UNIVERSIDAD NACIONAL DE INGENIERIA

Facultad de Ingeniería Económica, Estadistíca y CC.SS.

Curso: Modelos Lineales

Modelos Lineal General

Eduardo Marcos Sánchez

Modelos

Para representar e interpretar formalmente cualquier observación utilizamos modelos.

Un modelo es una representación abstracta de cómo creemos que se comporta el fenómeno en estudio.

Los modelos son herramientas para evaluar las hipótesis de investigación.

El modelo particular que nos ocupa son los ***modelos estadísticos***, los cuales se utilizan en general en el marco de un proceso deductivo (Ver el método hipotético deductivo). Se trata de explicar la ***variabilidad*** de un fenómeno particular tratando de encontrar y comprender los componentes de esta *variabilidad*.

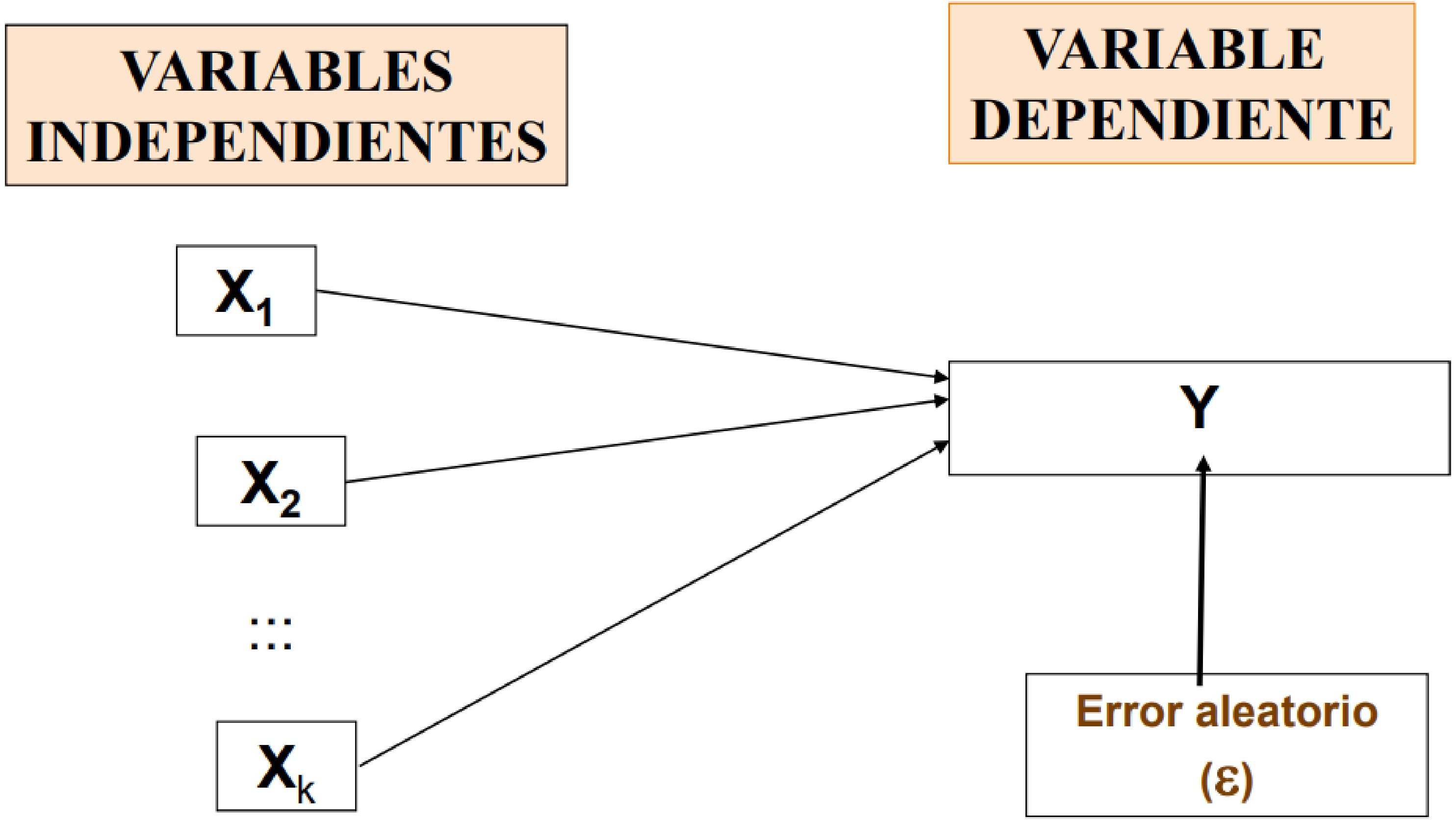
El método de análisis consiste en elaborar un cierto número de hipótesis sobre los factores que explican la variabilidad del fenómeno en estudio y estas hipótesis son verificadas a partir de los datos observados

Modelo Lineal General

En el contexto de la Estadística, se asume que toda observación está formado por dos componentes, la ***componente sistemática***, que representa la tendencia que siguen los datos y que puede ser modelado mediante una función matemática (funciones lineales y no lineales). La segunda componente corresponde a comportamientos en los que se observan sus manifestaciones pero no son medibles y solo se pueden representar mediante modelos probabilísticos (distribución normal, poisson, binomial, etc.), denominada componente estocástica.

El ***modelo lineal general***, constituye la base de la mayoría de los modelos que han surgido a partir de este. Por ejemplo los modelos no lineales, modelos lineales generalizados, modelos de efectos mixtos, modelos lineales aditivos, entre otros

Modelo Lineal General



Modelo Lineal General

Modelo Lineal General

Modelo de Rango Completo

Def. Las columnas de la matriz son linealmente indep. (rangos)

Modelo de Rango Incompleto

Def. Las columnas de la matriz son linealmente dep. (rangos)

Modelo Lineal General

Definición.-

El modelo lineal general se define como:

$$Y = \sum^{k} \beta_j x_j + \varepsilon$$

Donde:

(1) Y es una variable aleatoria continua, observable

(2) Las variable independientes $x_1, x_2, \ldots, x_k$ son fijas y conocidas y pueden ser cuantitativas (numéricas) o cualitativas (categóricas)

(3) Los βj son parámetros (constantes) desconocidas definidas en un espacio paramétrico Ω_β;

(4) ε es una variable aleatoria inobservable con distribución de probabilidades conocida $E(\varepsilon)=0$, $V(\varepsilon)=\sigma^2$), $cov(\varepsilon_i \varepsilon_j)=0$ para todo $i \neq j$

Modelo Lineal General

Forma matricial del modelo

Definición.-

Supóngase que, se selecciona una muestra aleatoria de tamaño n dado por $\{(y_i, x_{i1}, x_{i1}, \ldots, x_{ik}): i=1,2,\ldots,n\}$ de una población de la forma , entonces la observación de yi para cada individuo se puede escribir como:

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \ldots + \beta_k x_{1k} + \varepsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_{12} + \ldots + \beta_k x_{2k} + \varepsilon_2 \\
\vdots \quad & \vdots \quad \vdots \\
y_n &= \beta_0 + \beta_1 x_{n2} + \ldots + \beta_k x_{nk} + \varepsilon_n
\end{aligned}
$$

Modelo Lineal General

Forma matricial del modelo

Expresando matricialmente el sistema de ecuaciones, tenemos

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1k} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n0} & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

El modelo lineal general se expresa como:

$$\mathbf{y}_{nx1} = \mathbf{X}_{nx(k+1)}\, \boldsymbol{\beta}_{(k+1)x1} + \boldsymbol{\varepsilon}_{nx1}$$

Donde:

(1) $\mathbf{y}_{nx1}$ es un vector de v.a. observables

(2) $\mathbf{X}_{nx(k+1)}$ matriz de números observables (Los elementos de X no son variables aleatorias.

(3) $\boldsymbol{\beta}_{(k+1,1}$ es un vector de parámetros desconocidos definido en un espacio

(4) $\boldsymbol{\varepsilon}_{nx1}$ es un vector aleatorio continuo, inobservable con $E(\boldsymbol{\varepsilon}_{nx1})=0$ y $cov[\boldsymbol{\varepsilon}_{nx1}]=\Sigma$

Modelo Lineal General

Forma matricial del modelo

Notas.-

(1) En la especificación(4) la cov($\boldsymbol{\varepsilon}_{nx1}$)=$\Sigma$, es igual a la cov($\mathbf{y}_{nx1}$) existe y se denota como $\boldsymbol{\Sigma}$.

(2) En muchos casos, como parte del modelo, se establecen suposiciones adicionales, acerca de la distribución de probabilidades de $\boldsymbol{\varepsilon}_{nx1}$ (o de $\mathbf{Y}$) acerca de la estructura de $\boldsymbol{\Sigma}$.

(3) Usualmente se asume que $\mathbf{x}_{0i}$=**1**.

Modelo Lineal General

Ejm: Modelo de Rango Completo

Y : Calificación respecto a la satisfacción en el trabajo.

X_1: Calificación en un test de inteligencia

X_2: Índice de adaptación personal

Supongamos que el vector $(Y,X_1,X_2) \sim N_3(\boldsymbol{\mu},\boldsymbol{\Sigma})$

Recordemos que $Y/X_1=x_1,X_2=x_2) \sim N(\mu,\sigma^2)$

Además, $E(Y/X_1=x_1,X_2=x_2) = \mu = \beta_0+\beta_1x_1+\beta_2x_2$

El modelo de regresión lineal para cada una de las n=31 observaciones se expresaría como:

$$y_i = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2}+\varepsilon_i \qquad i=1,2,\ldots,31$$

La componente aleatoria, ε, se distribuye igual que la variable dependiente, esto es,

$E(\varepsilon)=0$, Var $(\varepsilon)=\sigma^2$, $cov(\varepsilon_i,\varepsilon_j)=0$; además, la distribución de ε es normal.

Modelo Lineal General

Ejm: Modelo de Rango Completo

$$E(Y/X_1=x_1, X_2=x_2) = \mu = \beta_1 x_1$$

Calificación respecto a la satisfacción en el trabajo (Y)	Calificación en un test de inteligencia (x_1)	Indice de adaptación personal (X_2)
54	15	8
37	13	1
30	15	1
48	15	7
37	10	4
37	14	2
31	8	3
49	12	7
43	1	9
12	3	1
30	15	1
37	14	2
61	14	10
31	9	1
31	4	5

$$\mu_1=E(Y_1/X_1=15, X_2=8)= \beta_0+\beta_1 15+\beta_2 8$$

$$\mu_2=E(Y_2/X_1=13, X_2=1)= \beta_0+\beta_1 13+\beta_2(1)$$

::::

$$\mu_{15}=E(Y_{15}/X_1=4, X_2=5)= \beta_0+\beta_1 4+\beta_2 5$$

Modelo Lineal General

Ejm: Modelo de Rango Completo

Un estudio fue conducido para examinar aquellas posibles variables relacionadas con la satisfacción en el trabajo de los empleados sin un grado profesional, de los hospitales. Una muestra aleatoria de 15 empleados produjo los siguientes resultados:

Calificación respecto a la satisfacción en el trabajo (Y)	Calificación en un test de inteligencia (x_1)	Indice de adaptación personal (X_2)
54	15	8
37	13	1
30	15	1
48	15	7
37	10	4
37	14	2
31	8	3
49	12	7
43	1	9
12	3	1
30	15	1
37	14	2
61	14	10
31	9	1
31	4	5

Modelo Lineal General

Ejm: Modelo de Rango Completo

El modelo expresado matricialmente es dado por:

$$\begin{pmatrix} 54 \\ 37 \\ \vdots \\ 31 \end{pmatrix}_{15x1} = \begin{pmatrix} 1 & 15 & 8 \\ 1 & 13 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 4 & 5 \end{pmatrix}_{15x3} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}_{3x1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{15} \end{pmatrix}_{15x1}$$

El rango máximo de la matriz de diseño, X es 3, (En este caso es exactamente 3)

Modelo Lineal General

El modelo de regresión lineal tiene un conjunto de supuestos formulados en términos de los errores, ε

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

Donde :
ε : Es la componente estocástica y es inobservable

Supuestos del modelo:

1) $E(\varepsilon)=0$ (insesgamiento)
2) $V(\varepsilon)=\sigma^2$ Varianza constante (homocedasticidad)
3) $Cov(\varepsilon_j\ \varepsilon_k)=0$ (Errores incorrelacionados)
4) $\varepsilon \sim N(0, \sigma^2)$ (Ruido blanco)
5) Las variables explicativas son fijas, conocidas e independientes entre si.

Modelo Lineal General

Los supuestos también se pueden expresar en función de la variable respuesta

a) $E(Y / X_1 = x_1, X_2 = x_2, ..., X_k = x_k) = \mu = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + ... + \beta_k x_k$ Insesgamiento

b) $V(Y / X_1 = x_1, X_2 = x_2, ..., X_k = x_k) = \sigma^2$ Homocedasticidad

c) $Cov(y_j\ y_k)=0$ (Observaciones de la respuesta incorrelacionados)

d) $Y \sim N(\mu, \sigma^2)$

e) Las k variables explicativas son fijas, conocidas e independientes entre si.

Modelo Lineal General

El vector de coeficientes $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$ y σ^2 , son desconocidos y para estimar la variable respuesta o realizar pronósticos; es necesario estimar estos parámetros.
Existen varios métodos de estimación:

Enfoque clásico

- Método de mínimos cuadrados ordinarios
- Método de máxima verosimilitud
- Métodos robustos
- Métodos no paramétricos
- Redes neuronales

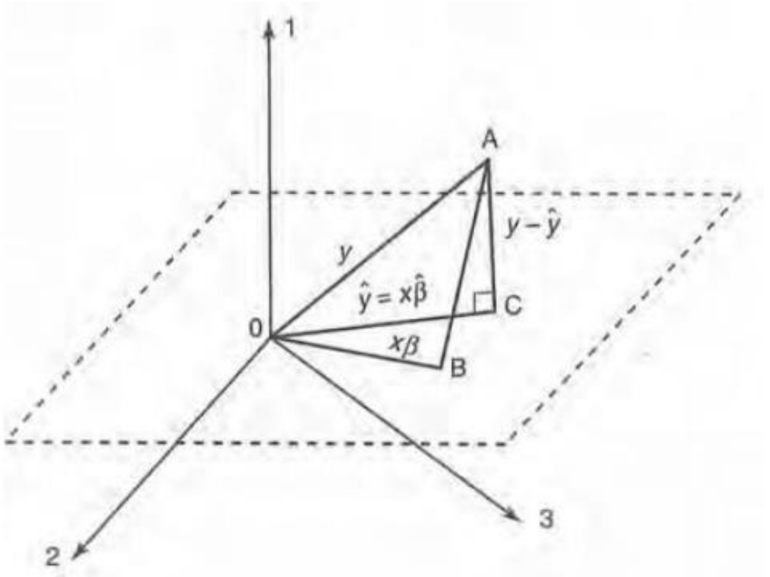
Enfoque bayesiano

- Métodos bayesianos

Estimación de parámetros: Mínimos Cuadrados Ordinarios (MCO)

El método de mínimos cuadrados tiene por finalidad, obtener estimaciones de los parámetros del modelo de regresión que minimicen la suma de cuadrados de los errores del modelo (ε), para ello se construye la función objetivo

$$s(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon`\varepsilon = (y - X\beta)'(y - X\beta)$$



Al proyectar el vector $\boldsymbol{y}$ sobre el subespacio vectorial formado por las variables explicativas, **S**, se busca la distancia más corta del vector al subespacio **S**. En este caso la distancia más corta es de **A** a **C** y corresponde a la proyección ortogonal de $\boldsymbol{y}$ sobre **S**

Estimación de parámetros: Mínimos Cuadrados Ordinarios (MCO)

El estimador de mínimos cuadrados tiene como objetivo encontrar el vector de parámetros que de un hiperplano que pase lo más cerca posible de la nube de puntos producida por los datos, es decir, se trata de encontrar el vector $\beta_{(k+1)\times 1}$ que minimice la distancia (euclidiana) entre los datos observados y el modelo propuesto, esto lleva a:

$$\underset{\beta \in \mathbb{R}^p}{Min}\ \varepsilon`\varepsilon = \underset{\beta \in \mathbb{R}^p}{Min}\ (y - X\beta)`(y - X\beta)$$

Utilizando el cálculo, podemos encontrar puntos críticos (vectores β) y luego averiguar cuál minimiza la distancia entre el vector de observaciones y el modelo propuesto

Estimación de parámetros: Mínimos Cuadrados Ordinarios (MCO)

Estimación de β mediante el método MCO

Dado un modelo lineal general

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

El método Mínimos Cuadrados Ordinarios consiste en buscar aquel β que minimize $\sum_{i=1}^{n} e_i^2$, i.e.

se obtiene que $\hat{\beta}_{MCO}$ forma parte de un **sistema de ecuaciones normales**:

$$\mathbf{X}'\mathbf{X}\hat{\beta}_{MCO} = \mathbf{X}'\mathbf{Y}$$

Desde que $\mathbf{X}$ es una matriz de rango completo se concluye que

$$\hat{\beta}_{MCO} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) = y'y - y'X\beta - \beta'X'y + \beta'X'X\beta$$

~~$y'X\beta$, $\beta'X'y$ =~~

$$= \underbrace{y'y - y'X\beta - \beta'X'y + \beta'X'X\beta = y'y - 2\beta'X'y + \beta'X'X\beta}$$

$$\frac{\partial\varepsilon'\varepsilon}{\partial\beta} = \frac{\partial}{\partial\beta}\left[\quad\right]_{\beta=\hat\beta} = 0$$

$$= \frac{\partial}{\partial\beta}[y'y] - \frac{\partial}{\partial\beta}[2\beta'X'y] \ldots\ldots = 0$$

$$= 0 + 2X'y - 2X'X\hat\beta = 0$$

$$2X'y - 2X'X\hat\beta = 0$$

$$X'y = X'X\beta$$

$X_{n \times (k+1)} \rightarrow$ R.C $\quad X'X$ sería de Rango$(k+1 = p)$

$$X'y = X'X\hat\beta$$

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

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

Propiedad de Mínimos Cuadrados Ordinarios (MCO)

- $\hat{\beta}_{MCO}$ es un estimador insesgado de β.
- $Var(\hat{\beta}_{MCO}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.
- Las variables explicativas son ortogonales al vector de residuos MCO, i.e. $\mathbf{X}'\hat{\varepsilon} = \mathbb{O}$.
- El predictor es ortogonal al vector de residuos MCO, i.e. $\hat{\mathbf{Y}}'\hat{\varepsilon} = \mathbb{O}$.
- $\hat{\beta}_{MCO}$ y $\hat{\varepsilon}$ están incorrelacionados.
- La suma de cuadrados residual se puede expresar como $\mathbf{Y}'\mathbf{Y} - \hat{\beta}'_{MCO}\mathbf{X}'\mathbf{Y}$, esto debido a que $\hat{\mathbf{Y}}'\hat{\mathbf{Y}} = \hat{\mathbf{Y}}'\mathbf{Y}$.
- El vector de residuos MCO es una transformación lineal del término
- $\mathbb{E}(\hat{\varepsilon}) = \mathbb{O}$ y $Var(\hat{\varepsilon}) = \sigma^2 M_{\mathbf{X}}$, donde $M_{\mathbf{X}} = \mathbb{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
- Si el modelo posee término independiente, la suma de residuos es cero.

- Linealidad: El estimador MCO de es lineal. La linealidad consiste en poder escribir el estimador como una combinación lineal fija de los valores de la variable endógena.
- Eficiencia: El estimador MCO de β es eficiente. Es decir, tiene varianza mínima dentro de la familia de estimadores lineales e insesgados de β

Propiedad de Mínimos Cuadrados Ordinarios (MCO)

Propiedades del método MCO

- $s^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-p-1}$ es un estimador insesgado para σ^2
- Dadas las siguientes sumas:
  - Suma de Cuadrados Total (SCT) $:= \sum_{i=1}^{n}(y_i - \overline{y})^2$
  - Suma de Cuadrados Explicada (SCE) $:= \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$
  - Suma de Cuadrados Residual (SCR) $:= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

  Si el modelo lineal general posee término independiente, entonces

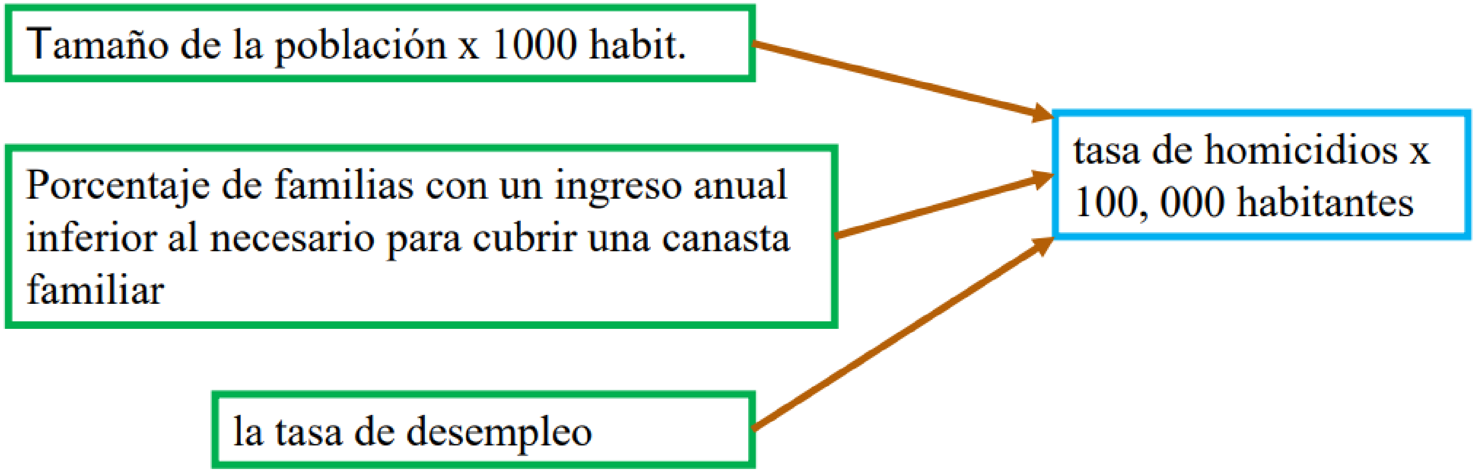
$$SCT = SCE + SCR$$

Teorema de Gauss - Markov

El estimador $\hat{\beta}_{MCO}$ es el mejor estimador lineal insesgado para β, i.e. cualquier otro estimador tiene matriz de covarianzas mayor que la del estimador MCO.

Mínimos Cuadrados Ordinarios (MCO)

Ejemplo.
Un sociólogo está investigando el incremento de la incidencia de homicidios en los últimos años en un país. Él considera que la tasa de homicidios x 100, 000 habitantes está asociada con el tamaño de la población x 1000 habit. el porcentaje de familias con un ingreso anual inferior al necesario para cubrir una canasta familiar y la tasa de desempleo. Los datos corresponden a una muestra de las ciudades más grandes de un país.



Mínimos Cuadrados Ordinarios (MCO)

Ejemplo (cont.):

Supóngase que la tasa de homicidios y el conjunto de variables que el investigador ha considerado se relacionan como:

$$E(Y / X_1 = x_1, X_2 = x_2, X_3 = x_3) = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \beta_3 x_3 \quad .$$

Donde :

Y: Tasa de homicidios (Variable dependiente)

X_1: Tamaño de la población (x1000 habitantes)

X_2: Porcentaje de familias con ingresos bajos (%)

X_3: Tasa de desempleo (%)

$(\beta_0, \beta_1, \beta_2, \beta_3)` \in \Omega_\beta = \Re^4$ son parámetros desconocidos.

Datos

Y	X1	X2	X3
11.2	587	16.5	6.2
13.4	643	20.5	6.4
40.7	635	26.3	9.3
5.3	692	16.5	5.3
24.8	1248	19.2	7.3
12.7	643	16.5	5.9
20.9	1964	20.2	6.4
35.7	1531	21.3	7.6
8.7	713	17.2	4.9
9.6	749	14.3	6.4
14.5	7895	18.1	6
26.9	762	23.1	7.4
15.7	2793	19.1	5.8
36.2	741	24.7	8.6
18.1	625	18.6	6.5
28.9	854	24.9	8.3
14.9	716	17.9	6.7
25.8	921	22.4	8.6
21.7	595	20.2	8.4
25.7	3353	16.9	6.7

Observación.- Se supone que las variables X_1, X_2, X_3 son conocidos y permitirían estimar y pronosticar los valores desconocidos de la variable respuesta, Y (Tasa de homicidios).

Mínimos Cuadrados Ordinarios (MCO)

Ejemplo (cont.):

Se trata de estimar los parámetros del modelo propuesto para explicar la tasa de homicidios y el conjunto de variables que el investigador considera que están relacionadas, utilizando el estimador de mínimos cuadrados.

$$(X`X)=\begin{pmatrix} 20 & 28660 & 394.4 & 138.7 \\ 28660 & 96220758 & 547985.8 & 189636.7 \\ 394.4 & 547985.8 & 7977.3 & 2795.8 \\ 138.7 & 189636.7 & 2795.8 & 989.57 \end{pmatrix}$$

$$(X`X)^{-1}=\begin{pmatrix} 2.333282e+00 & -6.782529e-05 & -6.487808e-02 & -1.307416e-01 \\ -6.782529e-05 & 1.921950e-08 & -7.999924e-07 & 8.083576e-06 \\ -6.487808e-02 & -7.999924e-07 & 1.497456e-02 & -3.306040e-02 \\ -1.307416e-01 & 8.083576e-06 & -3.306040e-02 & 1.111909e-01 \end{pmatrix}$$

$$(X`y)=\begin{pmatrix} 411.4 \\ 568073.5 \\ 8624 \\ 3049.06 \end{pmatrix}$$

$$\hat{\beta}=(X`X)^{-1}(X`y)=\begin{pmatrix} 36.76 \\ 0.00076 \\ 1.19 \\ 4.72 \end{pmatrix}$$

Modelo ajustado

$$\hat{y}_i = -36.76 + 0.00076x_{i1} + 1.19x_{i2} + 4.72x_{i3} \quad ; \quad i=1,2,\dots,20$$

Medida de bonda de ajuste

- El objetivo es buscar un modelo que se ajuste correctamente a los datos y además sea el más sencillo posible.
- Para el caso de la bondad de ajuste, nótese que se obtiene un buen ajuste del modelo cuando los residuales son pequeños.

Coeficiente de Determinación

Dado un modelo lineal general de la forma **Y = Xβ + ε** con término independiente, el coeficiente de determinación se define como

$$R^2 = 1 - \frac{SCR}{SCT} = \frac{SCE}{SCT}$$

Si el modelo no posee término independiente (β_0= 0) entonces se utiliza una modificación del R^2 de la forma

$$R_0^2 = 1 - \frac{SCR}{\sum_{i=1}^{n} y_i^2}$$

Medida de bonda de ajuste

El R^2 no indica:

- Las variables regresoras son una causa de los cambios en la variable dependiente.
- Existe sesgo por variables regresoras omitidas.
- Ha sido escogido el conjunto más apropiado de variables explicativas.
- Hay presencia de colinealidad en los datos, pero sí te da una primera pista.
- El modelo puede ser mejorado por transformaciones en las variables regresoras.
- Hay suficientes observaciones muestrales para dar una conclusión solida

<u>Coeficiente de Determinación ajustado</u>

- El coeficiente de determinación ajustado es una generalización del R^2 tradicional. Este indicador de bondad de ajuste además de tomar en cuenta el ajuste del modelo penaliza el uso de más variables regresoras.

$$R^2_{adj} = 1 - \frac{SCR/(n-p-1)}{SCT/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

- El R^2_{adj} es una medida para comparar modelos. Además el $R^2_{adj} \leq R^2$

Modelo lineal general

Ejercicio:

- Pruebe que dado un modelo lineal general con intercepto de la forma Y = Xβ + ε, entonces la suma de cuadrados explicada se puede representar como

$$SCE = \hat{\boldsymbol{\beta}}_{\mathbf{1}} X_c' X_c \hat{\boldsymbol{\beta}}_{\mathbf{1}} = \hat{\boldsymbol{\beta}}_{\mathbf{1}} X_c' \mathbf{Y}$$

- Pruebe que la suma de cuadrados residual de un modelo lineal general no cambia si usamos la versión centrada del mismo modelo.

- Sea el modelo lineal general Y = Xβ + ε con p variables explicativas y n observaciones. Suponiendo que ε posee distribución normal multivariada, halle la distribución de

$$W = \frac{\hat{\boldsymbol{\beta}}_{\mathbf{1}} X_c' Y/p}{(Y'(\mathbb{I}_n - \frac{1}{n}\mathbb{J}_n)Y - \hat{\boldsymbol{\beta}}_{\mathbf{1}} X_c' Y)/(n-p-1)}$$

- Pruebe que si agregamos una nueva variable explicativa a un modelo lineal general, entonces el R^2 mejora. ¿Que ocurriría con el $R^2_{adj} R2\ adj$?

Modelo lineal general

Ejercicio:

- Para el modelo Y = β0 + β1X + β2X + e se tienen los siguientes datos

$$n = 12, \qquad SCT = 104'9167,$$

$$(X^tX)^{-1} = \begin{pmatrix} 0'6477 & -0'041 & -0'0639 \\ -0'041 & 0'0071 & -0'0011 \\ -0'0639 & -0'0011 & 0'0152 \end{pmatrix}, \qquad X^tY = \begin{pmatrix} 91 \\ 699 \\ 448 \end{pmatrix}$$

a) Ajustar el modelo por el método de MCO y calcular el coeficiente de determinación.

- En un estudio de los determinantes de la inversión se usaron 20 datos anuales, correspondientes a las siguientes variables: inversión anual en millones de soles (Y), tipo de interés en porcentaje (X1) y variación anual de PIB en millones de soles (X2). Se dispone de la siguiente información:

-

$$\begin{array}{lll} \sum X_{1t} = 100 & \sum X_{2t} = 24 & \sum Y_t = 5 \\ \sum X_{1t}Y_t = -255 & \sum X_{2t}Y_t = 146 & \sum X_{1t}X_{2t} = 100 \\ \sum X_{1t}^2 = 680 & \sum X_{2t}^2 = 48'8 & \sum \left(Y_t - \overline{Y}\right)^2 = 1200 \end{array}$$

Modelo lineal clasico

El modelo lineal clásico es aquel modelo lineal que trabaja con los mismos supuestos del modelo lineal general pero ahora se agrega el supuesto de que los términos de perturbación aleatoria se distribuyen normalmente por cada observación, i.e. $e_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ Entonces podemos plantear el modelo lineal clásico de la forma matricial

$$Y = X\beta + \varepsilon$$

bajo los supuestos de que:

- $\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbb{I}_n)$
- X es no estocástica y $ran(X) = p + 1$

Observación: Note que si $\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbb{I}_n)$ entonces $Y \sim \mathcal{N}_n(X\beta, \sigma^2 \mathbb{I}_n)$.

$$f(Y|\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\}$$

CASO PARTICULAR
MODELO DE REGRESIÓN LINEAL SIMPLE

El modelo de regresión lineal, por lo tanto se puede expresar como:

$$E / Y / X = x) = \beta_0 + \beta_1 x \qquad ; \; (\beta_0, \beta_1) \in \mathbb{R}^2$$

el cual corresponde a la ecuación de una recta donde β_0 es el intercepto y β_1 es la pendiente.

En el contexto del análisis de regresión las constantes son denominadas ***coeficientes de regresión* parciales** y tienen una interpretación práctica.
Por ejemplo en la regresión lineal:
β_1: Se interpreta como el cambio en la media de Y producido por un cambio de una unidad en X.
β_0: Se interpreta como la media de Y cuando X=0. si la variable X no toma valor cero, este coeficiente no tiene interpretación.

Observación.- El intercepto se interpreta como respuesta media, únicamente, en el caso que las variables X1 ,X2 ,…,Xk tengan media cero. En otro caso, no son interpretables.

CASO PARTICULAR
MODELO DE REGRESIÓN LINEAL SIMPLE

Supóngase que, se selecciona una muestra aleatoria de tamaño n con elementos, {(xi ,yi): i=1,2,…,n}, extraída de una población en la cual las variables X e Y se relacionan de acuerdo al modelo de regresión lineal, por lo tanto, para cualquier observación de la muestra se verifica que

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad ; i=1,2,\ldots,n$$

Supuestos:

a) $E(\varepsilon_i)=0$

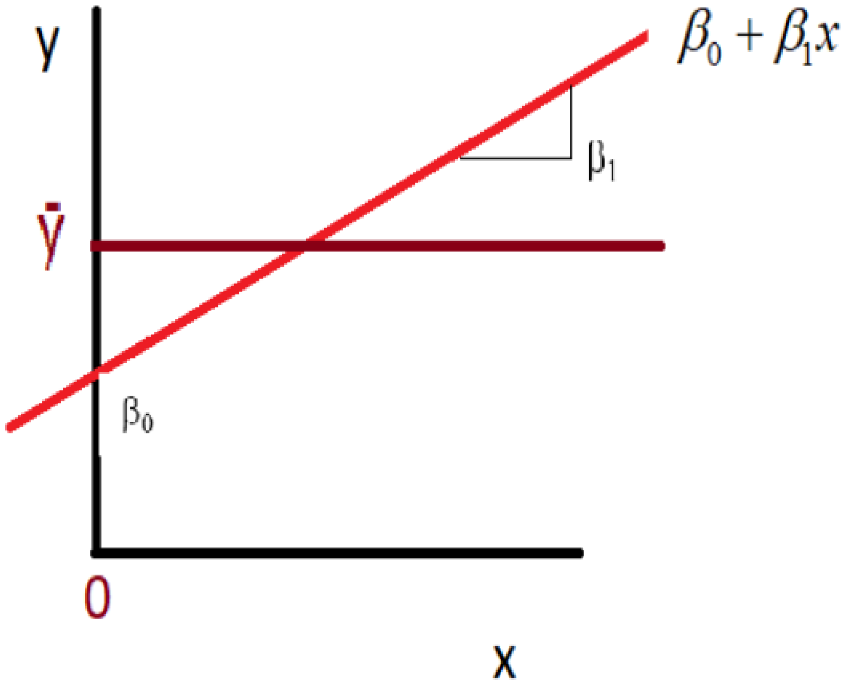
b) $V(\varepsilon_i)=\sigma^2$; constante (homocedasticidad)

c) $Cov(\varepsilon_k\ \varepsilon_m)=0$ (Errores incorrelacionados

d) $\varepsilon_i \sim N(0, \sigma^2)$ (Normalidad de los errores)

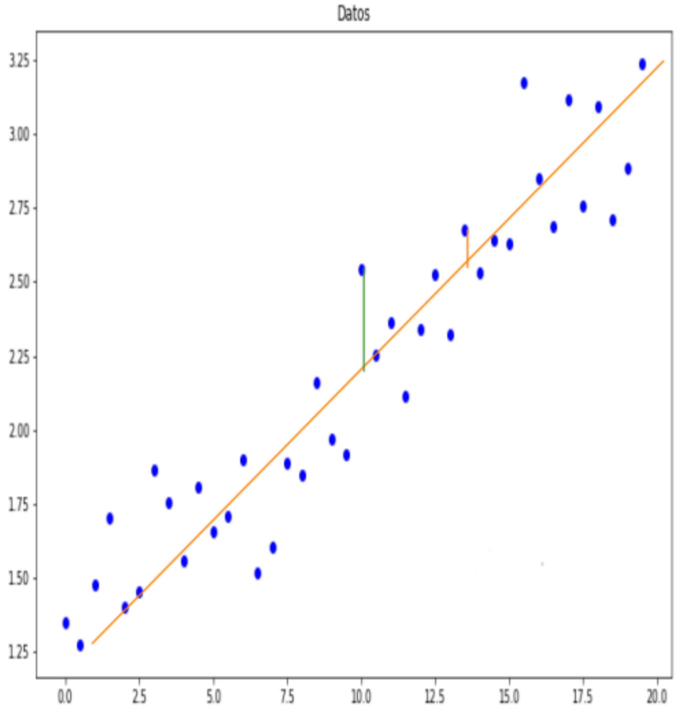
CASO PARTICULAR
MODELO DE REGRESIÓN LINEAL SIMPLE

Modelo



DATOS OBSERVADOS (x, y)

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad ; i=1,2,\ldots,n$$



MÉTODO DE ESTIMACIÓN DE MÍNIMOS CUADRADOS ORDINARIOS (Modelo lineal simple)

Modelo propuesto

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Error aleatorio

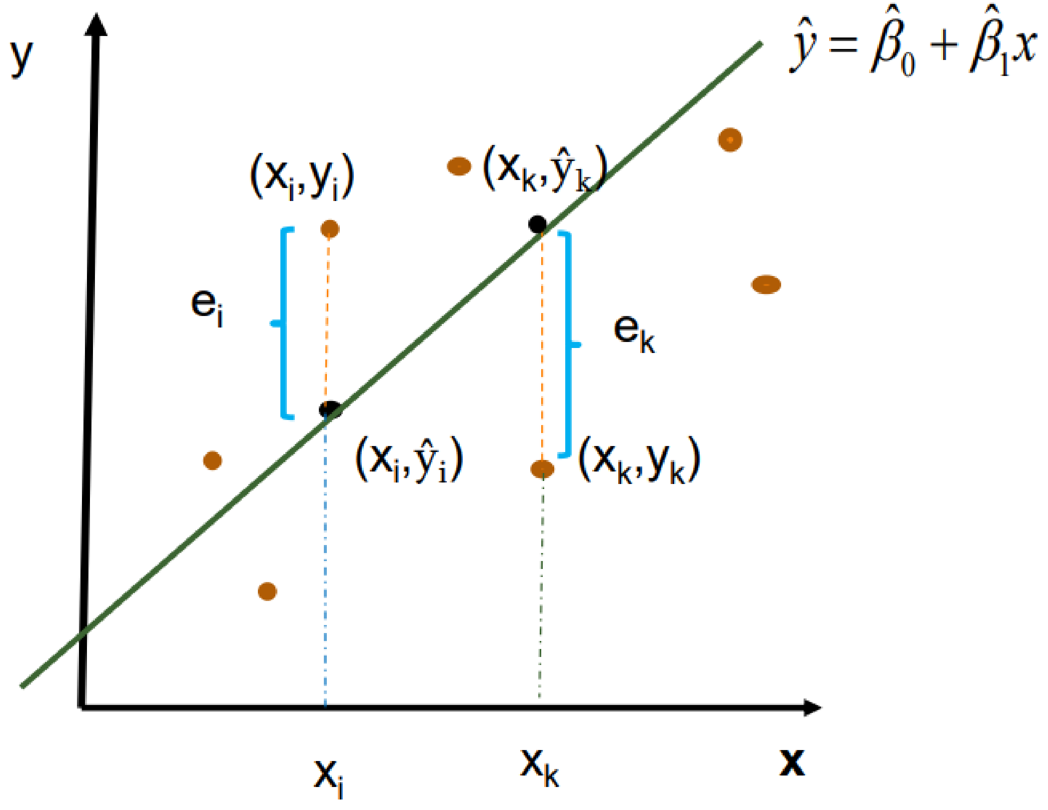
$$\varepsilon_i = y_i - \beta_0 + \beta_1 x_i$$

Modelo ajustado

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Residuo

$$e_i = \hat{y}_i - \hat{\beta}_0 + \hat{\beta}_1 x_i$$



MÉTODO DE ESTIMACIÓN DE MÍNIMOS CUADRADOS ORDINARIOS (Modelo lineal simple)

La función S(β) es continua y derivable en el vector β=($\beta 0$, $\beta 1$), por lo que podemos derivar e igualar a cero y obtener puntos críticos, y finalmente demostrar que :

$$\hat{\beta}_1 = \frac{Cov(x,y)}{S_x^2}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Observación: La recta de regresión pasa siempre por el centro de gravedad de la nube de puntos, es decir por el punto $(\overline{x}, \overline{y})$

ESTIMACIÓN DE LA VARIANZA DEL ERROR

Si se considera que los residuos corresponden a una realización de los errores del modelo podemos estimar la varianza σ^2 como:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(e_i - \bar{e})^2}{n-p} = \frac{\sum_{i=1}^{n} e_i^2}{n-p}$$

Expresándolo matricialmente :

$$\hat{\sigma}^2 = \frac{e`e}{n-p} = \frac{[(\text{I-H})\text{y}]`[(\text{I-H})\text{y}]}{n-p} = \frac{y`(I-H)`(I-H)y}{n-p}$$

$$\hat{\sigma}^2 = \frac{e`e}{n-p} = \frac{y`(I-H)y}{n-p}$$

La varianza estimada de los errores, se obtiene a partir de los residuos del ajuste.

MÉTODO DE ESTIMACIÓN MÁXIMO VEROSIMIL

El método de estimación máximo verosimilitud trata de encontrar el vector de parámetros θ que maximiza la función de verosimilitud L(θ) ($\theta \in \Theta$), de modo que

$$L(\hat{\theta}) \geq L(\theta) \quad \text{para todo} \quad \theta \in \Theta$$

Observaciones.-

(1) La función de verosimilitud es una función del parámetro y no de la variable aleatoria.

(2) La función de verosimilitud no es una función de probabilidad.

(3) Puede ocurrir que la función de verosimilitud no tenga un único máximo y en algunos casos, puede no ser finita incluso podría no existir

MÉTODO DE ESTIMACIÓN MÁXIMO VEROSIMIL

Para obtener estimadores máximos verosímiles es necesario conocer la distribución de probabilidades de los errores del modelo de regresión lineal dado por

$$y_{nx1} = X_{nx(k+1)}\beta_{(k+1)x1} + \varepsilon_{nx1}$$

Supuestos (expresados en términos de Y):

1) $E(\mathbf{Y})=\mu=\mathbf{X}\boldsymbol{\beta}$
2) $V(\mathbf{Y})=\sigma^2\mathbf{I}$
3) $Cov(\mathbf{Y})=\mathbf{0}$
4) **$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$**

Es decir,

$$f_Y(y\ ;\ \mu,\sigma^2 I) = \prod_{i=1}^{n} f_{Y_i}(y_i;\ \mu_i,\sigma^2)$$

donde

$f_{Y_i}(y_i;\ \mu_i,\sigma^2)$ Es la función de densidad de una distribución normal.

MÉTODO DE ESTIMACIÓN MÁXIMO VEROSIMIL

Supóngase una muestra aleatoria, de tamaño n , $\{(y_i, x_{i1}, \ldots, x_{ik})$: i=1,2,...,n} podemos formular la función de verosimilitud como

$$L(\beta, \sigma^2 : y; X) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[\frac{1}{2\sigma^2}(y - X\beta)`(y - X\beta)\right]$$

$$\hat{\beta} = (X`X)^{-1} X`y$$ Siempre que $(X`X)^{-1}$ exista.

Es el estimador máximo verosímil, de los coeficientes del modelo de regresión

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})`(y - X\hat{\beta})}{n}$$

Es el estimador máximo verosímil de la varianza de los errores

MÉTODO DE ESTIMACIÓN MÁXIMO VEROSIMIL

El siguiente paso es verificar que $\hat{\beta}$ y $\hat{\sigma}^2$ son los estimadores máximos verosímiles

de los parámetros del modelo de regresión.

El siguiente paso es evaluar las propiedades de estos estimadores y determinar sus funciones de distribución

1. Se garantiza que el estimador $\hat{\beta}$ es suficiente, pues la distribución normal es miembro de la familia exponencial

2. Insesgamiento

$$E(\hat{\beta}) = E((X`X)^{-1}X`y) = (X`X)^{-1}X`E(y)$$
$$= (X`X)^{-1}X`X\beta$$
$$E(\hat{\beta}) = \beta$$

Luego, el estimador es insesgado.

MÉTODO DE ESTIMACIÓN MÁXIMO VEROSIMIL

3. Varianza del estimador de β

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

4. Distribución del estimador máximo verosímil (mínimo cuadrático) del vector de parámetros, β.

El estimador $\hat{\beta} = (X'X)^{-1}X'y = Ay$ es una forma lineal con matriz $\mathrm{A} = (X'X)^{-1}X'y$ bajo el supuesto que, el vector, **y**, tiene distribución normal, entonces A**y** tiene distribución normal con esperanza, β y varianza $(X'X)^{-1}\,\sigma^2$

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

MÉTODO DE ESTIMACIÓN MÁXIMO VEROSIMIL

Ahora veamos las propiedades del estimador de σ^2 , la varianza del error

1. Se garantiza que el estimador $\hat{\sigma}^2$ es suficiente, puesto que, la distribución normal es miembro de la familia exponencial.
2. Insesgamiento

$$E(n\bar{\sigma}^2) = E\left((y`(I-H)y\right)$$

Bajo el supuesto que y~N(Xβ, Iσ²), la esperanza de la forma cuadrática y`(I-H)y es dado por :

$$E(y`(I-H)y) = \left[Traza(\sigma^2 I.(I-H) + \left(X\beta\right)`(I-H)(X\beta)\right]$$

$$= \left[(\text{n-p})\sigma^2 + (\beta`X`IX\beta - \beta`X`HX\beta\right]$$

$$= \left[(\text{n-p})\sigma^2 + (\text{y`y- y`y})\right]$$

$$E(y`(I-H)y) = (n-p)\sigma^2$$

El estimador máximo verosímil es sesgado

MÉTODO DE ESTIMACIÓN MÁXIMO VEROSIMIL

3. Distribución de probabilidades de $\hat{\sigma}^2$

Bajo el supuesto que y~N(Xβ, Iσ²), la distribución de la forma cuadrática y\`(I-H)y es dado por :

$$\frac{y`(I-H)y}{\sigma^2} \sim \chi^2_{n-p}$$

Rango(I-H)=n-p=n+(k+1) y parámetro de no centralidad

$$\lambda = \frac{1}{2\sigma^2}(X\beta)`((I-H)(X\beta) = 0$$

Esto se verifica porque I(I-H)= I-H es idempotente (AV idempotente).

PROPIEDADES DE LA RESPUESTA MEDIA ESTIMADA

La respuesta media estimada es dada por :

$$\hat{y} = X\hat{\beta}$$

1. Insesgamiento

$$E(\hat{y}) = E(X\hat{\beta}) = XE(\hat{\beta})$$
$$= X\beta$$

El estimador de la respuesta media es insesgado.

2. Varianza de la respuesta media estimada para un vector $x_0^t 0 = (1, x_{01}, x_{02}, ..., x_{0k})$

$$V(\hat{y}_0) = V(x_0^t \hat{\beta}) = V(x_0^t (X`X)^{-1} X`y)$$
$$= x_0^t (X`X)^{-1} X`V(Y)(x_0^t (X`X)^{-1} X`)`$$
$$= \sigma^2 x_0^t (X`X)^{-1} x_0$$

PROPIEDADES DE LA RESPUESTA MEDIA ESTIMADA

3. Bajo el supuesto que y~N(**X**β, σ²**I**) , La distribución de la respuesta media estimada dada una observación $\mathbf{x}_{nx(k+1)}$ es normal.

$$\hat{y}_0 \sim N(x_0^t \beta,\ x_0^t (X`X)^{-1} x_0 \sigma^2)$$

Sea **Y** = **X**β + ε; donde ε~N(**0**,σ²**I**), los siguientes resultados se verifican

1. $\hat{\beta} = (X`X)^{-1} X`y$ es el estimador máximo verosímil para β.
2. $\hat{\sigma}^2 = \frac{y`(I-H)y}{n-p}$ es el estimador máximo verosímil (corregido), insesgado para σ^2
3. $\hat{\beta}$ Se distribuye como una N(β, Cσ²), donde $C=(X`X)^{-1}$
4. $\frac{(n-p)\hat{\sigma}^2}{\sigma^2}$ Se distribuye como una χ^2_{n-p}
5. $\hat{\beta}$ y $\hat{\sigma}^2$ Son independientes
6. $\hat{\beta}$ y $\hat{\sigma}^2$ Son estadísticas suficientes para **β** y σ²
7. $\hat{\beta}$ y $\hat{\sigma}^2$ Son estadísticas completas

Propiedades estadísticas en un Modelo lineal clásico

- Los estimadores para β obtenidos mediante el método MCO y MV son los mismos.

- Los estimadores para $\hat{\sigma}^2$ obtenidos mediante el método MCO y MV son distintos.
  De hecho $\hat{\sigma}^2_{MV} = \frac{n-p-1}{n}\hat{\sigma}^2_{MCO}$.
- $\hat{\beta}_{MV} = \hat{\beta}_{MCO} \sim \mathcal{N}_{p+1}(\beta, \sigma^2(X'X)^{-1})$.

- Los estimadores $\hat{\beta}_{MCO}$ y $\hat{\sigma}^2_{MCO}$ son independientes. También lo son $\hat{\beta}_{MV}$ y $\hat{\sigma}^2_{MV}$.

- $\frac{n-p-1}{\sigma^2}\hat{\sigma}^2_{MCO} = \frac{n}{\sigma^2}\hat{\sigma}^2_{MV} \sim \chi^2(n-p-1)$.
- $\hat{\sigma}^2_{MV}$ es una estimador más eficiente para σ^2 que $\hat{\sigma}^2_{MCO}$.
- $\hat{\sigma}^2_{MV}$ es un estimador sesgado para σ^2; sin embargo, es un estimador asintóticamente insesgado para σ^2.

DESCOMPOSICIÓN DE LA VARIACIÓN TOTAL

Si solamente consideramos el numerador de la varianza, obtenemos la variación total de los datos (SCT o SST)

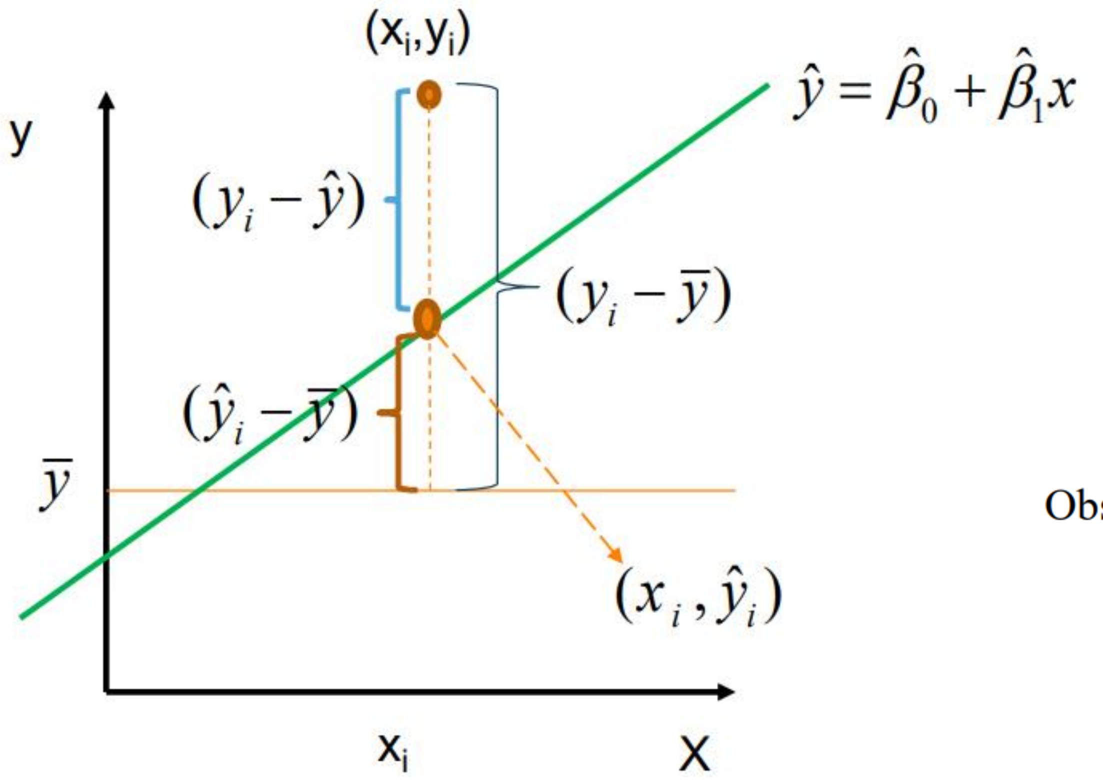
$$SCT = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

De mismo modo, podríamos encontrar la variación explicada por el modelo (SCR) y la variación no explicada (SCE) de modo que la variación total pueda descomponerse en :

$$SCT = SCR + SCE$$

DESCOMPOSICIÓN DE LA VARIACIÓN TOTAL

Gráficamente:



Observar que la distancia de cada observación y_i se descompone en :

$$(y_i - \overline{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})$$

DESCOMPOSICIÓN DE LA VARIACIÓN TOTAL

Por lo tanto, podemos particionar la variación total de los datos correspondientes a la variable respuesta, en dos componentes:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

$$\text{SCT} \quad = \quad \text{SCR} \quad + \quad \text{SCE}$$

Esto es, una porción de la variación total de los datos de la variable respuesta (SCT) se explica por el modelo ajustado y se denomina ***suma de cuadrados de la regresión*** (SCR) y la porción restante es la que queda sin explicar y se denomina ***Suma de cuadrados residual o suma de cuadrados del error*** (SCE)

DESCOMPOSICIÓN DE LA VARIACIÓN TOTAL

Expresando estos resultados matricialmente se tiene:

$$(y-\overline{y})`(y-\overline{y}) = (\hat{y}-\overline{y})`(\hat{y}-\overline{y}) + (y-\hat{y})`(y-\hat{y})$$

$$y`y - n\,\overline{y}^2 = \left(\hat{\beta}`X`y - n\,\overline{y}^2\right) + (y-\hat{y})`(y-\hat{y})$$

$$y`y = \hat{\beta}`X`y + (y-\hat{y})`(y-\hat{y})$$

De aquí se tiene:

$$\text{SCT} = y`y - n\,\overline{y}^2$$

$$\text{SCR} = \hat{\beta}`X`y - n\,\overline{y}^2$$

Para calcular la variación residual consideremos

$$\text{SCE} = y`y - n\,\overline{y}^2 - \left(\hat{\beta}`X`y - n\,\overline{y}^2\right)$$

$$SCE = y`y - \hat{\beta}`X`y$$

Inferencia acerca de los estimadores

Intervalo de confianza

Dicha distribución es simétrica presentando mayor dispersión que la curva Normal estándar para un tamaño muestral ”n” pequeño. A medida que ”n” aumenta (n>100) es prácticamente igual que la distribución Normal.

$$\hat{\beta}_i \pm S_{\hat{\beta}_i} t_{n-k}$$

donde $S_{\hat{\beta}_i}$ es la desviación típica estimada para el coeficiente, que se obtiene de la matriz de varianzas y covarianzas de los estimadores MCO $\hat{Var}(\hat{\beta}) = (X'X)^{-1}\hat{\sigma}^2$, calculado $\hat{\sigma}^2$ a partir de:

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

Inferencia acerca de los estimadores

Contraste de hipótesis sobre los coeficientes

Cuando se tiene un modelo de regresión con *k* variables en la matriz de diseño *X* (la primera columna de X son unos y no se cuenta como variable), es usual que nos interese estudiar

$$H_0 : \beta_j = \beta_{j0},$$

frente a una de las tres siguientes hipótesis alternas:

$$H_A : \beta_j < \beta_{j0}, \quad H_A : \beta_j \neq \beta_{j0}, \quad H_A : \beta_j > \beta_{j0},$$

para algún j=0,1,2,…,k

Para estas pruebas el estadístico de prueba está dado por

$$t_0 = \frac{\hat{\beta}_j - \beta_{j0}}{s.e.(\hat{\beta}_j)},$$

y bajo la hipótesis nula cierta, $t_0 \sim t_{n-k-1}$.