescribir, sustituyendo los EMV obtenidos y quedaría:

$$F(\mathbf{Y}) = \frac{n-k}{k-1}\left(\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y})^2 - \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2}\right) (\rightsquigarrow F(k-1, n-k) \text{ bajo } H_0).$$

8.5.2. Descomposición de la variabilidad de las observaciones

Para estudiar la descomposición de la variabilidad, primero se debe recordar que estamos considerando el modelo lineal $\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$, cuyo modelo estimado es $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\mu}}$, es decir $\widehat{Y}_{ij} = \widehat{\mu}_i = \bar{Y}_i$. Luego, el error que se comete al estimar Y_{ij} por $\widehat{Y}_{ij}$ viene dado por $R_{ij} = Y_{ij} - \bar{Y}_i$.

Por otro lado, anteriormente se ha estudiado que la variabilidad total de una serie de observaciones aleatorias, $Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{k1}, \ldots, Y_{kn_k}$, viene determinada por:

$$VT = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2$$

Dichas observaciones aleatorias son las m.a.s., independientes, de las variables $Y_i \rightsquigarrow \mathcal{N}(\mu_i, \sigma^2)$.

Al igual que anteriormente, la VT se puede descomponer como suma de dos términos,

$$VT = VE + VNE$$

- La variabilidad explicada (VE) por el modelo, que mide las variabilidades entre los grupos (también se llama variabilidad entre grupos):

$$VE = \sum_{i=1}^{k} n_i (\bar{Y}_i - \bar{Y})^2.$$

- La variabilidad no explicada (VNE) por el modelo, que es la suma de las variabilidades dentro de cada grupo:

$$VNE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} R_{ij}^2.$$

Teniendo en cuenta estos nuevos términos, el estadístico del contraste se puede reescribir como:

$$F(\mathbf{Y}) = \frac{n-k}{k-1} \left(\frac{VT - VNE}{VNE} \right) = \frac{n-k}{k-1} \frac{VE}{VNE} = \frac{VE/(k-1)}{VNE/(n-k)} = \frac{S_E^2}{S_R^2}$$

donde

$$S_E^2 = \sum_{i=1}^{k} n_i (\bar{Y}_i - \bar{Y})^2/(k-1)$$

denota la varianza entre los grupos y

$$S_R^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2/(n-k)$$

la varianza residual.

Por tanto, el test de la razón de verosimitud quedaría:

$$\varphi(\mathbf{Y}) = \begin{cases} 1 & \text{si } \dfrac{S_E^2}{S_R^2} > F_{k-1,n-k;\alpha} \\ \\ 0 & \text{si } \dfrac{S_E^2}{S_R^2} \leq F_{k-1,n-k;\alpha} \end{cases}$$

Nota: En la práctica, si el problema planteado no especifica un nivel de significación, se trabaja con el denominado $p - nivel$ o $p - valor$ asociado a dos datos:

$$p = P(F(k-1, n-k) > F_{exp})$$

siendo F_{exp} el valor del estadístico $F(\mathbf{Y})$ obtenido de los datos concretos:

- Si p es grande (0.15 o mayor), no se rechaza H_0.
- Si p es pequeño (se suele considerar $\alpha = 0{,}05$ o menor), se rechaza H_0.
- Para valores intermedios hay que tratar cada situación en particular aunque, normalmente, es aconsejable tomar más datos y rehacer los cálculos.