MODELO LINEAL

NOTAS DE CLASE

1. MODELO LINEAL GENERAL

1.1 Definición y Supuestos del Modelo
Se supone que existe una relación lineal entre una variable dependiente **Y** y **k-1** variables independientes $\mathbf{X_2, X_3, \ldots X_k}$. Existe además un término de perturbación aleatorio **u** conocida como error. Si se dispone de una muestra de tamaño **n** de observaciones en **Y** y en $\mathbf{X_2, X_3, \ldots X_k}$ se cumplen n ecuaciones de la forma:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \ldots\ldots \beta_k X_{ki} + u_i$$
$$i = 1,2,\ldots\ldots n$$

Los coeficientes $\boldsymbol{\beta_i}$ y los parámetros de la distribución de los errores $\mathbf{u_i}$ son desconocidos, se tiene entonces como problema inicial, disponer de estimaciones de tales parámetros. En forma matricial, el modelo es más manejable y se puede expresar como sigue:

$$\boxed{Y = X\beta + u}$$

Donde

$$\mathbf{Y}= \begin{bmatrix} Y_1 \\ Y_2 \\ \\ . \\ Y_n \end{bmatrix} \quad \mathbf{X}= \begin{bmatrix} 1 & X_{12} & X_{13} & X_{1k} \\ 1 & X_{22} & X_{23} & X_{2k} \\ \\ 1 & X_{n2} & X_{n3} & X_{nk} \end{bmatrix} \quad \boldsymbol{\beta}= \begin{bmatrix} \beta_1 \\ \beta_2 \\ \\ . \\ \beta_k \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \\ . \\ u_n \end{bmatrix}$$

Los **unos** en la primera columna de la matriz X se incluyen para dar lugar al intercepto $\boldsymbol{\beta_1}$. Equivale a considerar una $X_{lj} = 1$ para toda j. Observe que los subíndices de X rompen con la costumbre de utilizar el primer subíndice para el renglón y el segundo para la columna.

Se incluyen una serie de supuestos básicos que son importantes para cumplir con el proceso de estimación e inferencia que se deriven del modelo.

$$E(u) = 0$$
$$E(uu') = \sigma^2 I_n$$
$$X \text{ es una matriz de números sin propiedades probabilísticas}$$
$$\text{X es una matriz de rango k} < \text{n}$$

Como el vector de errores **u** es de orden nx1, su vector transpuesto **u'** es de orden 1xn y el producto **uu'** es una matriz cuadrada simétrica de orden nxn y su valor esperado corresponde a una matriz diagonal con valor igual a σ^2 constante en los elementos de la diagonal. Ello significa que los errores tienen varianza constante, propiedad que se conoce como homoscedasticidad. Los ceros fuera de la diagonal determinan que los errores no son correlacionados. Estos errores son variables no observables.

$$\mathbf{E(uu')}= \begin{bmatrix} E(u_1^2) & E(u_1u_2) & .. & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & .. & E(u_1u_2) \\ .. & .. & .. & \\ E(u_nu_1) & E(u_nu_2) & .. & E(u_n^2) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & .. & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

El que los valores de X sean números sin propiedades aleatorias implica que las variaciones aleatorias de Y son debidas únicamente a las variaciones de los errores $\mathbf{u_i}$. La última propiedad implica que el número de observaciones n, excede a los k parámetros considerados en el modelo. Podemos considerar también que X1 siempre toma el valor 1, con lo cual queda definida la primera columna de la matriz X. Si una variable Xj fuera combinación de una o más de las restantes variables X, entonces el rango de la matriz X será menor que k y ello dará lugar a que no exista la inversa de la matriz producto X'X y como consecuencia que tampoco exista solución para la estimación de los parámetros.

1.2 Estimadores de Mínimos Cuadrados.
El problema que se debe resolver inicialmente es el de obtener un vector de estimadores de los coeficientes del modelo:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \\ . \\ \hat{\beta}_k \end{bmatrix}$$

El modelo que se ajustará tendrá la forma

$$\boxed{Y = X\hat{\beta} + e}$$

Donde **e** es un vector columna (nx1) de valores residuales observables. Los residuales se pueden expresar en forma alternativa por un despeje simple como

$$\boxed{e = Y - X\hat{\beta}}$$

La suma de cuadrados de los residuales es equivalente al producto del vector **e'** por **e**.

$$\begin{aligned} \sum_{i=1}^{n} e_i^2 &= e'e \\ &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= (\mathrm{Y}' - \hat{\beta}'\mathrm{X}')(\mathrm{Y} - \mathrm{X}\hat{\beta}) \\ &= \mathrm{Y}'\mathrm{Y} - 2\hat{\beta}'\mathrm{X}'\mathrm{Y} + \hat{\beta}\mathrm{X}'\mathrm{X}\hat{\beta} \end{aligned}$$

Observe que la forma de construir el doble producto se apoya en que cada sumando es un escalar y en consecuencia equivale a su transpuesto.

Se busca un vector de coeficientes que minimice la suma de cuadrados de los residuales. Para minimizar esta suma de cuadrados, se toma derivada respecto al vector de coeficientes estimados.

$$\frac{\delta}{\delta\hat{\beta}} e'e = -2X'Y + 2X'X\hat{\beta}$$

La derivada se iguala al vector cero y se tendrá entonces

$$X'X\hat{\beta} = X'Y$$

Se toma la inversa de X'X, suponiendo que exista y se premultiplicá en ambos miembros para obtener la estimación del vector de coeficientes β de mínimos cuadrados ordinarios (MCO).

$$\boxed{\hat{\beta} = (X'X)^{-1} X'Y}$$

1.3 Propiedades de los Estimadores de Mínimos Cuadrados

Primero se verificará que el vector de estimadores de mínimos cuadrados es insesgado. Para esta verificación se parte de la expresión anterior y se sustituye el vector Y por la forma del modelo:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1} X'Y \\ &= (X'X)^{-1} X'(X\beta + u) \\ &= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'u\end{aligned}$$

El resultado es una expresión alternativa del vector de estimadores de los coeficientes.

$$\boxed{\hat{\beta} = \beta + (X'X)^{-1} X'u}$$

A continuación se toma valor esperado y se verifica el insesgamiento.

$$\begin{aligned}E\left(\hat{\beta}\right) &= E\left(\beta + (X'X)^{-1} X'u\right) \\ &= E\left(\beta\right) + E\left((X'X)^{-1} X'u\right) \\ &= E\left(\beta\right) + (X'X)^{-1} X'E(u) \\ &= \beta\end{aligned}$$

1.4 Varianza de los estimadores MCO

La matriz de varianzas y covarianzas de los estimadores de mínimos cuadrados se define como el valor esperado del producto del vector de diferencia entre el vector de estimadores y el de parámetros por su transpuesto:

$$V\left(\hat{\beta}\right) = E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right)$$

$$E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right) = \begin{bmatrix} E(\hat{\beta}_1 - \beta_1)^2 & E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) & .. & E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k) \\ E(\hat{\beta}_2 - \beta_2)(\hat{\beta}_1 - \beta_1) & E(\hat{\beta}_2 - \beta_2)^2 & .. & E(\hat{\beta}_2 - \beta_2)(\hat{\beta}_k - \beta_k) \\ .... & .... & .. & .... \\ E(\hat{\beta}_k - \beta_k)(\hat{\beta}_1 - \beta_1) & E(\hat{\beta}_k - \beta_k)(\hat{\beta}_2 - \beta_2) & .. & E(\hat{\beta}_k - \beta_k)^2 \end{bmatrix}$$

Se toma una expresión alternativa de la diferencia entre el vector de estimadores y el de parámetros.

$$\hat{\beta} - \beta = (X'X)^{-1} X'u$$

La matriz de varianzas y covarianzas de los estimadores se obtiene mediante el siguiente desarrollo algebraico:

$$V(\hat{\beta}) = E\Big((X'X)^{-1}X'u((X'X)^{-1}X'u)'\Big)$$
$$= E\Big((X'X)^{-1}X'uu'X(X'X)^{-1}\Big)$$
$$= \Big((X'X)^{-1}X'E(uu')X(X'X)^{-1}\Big)$$
$$= \Big((X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}\Big)$$
$$= \sigma^2(X'X)^{-1}$$

$$\boxed{V(\hat{\beta}) = \sigma^2(X'X)^{-1}}$$

Entonces basta multiplicar σ^2 por la matriz inversa de **X'X** para obtener la matriz de varianzas y covarianzas de los estimadores de los coeficientes MCO. La varianza de cualquier coeficiente estimado $\hat{\beta}_i$ se tiene disponible en el i-ésimo elemento de la diagonal principal de la última matriz y las covarianzas de $\hat{\beta}_i\ \hat{\beta}_j$ se ubican en el elemento i, j de la matriz referida.

Teorema de Gauss Markov
Hasta este punto sabemos que los estimadores MCO constituyen una combinación lineal insesgada. Si se prueba que son de varianza mínima, serán además estimadores MELI (Mejores Estimadores Lineales Insesgados). El teorema que prueba esta propiedad se conoce como Teorema de Gauss Markov.

Para probar que los estimadores de mínimos cuadrados son MELI, considérese la siguiente de manera general la combinación lineal:

$$c'\beta$$

Un estimador de esta combinación lineal en términos del vector de MCO

$$c'\hat{\beta} \quad \text{donde} \quad \hat{\beta} = (X'X)^{-1}X'Y$$

Este estimador es insesgado debido al insesgamiento del vector de estimadores de mínimos cuadrados (MCO).

$$E(c'\hat{\beta}) = c'\beta$$

y su varianza es entonces:

$$V(c'\hat{\beta}) = E((c'\hat{\beta} - c'\beta)(c'\hat{\beta} - c'\beta)')$$
$$= E((c'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'c')$$
$$= E(c'(X'X)^{-1}X'uu'X(X'X)^{-1}c)$$

$$= c'(X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}c$$
$$= \sigma^2 c'(X'X)^{-1}c$$

Considérese otro estimador alternativo de $c'\beta$

$b = a'Y$ donde **a** es un vector de constantes conocidas n x k

Se obtiene el valor esperado de este estimador

$$
\begin{aligned}
E(b) &= a'Y \\
&= E(a'(X\beta + u)) \\
&= E(a'X\beta + a'u) \\
&= a'X\beta
\end{aligned}
$$

Este estimador será insesgado para $c'\beta$ si y sólo si $a'X = c'$. Con este supuesto el nuevo estimador toma la forma:

$$b = c'\beta + a'u$$

La varianza del último estimador es

$$V(b) = E((b - c'\beta)(b - c'\beta)')$$

Como la diferencia $b - c'\beta = a'u$, entonces la varianza se expresa en forma equivalente como

$V(b) = E(a'uu'a)$ y por tanto

$$V(b) = \sigma^2 a'a$$

Se ha encontrado que

$$V(c'\hat{\beta}) = \sigma^2 c'(X'X)^{-1}c$$

Si se usa la equivalencia $a'X = c'$ entones la varianza anterior se puede escribir

$$V(c'\hat{\beta}) = \sigma^2 a'X(X'X)^{-1}X'a$$

Ahora se comparan las matrices de varianzas y covarianzas de ambos estimadores mediante su diferencia para determinar cual es más eficiente.

$$
\begin{aligned}
V(b) - V(c'\hat{\beta}) &= \sigma^2 a'a - \sigma^2 a'X(X'X)^{-1}X'a \\
&= \sigma^2 a'(I_n - X(X'X)^{-1}X')a \\
&= \sigma^2 a'Ma
\end{aligned}
$$

Donde

$$M = I_n - X(X'X)^{-1}X'$$

Se verifica que la matriz M es simétrica, idempotente y semidefinida positiva. Por lo tanto

$$\sigma^2 a'Ma \geq 0$$

y en consecuencia

$$V(b) - V(c'\hat{\beta}) \geq 0$$

Entonces la combinación lineal apoyada en los estimadores de MCO es más eficiente y por tanto MELI

1.5 Estimación de la Matriz de Varianzas y Covarianzas.

Se dispone de la forma que adopta la matriz de varianzas y covarianzas, pero no se dispone del procedimiento para estimar esta matriz a partir de los datos disponibles en muestra. Se parte de la expresión ya familiar para los residuales:

$$e = Y - X\hat{\beta}$$

El vector de residuales se pueden expresar en forma alternativa como

$$\begin{aligned} e &= X\beta + u - X(X'X)^{-1}X'(X\beta + u) \\ &= X\beta + u - X(X'X)^{-1}X'X\beta - X(X'X)^{-1}X'u \\ &= u - X(X'X)^{-1}X'u \\ &= (I_n - X(X'X)^{-1}X')u \\ e &= Mu \end{aligned}$$

Por lo tanto la suma de cuadrados de los residuales queda:

$$\begin{aligned} e'e &= u'M'Mu \\ &= u'Mu \quad \text{puesto que M es simétrica e idempotente} \\ &= Tr(u'Mu) \quad \text{por ser un escalar todo el producto} \\ &= Tr(uu'M) \quad \text{puesto que Tr(AB) = Tr(BA) con A,B conformables} \end{aligned}$$

Ahora se toma valor esperado de la suma de cuadrados de los residuales

$$\begin{aligned} E(e'e) &= E(Tr(uu'M)) \\ &= \sigma^2 Tr(M) \\ &= \sigma^2 Tr(I_n - X(X'X)^{-1}X') \\ &= \sigma^2 (Tr(I_n - Tr(X(X'X)^{-1}X')) \end{aligned}$$

$$= \sigma^2 (Tr(I_n - Tr(X'X(X'X)^{-1}))$$
$$= \sigma^2 (Tr(I_n) - Tr(I_k))$$
$$= \sigma^2 (n-k)$$

Entonces un estimador insesgado de la varianza del error será

$$\boxed{\hat{\sigma}^2 = \frac{e'e}{n-k}}$$

1.6 Normalidad de los Errores
Para llevar a efecto procedimientos inferenciales, tales como pruebas de significancia de hipótesis y cálculo de intervalos de confianza que involucren a los parámetros del modelo, es necesario contar con un supuesto de distribución.

Apoyándose en el teorema Central de Límite es razonable considerar que los errores **u_i**, cuyo valor es el resultado de múltiples factores desconocidos se distribuyan como una normal multivariada con vector medias nulo y matriz de varianzas y covarianzas dada por el producto $\sigma^2 I_n$.

El conjunto de hipótesis iniciales relativas a los errores se resumen entonces:

$$u_i \rightarrow N(0, \sigma^2 I_n)$$

Distribución de los Estimadores
El vector de estimadores de mínimos cuadrados para los coeficientes del modelo general se expresa como una combinación lineal de los errores y por tanto la normalidad de los errores se hereda al vector de estimadores, conservando sus propiedades de insesgamiento y varianza.

$$\hat{\beta} = \beta + (X'X)^{-1} X'u$$

El vector de estimadores sigue entonces una distribución normal multivariada

$$\hat{\beta} \rightarrow N(\beta, \sigma^2 (X'X)^{-1})$$

1.7 Estimadores de Máxima Verosimilitud
Es interesante comprobar que los estimadores de mínimos cuadrados que se obtuvieron sin hacer ningún supuesto de distribución, coinciden con los estimadores de máxima verosimilitud que suponen la normalidad de los errores.

Se toma como función de verosimilitud

$$L(\hat{\beta}) = \pi_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{u_i}{2\sigma^2}} \right)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n e^{-\frac{\sum_{i=1}^{n} u_i}{2\sigma^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n e^{-\frac{u'u}{2\sigma^2}}$$

Ahora se toma logaritmo natural de la función de verosimilitud

$$\ln\left(L(\hat{\beta})\right) = \ln\left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n - \frac{u'u}{2\sigma^2}$$

Como **Y = Xβ + u**

$$\ln\left(L(\hat{\beta})\right) = \ln\left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n - \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}$$

A continuación se deriva e iguala a cero la anterior expresión

$$\frac{\delta}{\delta\beta} \ln\left(L(\hat{\beta})\right) = \frac{\delta}{\delta\beta} \left(-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \right)$$

$$\frac{\delta}{\delta\beta} \ln\left(L(\hat{\beta})\right) = \frac{1}{2\sigma^2} \frac{\delta}{\delta\beta} \left(\mathrm{Y'\,Y} - 2\beta' \mathrm{X'\,Y} + \hat{\beta} \mathrm{X'\,X} \beta \right)$$

Pero esta optimización es la misma que se utilizó para los estimadores de mínimos cuadrados.

$$= -2X'Y + 2X'X\beta = 0$$

Se concluye que el vector de estimadores de MV coincide con el de MCO

$$\boxed{\hat{\beta} = (X'X)^{-1} X'Y}$$

Por las razones expuestas, cualquiera de los elementos del vector de estimadores de mínimos cuadrados se distribuye normalmente

$\hat{\beta}_i \rightarrow N(\beta_i, \sigma^2 a_{ii})$ i=1,2,……k

Donde $\mathbf{a_{ii}}$ es el elemento i-ésimo de la diagonal principal de la matriz $(X'X)^{-1}$

1.8 Inferencia sobre los parámetros.

El conocimiento de la varianza del error σ^2 permitiría probar hipótesis y calcular intervalos de confianza para los parámetros usando la normal, después de las adecuadas estandarizaciones. El problema es que se desconocer la varianza del error y hay que proceder con su estimación y con una distribución alternativa que considere este hecho. Esta distribución es la t de Student (Gosset)

Recuérdese la fórmula alternativa de la suma de residuales al cuadrado

$$\begin{aligned} e'e &= u'Mu \\ &= u'(I_n - X(X'X)^{-1}X')u \end{aligned}$$

Se ha mencionado que la matriz **M** es idempotente, simétrica y con traza **n-k** y de rango **n-k** . Ahora se trata de encontrar una matriz P ortogonal tal que **P'MP=** $\mathbf{D_{n-k}}$ donde $\mathbf{D_{n-k}}$ es una matriz diagonal **n-k** elementos mayores a cero y **k** elementos iguales a cero. La matriz **P** se utiliza también para definir una transformación del vector **u** a un vector **v**

u = P v **v= P' u** puesto que al ser ortogonal $\mathbf{P^{-1}}$ = **P'**

Entonces

$$\begin{aligned} e'e &= u'Mu \\ &= v'P'MPv \\ &= v'D_{n-k}v \\ &= \sum_{i=1}^{n-k} v_i^2 \end{aligned}$$

Las $\mathbf{u_i}$ se distribuyen normalmente con media cero y varianza σ^2 y gracias a las propiedades de las matrices ortogonales heredan su distribución a las $\mathbf{v_i}$. La suma de cuadrados de los residuales **e'e** se expresan con la ayuda de esta transformación como la suma de los cuadrados de **n-k** variables normales con media cero y varianza σ^2. Si la suma de cuadrados se divide previamente entre σ^2, entonces **e'e/** σ^2 se distribuye Ji cuadrada con **n-k** grados de libertad.

Si se tiene una variable Z que se distribuya normalmente y una variable W que se distribuya Ji cuadrada con r grados de libertad, independiente de Z, entonces la siguiente variable T se distribuye t de Student con r grados de libertad.

$$T = \frac{Z}{\sqrt{\frac{W}{r}}}$$

Se mencionó que cada estimador de coeficiente del modelo se distribuye normalmente con media igual al coeficiente parametral y varianza dada por el producto de la varianza del error multiplicada por el elemento a_{ii} de la diagonal principal de la matriz $(X'X)^{-1}$

$$\boxed{\hat{\beta}_i \rightarrow N(\beta_i, \sigma^2 a_{ii})}$$

Entonces se puede construir una Z normal estándar.

$$Z_i = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 a_{ii}}} \rightarrow N(0,1)$$

También una T que se distribuye t de Student con n-k grados de libertad para probar hipótesis sobre los coeficientes

$$T_i = \frac{\dfrac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 a_{ii}}}}{\sqrt{\dfrac{e'e}{\sigma^2 (n-k)}}} = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\dfrac{e'e}{(n-k)}}\sqrt{a_{ii}}}$$

La prueba de hipótesis más usual es Ho: $\beta = 0$ vs Ha: $\beta \neq 0$

El producto de radicales en el denominador constituyen el error estándar del respectivo estimador del coeficiente.

$$EE(\hat{\beta}_i) = \sqrt{\frac{e'e}{(n-k)}}\sqrt{a_{ii}}$$

Su cuadrado proporciona el estimador de la varianza del estimador del coeficiente

$$\hat{V}(\hat{\beta}_i) = \frac{e'e a_{ii}}{(n-k)}$$

Esta misma estadística permite construir intervalos de confianza para los coeficientes apoyados en la distribución t de Student con n-k grados de libertad.

$$\boxed{\hat{\beta} \pm t_{1-\alpha/2, n-k} \sqrt{\frac{e'e}{(n-k)}}\sqrt{a_{ii}}}$$

1.9 Coeficiente de Correlación Simple

La magnitud y dirección de la relación lineal entre dos variables X,Y se mide mediante el coeficiente de correlación producto momento de Pearson, el cual se define como el cociente de la covarianza de las dos variables entre el producto de sus desviaciones estándares. El coeficiente de correlación está acotado en el intervalo (-1,1). Sus valores extremos corresponden a una perfecta relación lineal directa y a una perfecta relación lineal inversa.

$$\rho = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

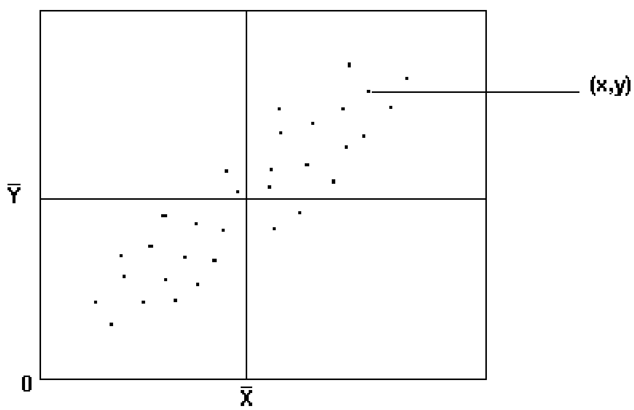
Para su cálculo empírico considere **n** observaciones de dos variables medidas en términos de sus desviaciones, lo que equivale a trasladar el origen de coordenadas al punto de medias.

$$x = X - \bar{X}$$
$$y = Y - \bar{Y}$$

En términos de las desviaciones, el coeficiente de correlación se calcula con la siguiente fórmula:

$$r = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}}$$

Si existe relación lineal directa, los productos xy se ubicarán de manera dominante en el primero y tercer cuadrantes. Sus signos positivos, darán lugar a una suma de productos de desviaciones positiva y desde luego a correlación positiva. Si la relación es inversa, entonces los puntos se ubicarán en el primero y cuarto cuadrantes con signos negativos en los productos y por tanto con correlación negativa.



En el análisis de regresión es usual que se calcule la matriz de coeficientes de correlación de todas las variables que intervienen en el modelo. La matriz de correlación es una matriz simétrica con unos en la diagonal principal.

$$\begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & 1 \end{bmatrix}$$

1.10 Coeficiente de Correlación Parcial

Los coeficientes de correlación simples miden la relación lineal entre dos variables, sin embargo, son perturbados por la presencia de las otras variables en el modelo que a su vez correlacionan con ellas. Esta limitación es resuelta mediante el coeficiente de correlación parcial, el cual mide la relación lineal entre dos variables manteniendo constantes las restantes variables.

Para simplificar su comprensión, considérese un modelo de 3 variables:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$
$$i = 1,2,\dots\dots n$$

A continuación considere la regresión de Y sobre X_3

$$Y_i = \hat{\beta}_{12} + \hat{\beta}_1 X_{3i} + k_i$$
$$i = 1,2,......n$$

Esta regresión, expresada en términos de desviaciones respecto a la media se presenta sin la presencia explícita del término constante donde k_i es el residual y éste a su vez, es la diferencia de **y** menos la influencia lineal de **x3**.

$$y_i = \hat{\beta}_1 x_{3i} + k_i$$
$$i = 1,2,......n$$

de donde k_i se expresa:

$$k_i = y_i - \hat{\beta}_{13} x_{3i}$$
$$i = 1,2,......n$$

En forma similar considere la regresión de X_2 sobre X_3 en términos de desviaciones respecto a la media donde l_i es el residual del modelo ajustado y expresa el valor de y una vez eliminada la influencia lineal de x_3.

$$x_{2i} = \hat{\beta}_{23} x_{3i} + l_i$$
$$i = 1,2,......n$$

y entonces

$$l_i = x_{2i} - \hat{\beta}_{23} x_{3i}$$
$$i = 1,2,......n$$

El coeficiente de correlación parcial de Y con X_2, una vez eliminada la influencia lineal de X_3 se calcula como el coeficiente de correlación simple entre los residuales k_i y l_i :

$$r_{12.3} = \frac{\sum_{i=1}^{n} k_i l_i}{\sqrt{\sum_{i=1}^{n} k_i^2}\sqrt{\sum_{i=1}^{n} l_i^2}}$$

El coeficiente de correlación parcial entre Y y X_2 se puede expresar en forma alternativa en términos de los coeficientes de correlación simple como sigue:

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

En forma análoga si se tienen k variables, los coeficientes de correlación parcial expresan la relación entre dos variables, después de restar la influencia lineal de las otras variables.

1.11 Coeficiente de Determinación R^2

Se retorna brevemente a la definición inicial del modelo cuyos valores de variables se miden en unidades originales y toman al origen **0** como referencia.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{2i} .... \beta_k X_{2i} + u_i$$
$$i = 1,2,......n$$

Ahora se promedia la expresión anterior

$$\overline{Y} = \beta_1 + \beta_2 \overline{X}_2 + \beta_3 \overline{X}_3 .... \beta_k \overline{X}_k + \overline{u}$$

Se resta la segunda expresión de la primera y se definen nuevas variables expresadas en desviaciones respecto a su media.

$$Y_i - \overline{Y} = (\beta_1 - \beta_1) + \beta_2 (X_{2i} - \overline{X}_2) + \beta_3 (X_{2i} - \overline{X}_3) .... \beta_k (X_{2i} - \overline{X}_k) + (u_i - \overline{u})$$

La diferencia trae como consecuencia que la ordenada al origen β_1 se anule, pues el modelo pasa ahora a través del origen.

$$y_i = \beta_2 x_{2i} + \beta_3 x_{2i} .... \beta_k x_{2i} + (u_i - \overline{u})$$
$$i = 1,2, ... n$$

De manera análoga la variable dependiente y se expresa en términos de los estimadores de mínimos cuadrados, las variables independientes en forma de desviaciones respecto a la media y los residuales, puesto que la media de los residuales siempre es nula.

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{2i} .... \hat{\beta}_k x_{2i} + e_i$$
$$i = 1,2, ... n$$

En forma matricial $y = X\beta + (u - \overline{u})$ y $y = X\hat{\beta} + e$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \\ . \\ y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} x_{21} & x_{31} & x_{k1} \\ x_{22} & x_{32} & x_{k2} \\ \\ x_{2n} & x_{3n} & x_{kn} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_2 \\ \beta_3 \\ \\ . \\ \beta_k \end{bmatrix} \qquad \mathbf{u} - \overline{u} = \begin{bmatrix} u_1 - \overline{u} \\ u_2 - \overline{u} \\ \\ . \\ u_n - \overline{u} \end{bmatrix}$$

Donde X es de orden nx(k-1) y **β** es (k-1)x1

Los residuales se pueden expresar como

$$\begin{aligned} e'e &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\ &= y\acute{y} - \hat{\beta}' X' X\hat{\beta} \\ &= y\acute{y} - \hat{\beta}' X' y \quad \text{puesto que} \quad \hat{\beta} = (X'X)^{-1} X' y \end{aligned}$$

El producto escalar yý representa la suma de cuadrados de las desviaciones de la variable dependiente respecto de su media y refleja la variabilidad total de la variable dependiente.

$$y' y = \sum_{i=1}^{n} y_i^2$$

La suma de cuadrados de desviaciones respecto de la media se expresa como la suma de cuadrados de los residuales más el producto $\hat{\beta}' X' y$.

$$y'y = \hat{\beta}'X'y + e'e$$

Si se divide ambos miembros de la igualdad entre y'y, entonces $\hat{\beta}'X'y$ /y'y representa la proporción de la variabilidad total de **Y** explicada por el modelo y e'e la parte no explicada por el modelo. Esta proporción se denomina R^2, el Coeficiente de Determinación.

$$1 = \frac{\hat{\beta}'X'y}{y'y} + \frac{e'e}{y'y}$$

$$R^2 = \frac{\hat{\beta}'X'y}{y'y}$$ o en forma alternativa $$\boxed{R^2 = 1 - \frac{eé}{y'y}}$$

1.12 Coeficiente de Determinación Ajustado por Grados de Libertad.
Si a la suma de cuadrados de los residuales y a la suma de cuadrados de la variable dependiente se los divide entre sus respectivos grados de libertad, entonces se tiene el Coeficiente de Determinación ajustado por grados de libertad. Esta última forma permite la comparación de modelos ajustados a los mismos datos con diferente número de variables.

$$\boxed{R^2 = 1 - \frac{eé/(n-k)}{y'y/(n-1)}}$$

Si en lugar de tomar desviaciones respecto a la media, se respetan las unidades originales, la R^2 sin corregir por los grados de libertad se expresa como sigue:

$$R^2 = \frac{\hat{\beta}'X'Y - (1/n)\left(\sum_{i=1}^{n} Y_i\right)^2}{Y'Y - (1/n)\left(\sum_{i=1}^{n} Y_i\right)^2}$$

El coeficiente de correlación múltiple R es la raíz cuadrada positiva del coeficiente de determinación y representa el grado de relación lineal existente entre la variable dependiente y la combinación lineal dada por el conjunto de variables independientes incluidas en el modelo.

1.13 Análisis de Varianza en Regresión.
Se ha mencionado que **$e'e/\sigma^2$** se distribuye como una Ji cuadrada con n-k grados de libertad, por su parte el producto $\hat{\beta}'X'y$ / **σ^2** bajo la hipótesis de que el vector de coeficientes parametrales **β** de las variables independientes sea nulo, se distribuye Ji cuadrada con k-1 grados de libertad en forma independiente, de modo que el siguiente cociente se distribuye **F** con k-1 y n-k grados de libertad y permite probar la hipótesis de nulidad del vector de coeficientes de las variables independientes **β** del modelo

$$F = \frac{\hat{\beta}'X'y/(k-1)}{e'e/(n-k)}$$

Esta prueba de hipótesis se presenta en forma de una tabla de análisis de la varianza de la variable dependiente, cuyas fuentes de variación son el modelo de regresión y la parte residual o no explicada por el modelo.

Tabla ANOVA para la Regresión

Fuente de Variación	Sumas de Cuadrados	Grados de Libertad	Cuadrados medios	Estadística F k-1,n-k
Modelo de Regresión	$\hat{\beta}' X' y$	k - 1	$\hat{\beta}' X' y /(k-1)$	$F = \dfrac{\hat{\beta}' X' y7(k-1)}{\acute{e}\acute{e}/(n-k)}$
Residual	$e'e$	n - k	$e'e /(n-k)$	
Total	$y'y$	n - 1		

1.14 Predicción
El modelo de regresión lineal múltiple tiene como objetivo la explicación de la relación lineal de la variable dependiente con el conjunto de variables independientes, pero también es frecuente que se lo utilice con fines predictivos del valor esperado de Y en el periodo n+1, dados valores de las X en el periodo n+1 . Considérese el vector **c** de valores de X

$$\mathbf{c} = \begin{bmatrix} 1 \\ X_{2,n+1} \\ X_{3,n+1} \\ . \\ X_{k,n+1} \end{bmatrix}$$

y que se desea predecir

$$E(Y_{n+1}) = E(c'\beta + u_{n+1}) = c'\beta$$

Hemos visto que $c'\hat{\beta}$ es el mejor estimador lineal insesgado de $c'\beta$. Por lo tanto la mejor estimación de Yn+1 será:

$$\hat{Y}_{n+1} = c'\hat{\beta} = \hat{\beta}_1 + \hat{\beta}_2 X_{2,n+1} + \hat{\beta}_3 X_{3,n+1} \dots\dots\dots + \hat{\beta}_k X_{k,n+1}$$

1.15.1 Intervalo de confianza para el valor medio de Y
Si además se quiere hacer una estimación por intervalo de confianza para el valor medio de Y dado c, entonces se debe conocer la distribución muestral de $c'\hat{\beta}$.

Por ser $\hat{\beta}$ insesgado, la combinación lineal resulta insesgada.

$$E(c'\hat{\beta}) = c' E(\hat{\beta}) = c'\beta$$

También se sabe que la matriz de varianzas y covarianzas tiene la propiedad:

$$V(c'\hat{\beta}) = \sigma^2 c'(X'X)^{-1}c$$

Como $\hat{\beta}$ tiene una distribución normal multivariante, entonces la combinación lineal $\hat{Y} = c'\hat{\beta}$ también se distribuye normal.

$$\hat{Y} = c'\hat{\beta} \rightarrow N(c'\beta, \sigma^2 c'(X'X)^{-1}c)$$

Al estandarizar se tiene

$$Z = \frac{c'\hat{\beta} - c'\beta}{\sqrt{\sigma^2 c'(X'X)^{-1}c}} \rightarrow N(0,1)$$

Puesto que se desconoce σ^2 se utiliza su estimador y la estadística resultante se distribuye t de Student con n-k grados de libertad.

$$T = \frac{c'\hat{\beta} - c'\beta}{\sqrt{\hat{\sigma}^2 c'(X'X)^{-1}c}} \rightarrow t_{n-k}$$

Entonces un intervalo de 100(1-α)% de confianza para el valor medio de Y_{n+1} tendrá la forma

$$\boxed{c'\hat{\beta} \pm t_{n-k,1-\alpha/2}\hat{\sigma}\sqrt{c'(X'X)^{-1}c}}$$

1.15.2 Intervalo de confianza para el valor puntual de Y

Si el Interés en la construcción del intervalo de confianza radica no el valor medio de Y, sino en el valor puntual de Y dado un vector **c**, entonces hay que hacer una ligera modificación al intervalo anterior.

Nuevamente se considera el valor estimado de Y

$$\hat{Y}_{n+1} = c'\hat{\beta} = \hat{\beta}_1 + \hat{\beta}_2 X_{2,n+1} + \hat{\beta}_3 X_{3,n+1} \ldots\ldots\ldots + \hat{\beta}_k X_{k,n+1}$$

El valor observado de Y_{n+1} es expresado por el modelo de la siguiente forma

$$Y_{n+1} = c'\beta + u_{n+1} = \beta_1 + \beta_2 X_{2,n+1} + \beta_3 X_{3,n+1} \ldots\ldots\ldots + \beta_k X_{k,n+1} + u_{n+1}$$

Se tienen entonces dos expresiones que hay que relacionar:

y $Y_{n+1} = c'\beta + u_{n+1}$

Se plantea la diferencia entre el valor dado por la predicción y el real

$$d = \hat{Y}_{n+1} - Y_{n+1}$$

$$d = c(\hat{\beta} - \beta) - u_{n+1}$$

Como $\hat{\beta}$ es insesgado, y la esperanza del error es cero, la esperanza de la diferencia también es cero.

$$E(d) = 0$$

La varianza de la diferencia es la siguiente suma de varianzas

$$V(d) = V(c\hat{\beta}) + V(u_{n+1})$$

$$= \sigma^2 c'(X'X)^{-1}c + \sigma^2 = \sigma^2(c'(X'X)^{-1}c + 1)$$

Se define en forma análoga al caso del valor medio la estadística T en función del estimador de la varianza del error. La estadística T se distribuye t de Student con n-k grados de libertad

$$T = \frac{\hat{Y}_{n+1} - Y_{n+1}}{\hat{\sigma}\sqrt{1 + c'(X'X)^{-1}c}} \rightarrow t_{n-k}$$

A partir de esta estadística se construye el intervalo de 100(1-α)% de confianza para la predicción puntual de Y_{n+1}

$$\boxed{\hat{Y}_{n+1} \pm t_{1-\alpha/2,n-k}\hat{\sigma}\sqrt{1 + c'(X'X)^{-1}c}}$$

1.16 Análisis de Residuales y Valores Influyentes.

El análisis de residuales en regresión múltiple constituye una valiosa herramienta de diagnóstico del modelo y para la identificación de valores extremos que tengan un efecto notable en el comportamiento del modelo ajustado y que eventualmente pudieran corresponder a observaciones espurias o errores de registro. Sin embargo hay que tener cautela al suprimir observaciones en forma muy liberal, pues se corre el riesgo de disminuir artificialmente la varianza natural del fenómeno en estudio. Existe una amplia gama de estadísticas asociadas a los residuales que se comentan a continuación.

Residuales no estandarizados. Corresponden a la diferencia entre el valor observado de la variable dependiente **Y** menos el valor estimado por el modelo ajustado.

$$\boxed{e_i = Y_i - \hat{Y}_i}$$

Residuales estandarizados. Es el cociente del residual entre la raíz cuadrada de la estimación de la varianza del error. En otros términos son residuales relativos a su desviación estándar. Se podrían considerar como valores extremos aquellos residuales estandarizados que sean mayores en valor absoluto a 2, esto es a 2 desviaciones. Los residuales estandarizados que superen esta cifra serán motivo de revisión.

$$\boxed{Ze_i = \frac{e_i}{\sqrt{\sum_{i=1}^{n} e_i / (n-k)}}}$$

Leverage. Literalmente se traduce como impulso. Refleja el impacto o influencia de la observación de referencia en los parámetros estimados del modelo y en las estimaciones hechas por el mismo. Su promedio es el cociente entre el número de parámetros estimados para el modelo, incluida la constante, entre el total de observaciones. Un valor de leverage que sea mayo a 2 veces el promedio es motivo de atención por parte del investigador.

Los valores de leverage corresponden a la diagonal de la llamada matriz gorro (hat) denotada por H en el siguiente desarrollo y es la matriz cuadrada nxn que premultiplicada por el vector de observaciones de la variable dependiente da lugar al vector de estimaciones de ahí su nombre:

$\hat{\beta} = (X'X)^{-1}X'Y$ es el vector de estimaciones de los coeficientes del modelo

$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y$ es el vector de estimaciones

$H = X(X'X)^{-1}X'$ es una matriz cuadrada nxn

$$\hat{Y} = HY$$

El Leverage de la observación i-ésima corresponde al elemento $\mathbf{h_{ii}}$ de la diagonal principal de la matriz **H**. Por ser H una matriz de gran dimensión, en la práctica los algoritmos de cálculo se orientan únicamente a la diagonal. Algunos paquetes, como el SPSS, calculan una versión corregida del leverage al sumarle 1/n y utilizar este valor corregido en el cálculo de otras estadísticas como la distancia de Cook. La suma de los elementos de la diagonal de la matriz H es igual a p, el número de parámetros y puesto que es de orden nxn, hay n elementos en la diagonal. Así el promedio es p/n. Como recomendación se analizarán con detalle aquellas observaciones cuyo leverage exceda en 2 veces el promedio, esto es 2p/n.

Residuales Studentizados. Se calcula como el cociente del residual entre la raíz cuadrada de uno menos el leverage de la observación. Los residuales Studentizados tienen una interpretación similar a la de los residuales estandarizados con la adición de una corrección en el cálculo de las varianzas asociadas a los residuales. (El leverage utilizado en las siguientes fórmulas se modifica con la adición de 1/n para hacerlo consistente con el procedimiento utilizado por SPSS)

$$\boxed{Se_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1-h_i)}}}$$

Distancia de Cook. Constituye una medida alternativa de la influencia de las observaciones. Se calcula como la suma de cuadrados de las diferencias entre las estimaciones de la variable dependiente menos la estimación efectuada después de haber suprimido la i-ésima observación del cálculo, dividida entre el producto del número de parámetros por el error estándar de la estimación. Una forma alternativa de cálculo en función del leverage y que para propósitos prácticos es más eficiente se presenta enseguida.

$$DCook_i = \frac{\sum_{j=1}^{n}\left(\hat{Y}_i - \hat{Y}_{j(i)}\right)^2}{p\hat{\sigma}}$$

$$\boxed{DCook_i = \left(\frac{e_i}{\hat{\sigma}\sqrt{1-h_i}}\right)^2 \frac{h_i}{1-h_i}\frac{1}{p}}$$

Distancia Mahalanobis. En el caso de regresión esta distancia se calcula como (n-1) veces el valor de leverage asociado a la observación. Como regla de aplicación, algunos autores recomiendan analizar aquellas observaciones cuya distancia de Mahalanobis sea mayor al valor crítico de una distribución Ji cuadrada con grados de libertad igual a el número de parámetros p-1 y probabilidad asociada de 0.001.

$$DMh_i = (n-1)h_i$$

Delta Beta (Dfbeta). Permiten valorar el cambio o incremento en el coeficiente de referencia en caso de que la observación sea eliminada del conjunto de datos. Si la Dfbeta es positiva, la pendiente se incrementa y se decrementa en caso contrario. La recomendación usual es considerar como valores extremos a aquellas observaciones cuya Dfbeta> $2/\sqrt{n}$

$$\left(\Delta\hat{\beta}_j\right)_i = \left(\hat{\beta}_j - \hat{\beta}_{j,i}\right) = \frac{(X'X)^{-1}X_i'e_i}{1-h_i}$$ donde $X_i^{''} = (1, X_{1i}, X_{2i} .... X_{ki}$

Delta Beta Estandarizadas (SDbeta). Es la Dfbeta estandarizada y tiene la ventaja de que su interpretación no se ve afectada por las unidades de medición.

$$S\left(\Delta\hat{\beta}_j\right)_i = \frac{\left(\hat{\beta}_j - \hat{\beta}_{j,i}\right)}{s_i\sqrt{(X'X)^{-1}{}_{ii}}}$$

Delta Fit (Dffit). Permiten valorar el cambio en la estimación de la variable dependiente **Y** correspondiente a la observación si la observación es eliminada del cálculo. Se suele calcular también una versión estandarizada que prescinde de las unidades de la variable dependiente Y.

$$Dffit_i = \frac{e_i h_i}{1-h_i}$$