UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL DE HUAMANGA

ESTIMACIÓN POR MÍNIMOS CUADRADOS ORDINARIOS (MCO)

Econometría I

Prof. William Canales Molina.

Material estrictamente académico

1. METODOLOGÍA DE ESTIMACIÓN MCO

- Dados el modelo de regresión poblacional y el estimado:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i = b_1 + b_2 X_i$$

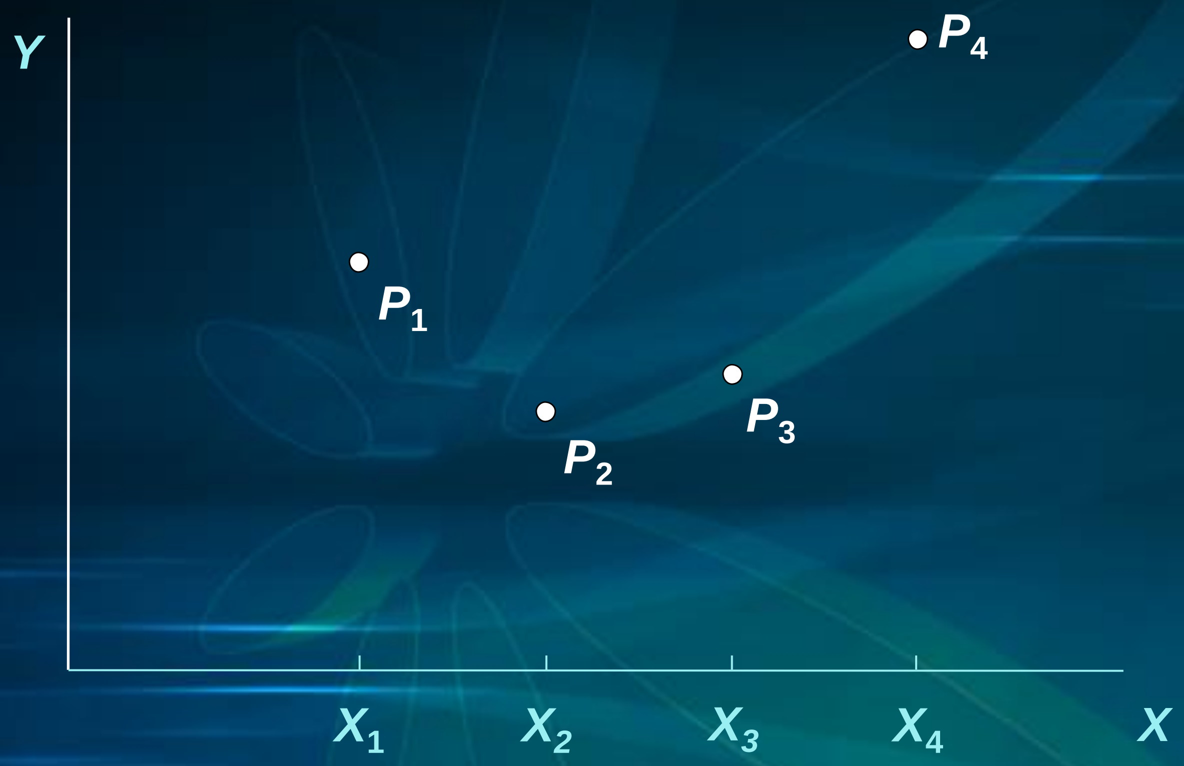
se definen los *residuos* de cualquier línea ajustada como:

$$e_i = Y_i - \hat{Y}_i = Y_i - (b_1 + b_2 X_i)$$

- Cada valor observado de la variable endógena puede expresarse como:

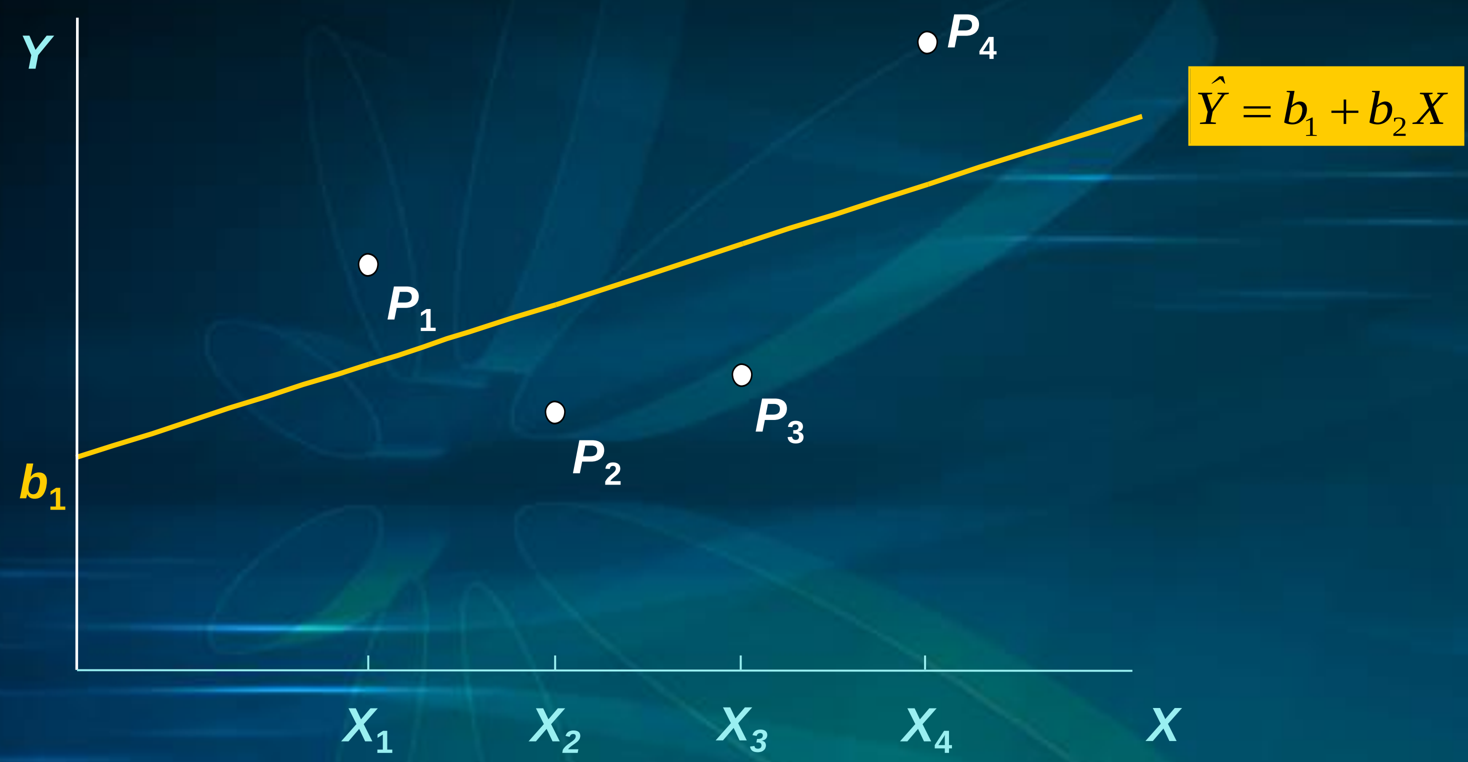
$$Y_i = \hat{Y}_i + e_i$$
$$Y_i = b_1 + b_2 X_i + e_i$$

Continuación…



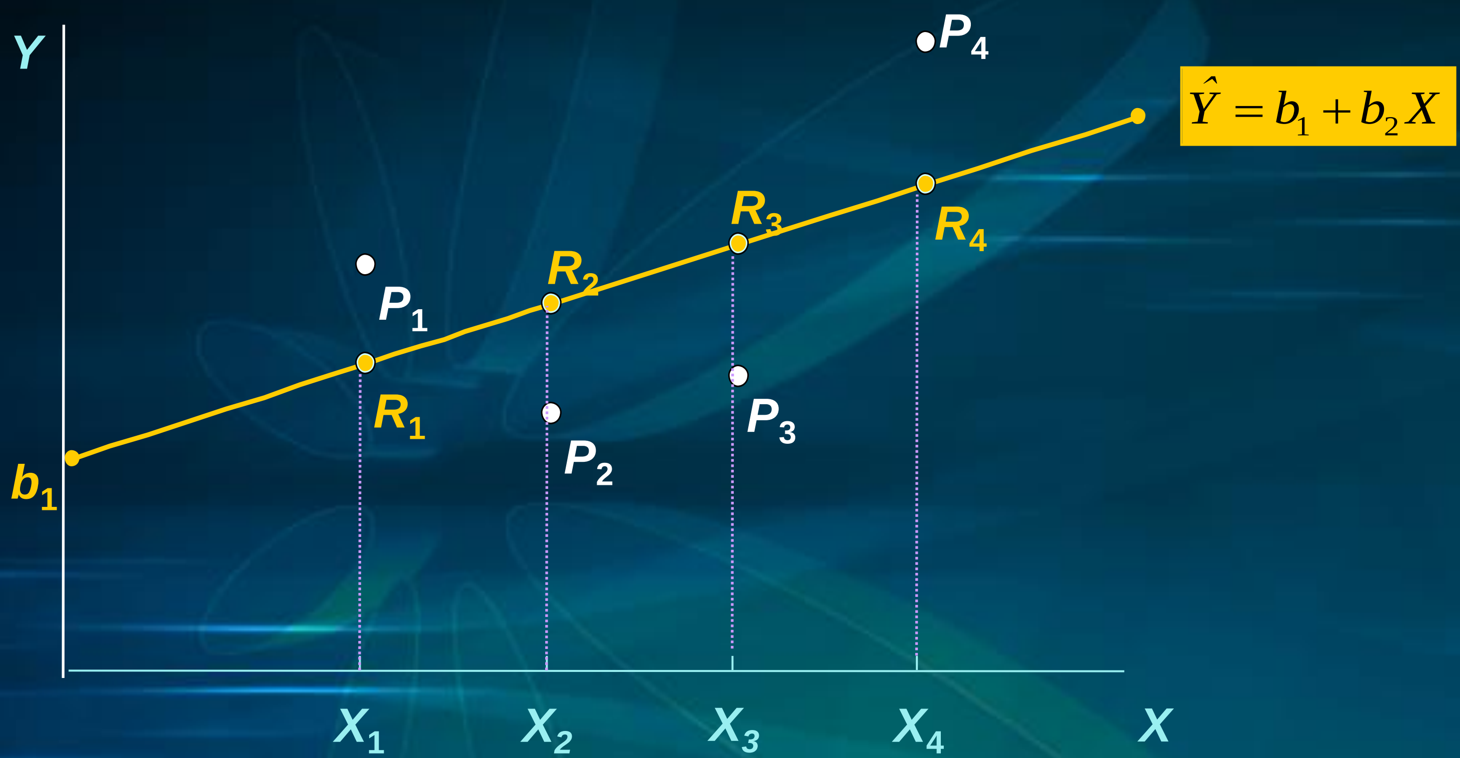
En la práctica sólo observamos los puntos P.

Continuación…



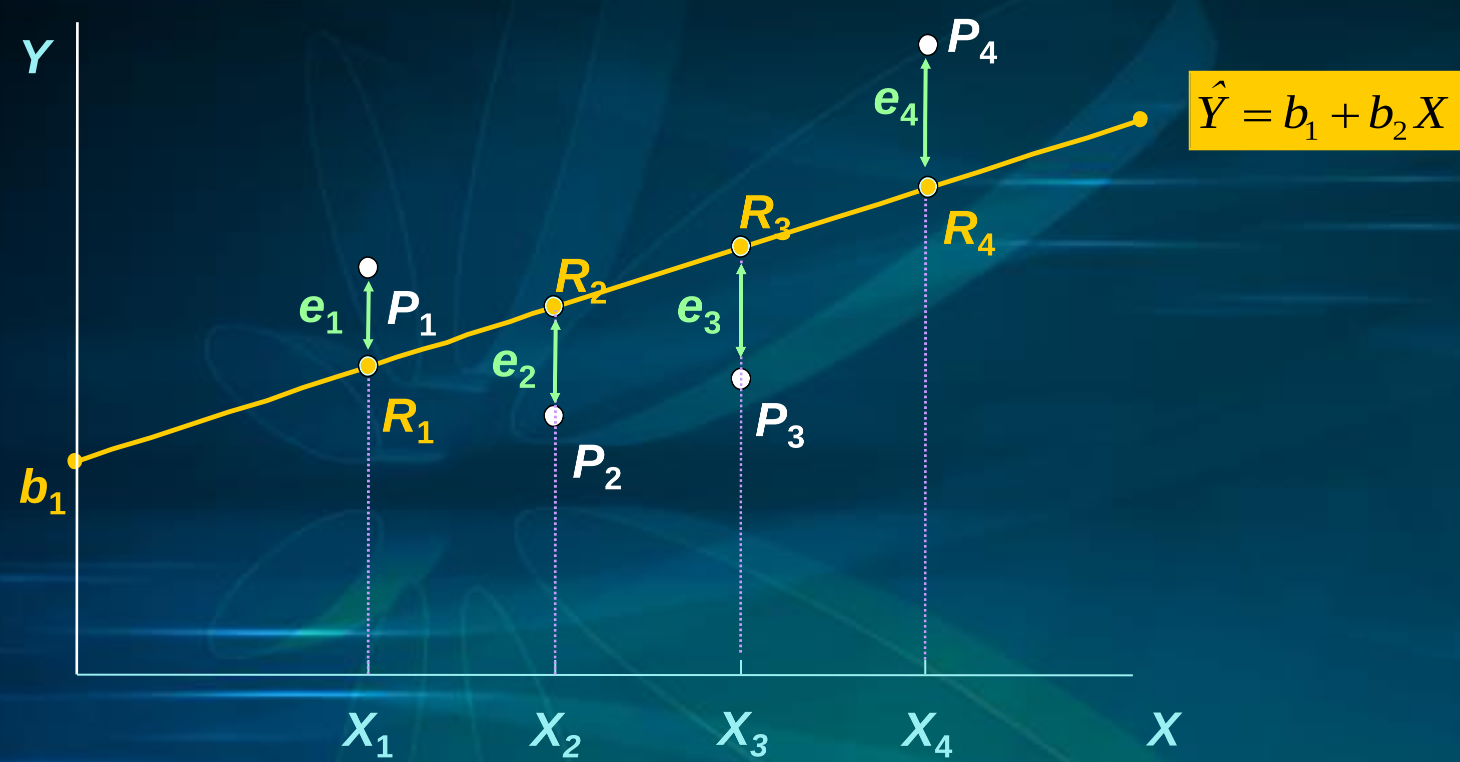
Podemos usar los puntos P para dibujar una línea recta que sea una aproximación de $Y = \beta_1 + \beta_2 X$. Esta línea estimada será $Y = b_1 + b_2 X$, donde b_1 es un estimador de β_1 y b_2 es un estimador de β_2.

Continuación…



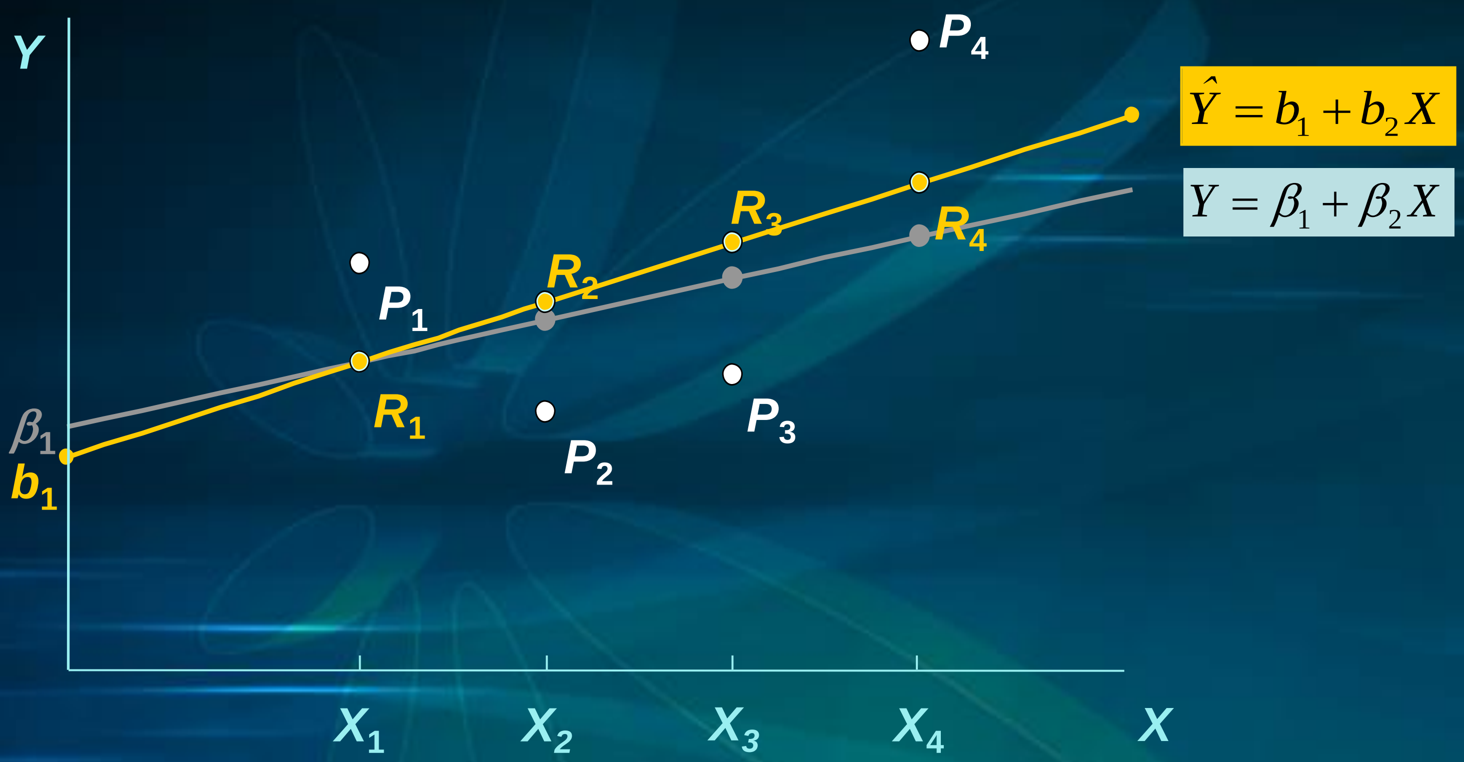
La linea se llama modelo ajustado y los valores de Y estimados se llaman valores ajustados de Y. Se representan por la altura de los puntos R.

Continuación…



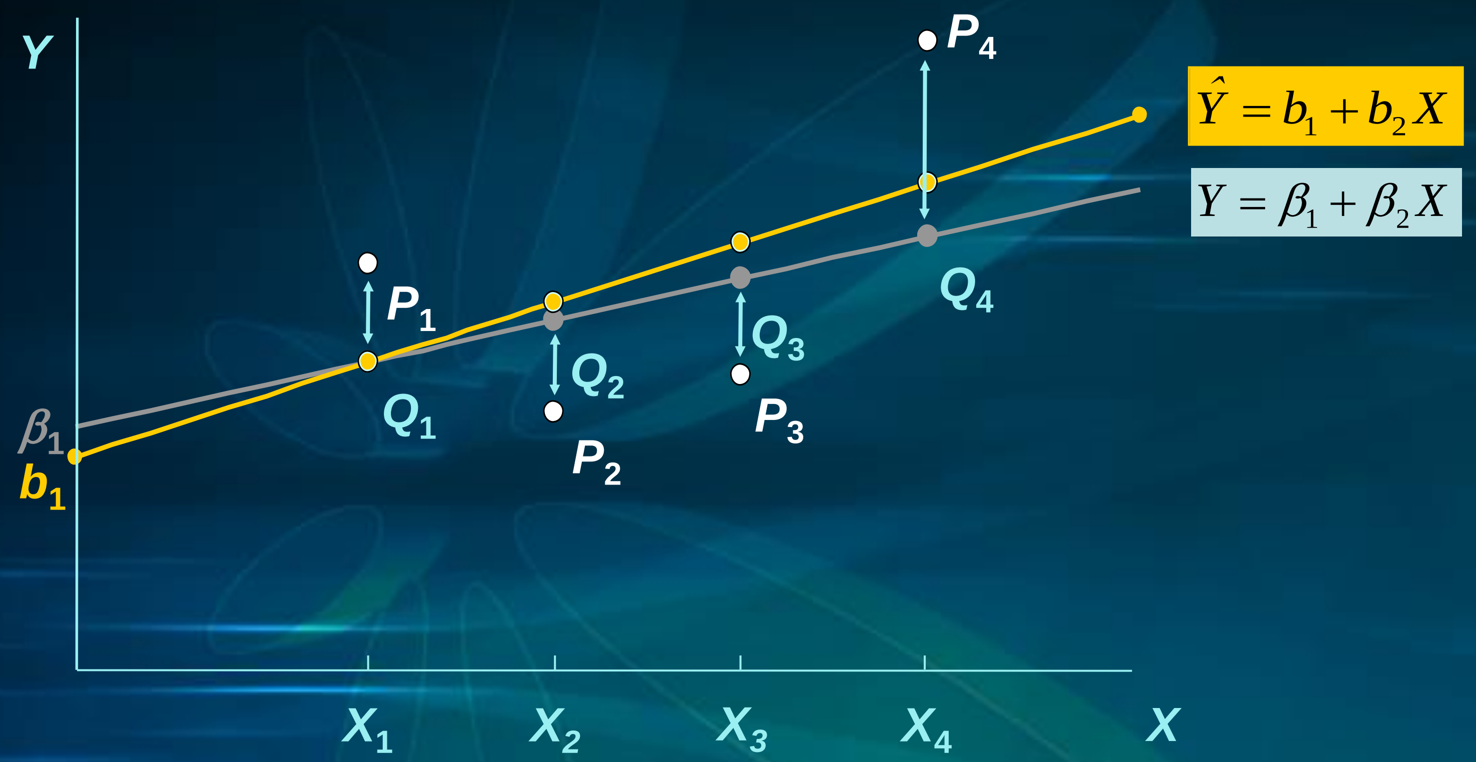
La discrepancia entre el valor observado y el valor ajustado de Y se conoce como residuo.

Continuación…



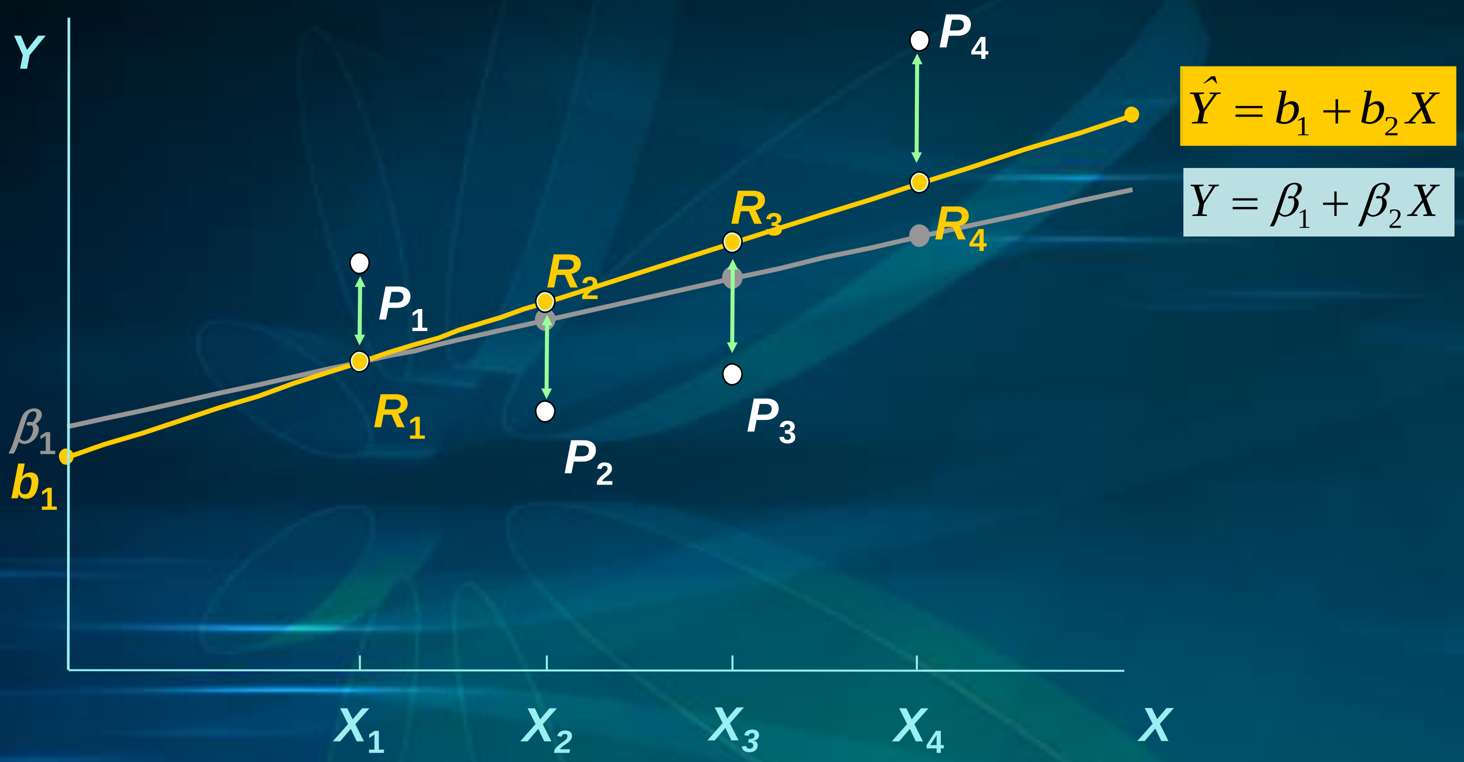
<u>Importante</u>: Note que el residuo no es igual al término de perturbación.

Continuación…



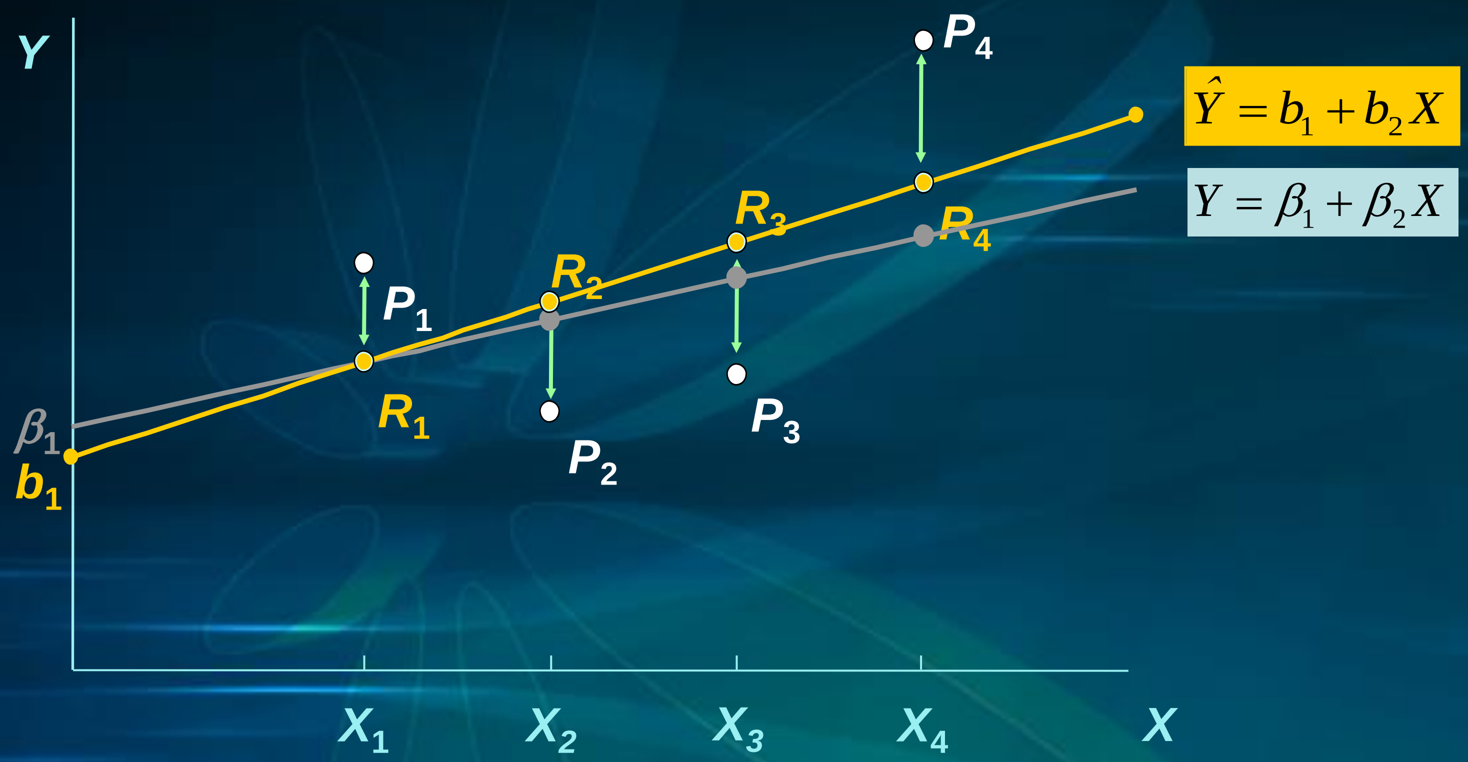
El término de perturbación en cada observación es el responsable de la divergencia entre el valor observado y el componente no aleatorio de la verdadera relación.

Continuación…



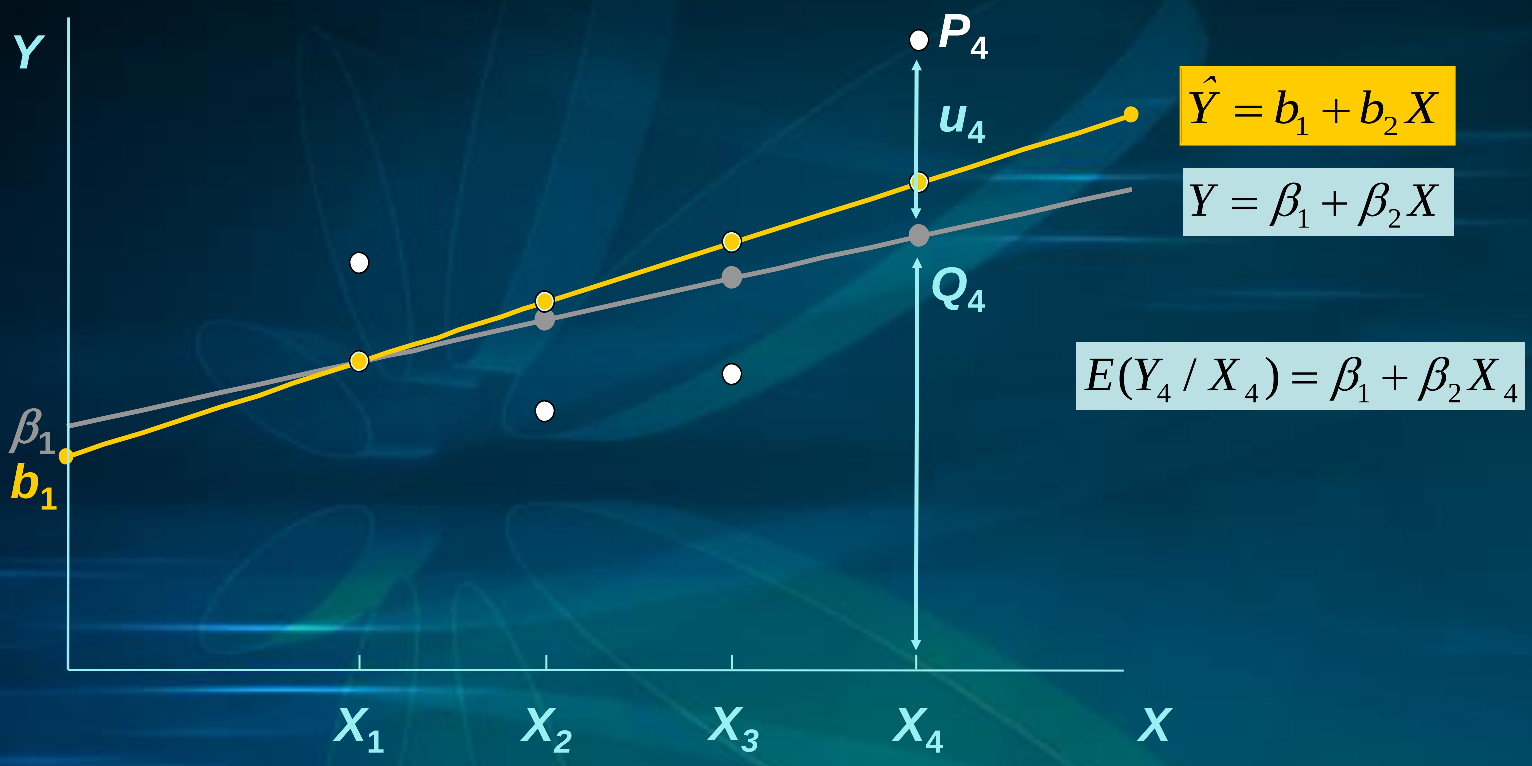
El residuo es la discrepancia entre el valor observado y el valor ajustado de la variable endógena.

Continuación…

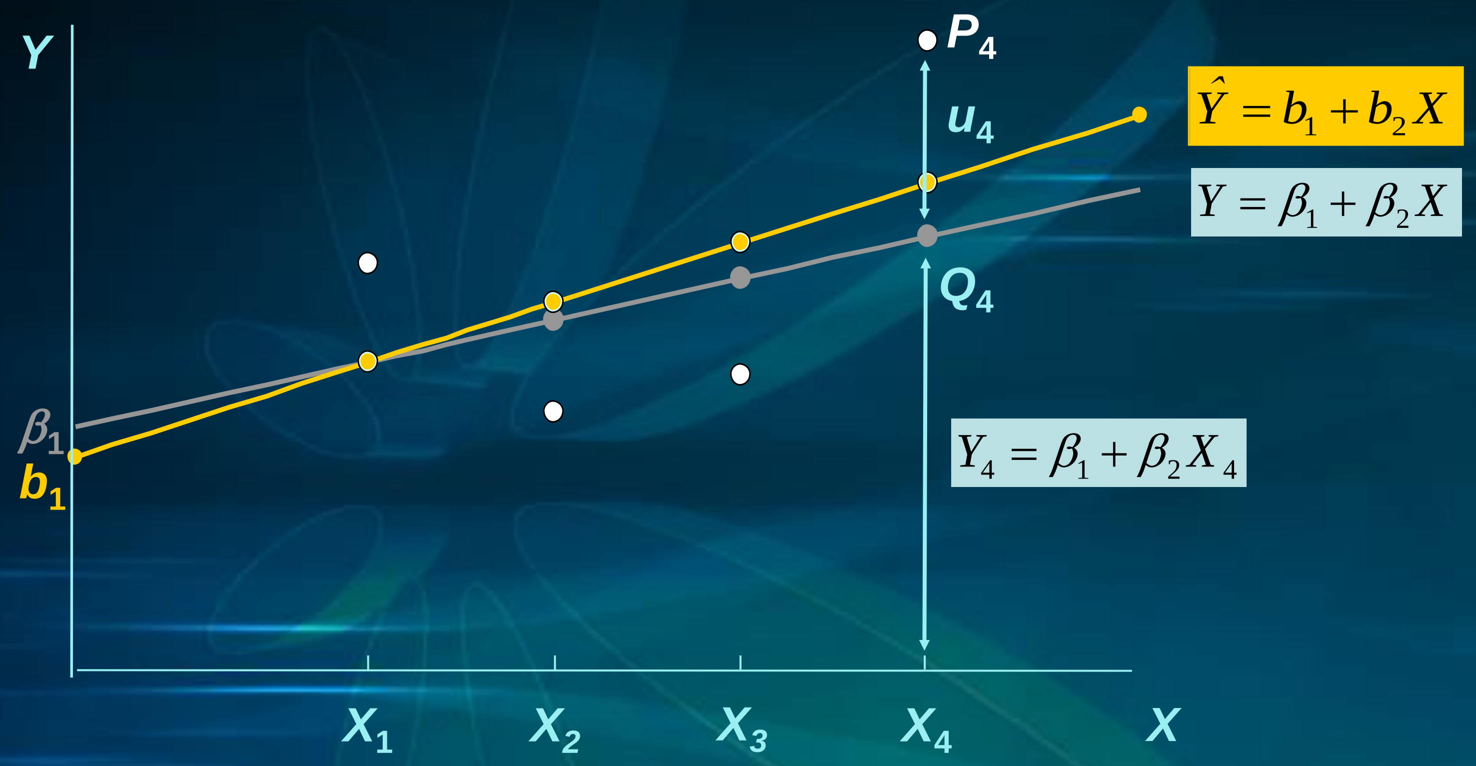


Si la regresión es "buena", los residuos y los valores del término de perturbación serán similares (ver primera y segunda obs.), pero son conceptualmente distintos.

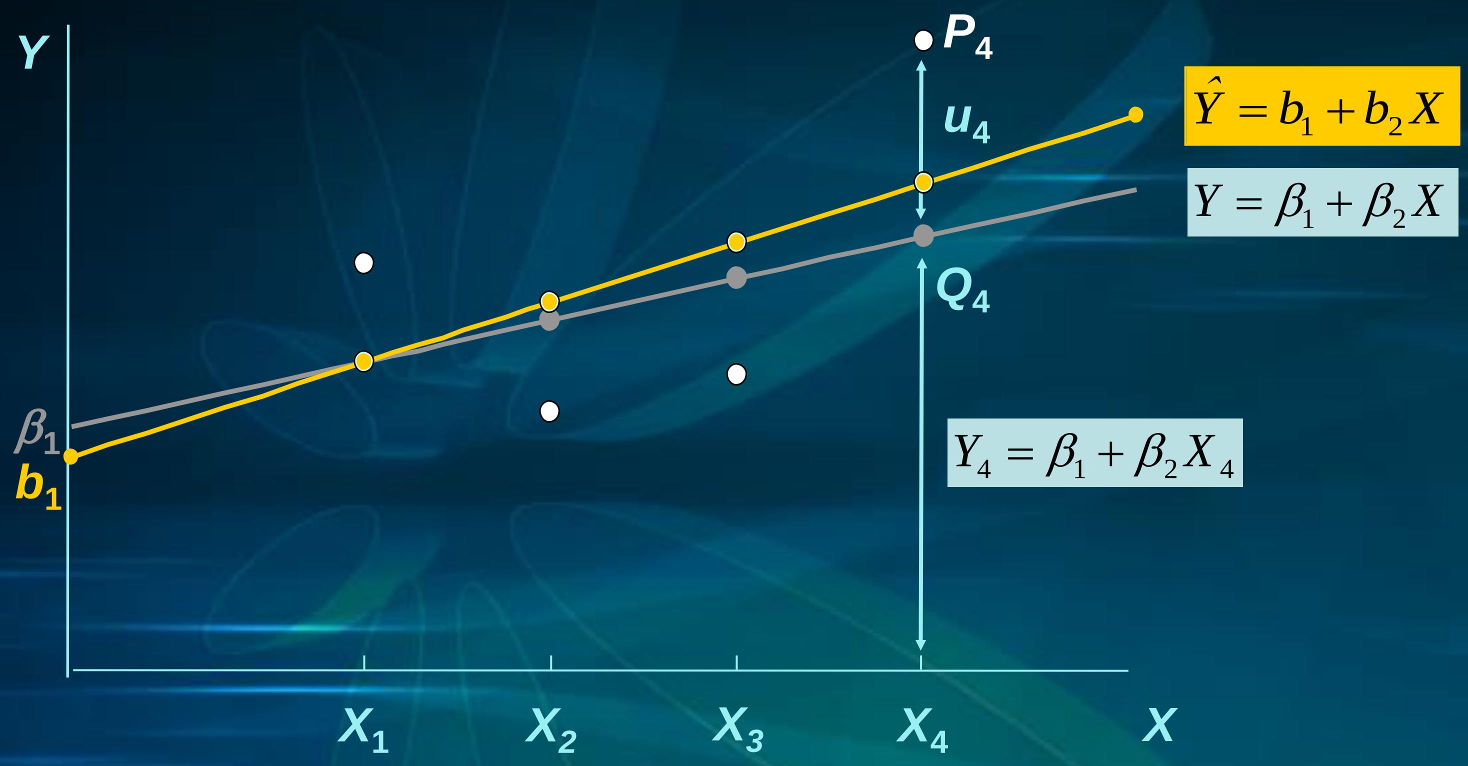
Continuación…



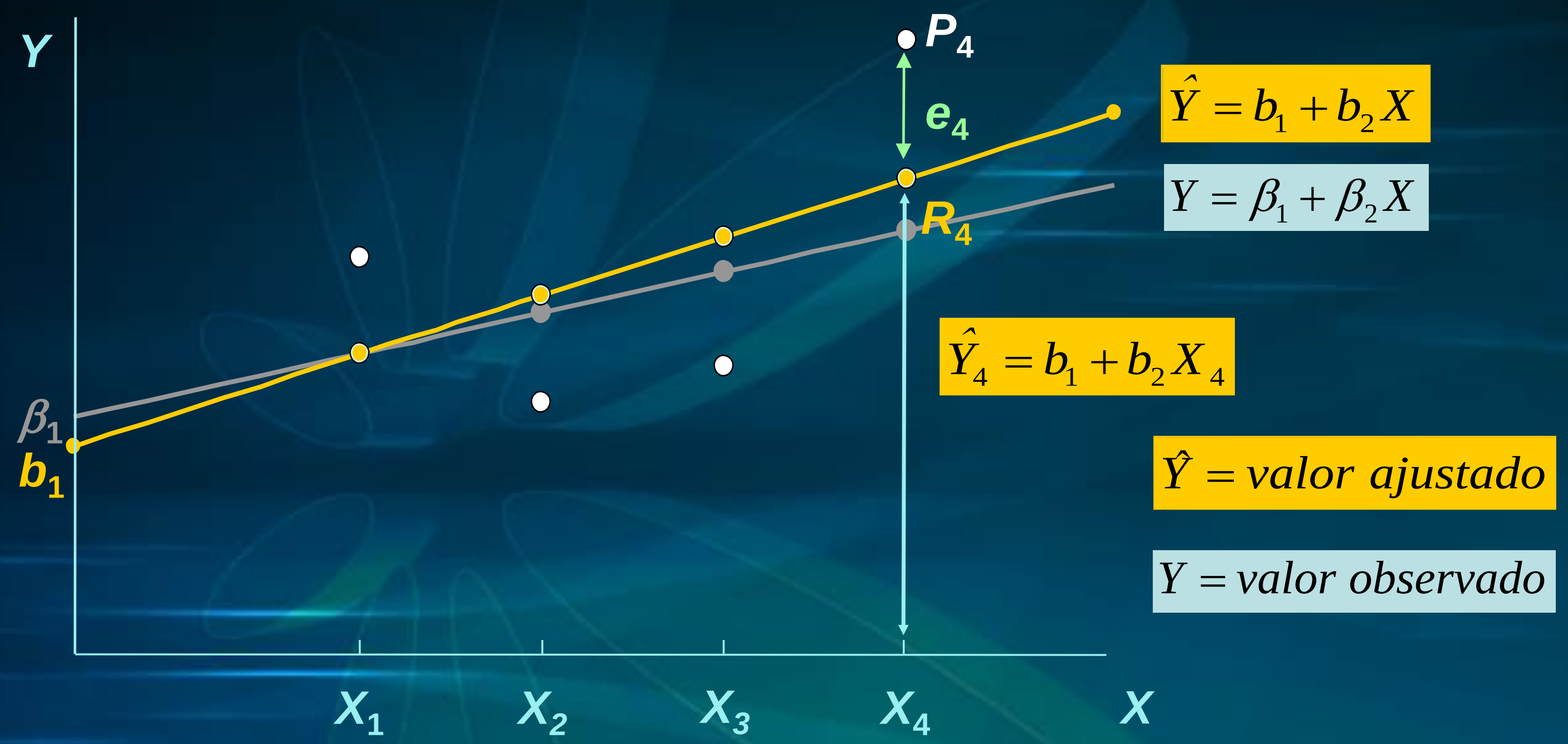
Las dos rectas serán analizadas en el curso. Cada una permite la descomposición del valor de Y. Esta descomposición se ilustrará con la cuarta observación.



Usando la relación teórica, Y puede descomponerse en un componente no estocástico y un componente aleatorio.



Es una descomposición teórica porque desconocemos los valores de β_1, β_2 o los valores del término de perturbación.



La otra descomposición es en referencia a la línea ajustada. En cada observación, el valor observado de Y es igual al valor ajustado de Y más el residuo. Esta es una descomposición operativa que se utilizará para usos prácticos.

Principio de mínimos cuadrados

- El ***Principio de Mínimos Cuadrados*** consiste en:

$$\underset{b_1,b_2}{Min}\ SCR = \sum_{i=1}^{n} e_i^2 = e_1^2 + \ldots + e_n^2$$

- Condiciones de Primer Orden.

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial b_1} = 0 \qquad \frac{\partial \sum_{i=1}^{n} e_i^2}{\partial b_2} = 0$$

- Ecuaciones Normales:

$$\sum_{i=1}^{n} e_i = 0 \qquad \sum_{i=1}^{n} x_i e_i = 0$$

Condiciones de primer y segundo orden

- **Primera condición de primer orden:**

$$\frac{\partial SCR}{\partial b_1} = 0 \quad \Rightarrow \quad -2\left(\sum Y_i - b_1 - b_2 X_i\right) = 0$$

$$\sum Y_i - b_1 n - b_2 \sum X_i = 0$$

$$b_1 = \overline{Y} - b_2 \overline{X}$$

$$\overline{X} = \frac{\sum X_i}{n}$$

$$\sum X_i = n\overline{X}$$

- **Segunda condición de primer orden:**

$$\frac{\partial SCR}{\partial b_2} = 0 \quad \Rightarrow \quad -2\left(\sum Y_i - b_1 - b_2 X_i\right) X_i = 0$$

Metodología de estimación…

$$\frac{\partial SCR}{\partial b_2}=0 \quad \Rightarrow \quad -2\left(\sum Y_i-b_1-b_2X_i\right)X_i=0$$

$$\sum((Y_i-b_1-b_2X_i)X_i)=0$$

$$\sum Y_iX_i-b_1\sum X_i-b_2\sum X_i^2=0$$

$$\sum Y_iX_i-\left(\overline{Y}-b_2\overline{X}\right)\sum X_i-b_2\sum X_i^2=0$$

$$\sum Y_iX_i-\overline{Y}\sum X_i+b_2\overline{X}\sum X_i-b_2\sum X_i^2=0$$

$$\sum Y_iX_i-n\overline{YX}+b_2n\overline{X}^2-b_2\sum X_i^2=0$$

$$b_2=\frac{\sum Y_iX_i-n\overline{YX}}{\sum X_i^2-n\overline{X}^2}=\frac{\sum\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}{\sum\left(X_i-\overline{X}\right)^2}$$

Continuación…

Demostración de Resultados Importantes:

$$\begin{aligned}\sum\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right) &= \sum X_iY_i-\sum X_i\bar{Y}-\sum \bar{X}Y_i+\sum \bar{X}\bar{Y}\\ &= \sum X_iY_i-\bar{Y}\sum X_i-\bar{X}\sum Y_i+n\bar{X}\bar{Y}\\ &= \sum X_iY_i-\bar{Y}\left(n\bar{X}\right)-\bar{X}\left(n\bar{Y}\right)+n\bar{X}\bar{Y}\\ &= \sum X_iY_i-n\bar{X}\bar{Y}\end{aligned}$$

Análogamente:

$$\begin{aligned}\sum\left(X_i-\bar{X}\right)^2 &= \sum\left(X_i^2-2\bar{X}X_i+\bar{X}^2\right)=\sum X_i^2-2n\bar{X}^2+n\bar{X}^2\\ &= \sum X_i^2-n\bar{X}^2\end{aligned}$$

Continuación…

- Estimador MCO de *beta1* $b_1 = \bar{Y} - b_2 \bar{X}$
- Estimador MCO de *beta2*:

$$SC3: \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 > 0$$

$$b_2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \qquad b_2 = r_{YX} \frac{\sqrt{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2}}{\sqrt{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}}$$

- Estos estimadores replican los poblacionales (normalidad):

Poblacional	***Muestral***
$\beta_{1_{YX}} = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X$	$b_1 = \bar{Y} - b_2 \bar{X}$
$\beta_{2_{YX}} = \rho \frac{\sigma_Y}{\sigma_X}$	$b_2 = r \frac{SE(Y)}{SE(X)}$

Estimador de la varianza

- ***Estimador de la varianza del término de perturbación.***
  - No se puede estimar a partir de una muestra de valores de ***u***, pues depende de los parámetros poblacionales no observados *beta1* y *beta2*.
  - Puede obtenerse un estimado a partir de los errores de estimación. El estimador es:

$$\hat{\sigma}^2_{MCO} \equiv s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$$

Modelo con K variables

- El modelo de regresión poblacional de "k" variables, para una observación o período de tiempo, se denota por:

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik} + u_i$$

$$y = X\beta + u$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

y el modelo de regresión estimado:

$$\hat{Y}_i = b_1 + b_2 x_{i2} + b_3 x_{i3} + \cdots + b_k x_{ik}$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \cdots + \hat{\beta}_k x_{ik}$$

Matricialmente

- Matricialmente, para las n observaciones:

$$y = X\beta + u$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{(n\times1)} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}_{(n\times k)} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k\times1)} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}_{(n\times1)}$$

- Entonces, dado $\hat{y} = Xb$, se tiene:

$$y = \hat{y} + e = Xb + e$$

Continuación…

- Suma de los cuadrados de los ***residuos*** o ***errores de estimación***:

$$SCR(b) \equiv \sum_{i=1}^{n} e_i^2 = e'e = \left(y - \hat{y}\right)'\left(y - \hat{y}\right)$$

$$SCR(b) \equiv y'y - 2b'X'y + b'X'Xb$$

$$SCR(b) \equiv y'y - 2y'Xb + b'X'Xb$$

- De acuerdo al ***Principio de Mínimos Cuadrados Ordinarios***, se ***minimiza SCR***, lo cual implica que se cumpla las C.P.O. y C.S.O:

Continuación…

$$\frac{\partial SCR}{\partial b} = 0 \quad \Rightarrow \quad (X'X)b = X'y$$

$$\frac{\partial^2 SCR}{\partial b \partial b'} = 2X'X \quad \textit{positivo definida}$$

Reglas de derivación de matrices

$$\frac{\partial (a'x)}{\partial x} = a; \quad \frac{\partial (Ax)}{\partial x} = A';$$

$$\frac{\partial (x'Ax)}{\partial x} = 2(x'A)' = 2Ax, \quad A: \textit{matriz simétrica}$$

$$\frac{\partial (x'Ax)}{\partial A} = x'x, \quad x'x: \textit{matriz cuadrada}$$

- Las ***Ecuaciones Normales*** se obtienen a partir de las Condiciones de Primer Orden (C.P.O.):

$$(X'X)b = X'y \qquad \Rightarrow \quad X'e = 0$$

Continuación…

- En el modelo de regresión lineal clásico de *k* variables se obtiene ***k*** ecuaciones normales: una para cada parámetro especificado en la ecuación de regresión.
- A partir de la CPO se obtiene el vector de ***estimadores de MCO del MRLCK***:

$$b = \hat{\beta}_{MCO} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'y$$

- Supuesto de "rango completo por columnas" es importante! Para derivar el estimador MCO sólo se necesita que X'X tenga inversa y ello se cumple sólo si se cumple el SC 3.

Continuación…

Estimador de la Varianza s².

- Es razonable obtener el estimador a partir de la suma de cuadrados de los residuos:

$$e = y - Xb$$
$$e = y - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$$
$$e = (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')y$$
$$e = \mathbf{M}y$$

- Donde ***M*** es una matriz ***simétrica*** e ***idempotente***; además:

$$\mathbf{M}' = \mathbf{M}; \quad \mathbf{M} = \mathbf{M}^2$$

$$Me = e$$

$$MX = 0$$

Continuación…

- Con estas propiedades, se tiene que:

$$\boldsymbol{e} = \boldsymbol{M}\boldsymbol{y} = \boldsymbol{M}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) = \boldsymbol{M}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{M}\boldsymbol{u}$$
$$\boldsymbol{e} = \boldsymbol{M}\boldsymbol{u}$$

- Elevando al cuadrado los residuos:

$$\mathbf{e'e} = (\mathbf{Mu})'\mathbf{Mu} = \mathbf{u'M'Mu}$$
$$\mathbf{e'e} = \mathbf{u'Mu}$$

- Aplicando la esperanza matemática a esta expresión y aprovechando las propiedades de la traza obtenemos:

Continuación…

$e = Mu$ (Definición del residuo del modelo multivariado.

$$E(e'e) = E\left((Mu)'Mu\right)$$

$$E(e'e) = E(u'M'Mu)$$

$$E(e'e) = E(u'Mu) \text{ (} M \text{ es simétrica e idempotente)}$$

$$E(e'e) = E[tr(u'Mu)] \text{ (un escalar es igual a su traza).}$$

$$E(e'e) = E[tr(uu'M)] = E[tr(Muu')] \text{ (propiedades del operador traza)}$$

$$E(e'e) = tr\ [E(Muu')] \text{ (linealidad del operador traza)}$$

$$E(e'e) = tr\ [ME(uu')] \text{(SC4f)}$$

$$E(e'e) = tr\ [M\sigma_u^2 I\ ] \text{ (SC5)}$$

$$E(e'e) = \sigma_u^2 tr\ [M\ ]$$

$$E(e'e) = \sigma_u^2(n-k) \text{ (propiedad de la matriz M)}$$

$$E\left(\frac{e'e}{n-k}\right) = \sigma_u^2$$

Continuación…

$$E(\boldsymbol{e}'\boldsymbol{e}) = \sigma_u^2(n-k)$$

$$E\left(\frac{\boldsymbol{e}'\boldsymbol{e}}{n-k}\right) = \sigma_u^2$$

- Así, el estimador de la varianza del término de perturbación, y de la regresión, es:

$$s^2 \equiv \hat{\sigma}_u^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k}$$

- El ***error estándar de la estimación*** o ***error estándar de la regresión***, ***s***, es la desviación estándar de los valores de Y alrededor del plano de regresión.

Características muestrales

- Características asociadas a un modelo bivariado:
  - La línea de regresión estimada pasa por los valores medios muestrales de las variables. $\overline{Y} = b_1 + b_2 \overline{X}$
  - El promedio observado es igual al promedio estimado: $\overline{Y} = \overline{\hat{Y}}$
  - Los *residuos* de MCO tienen correlación cero con la variable independiente (en la muestra). $Corr(X, e) = 0$
  - No existe correlación muestral entre los valores estimados de la variable endógena y los residuos. $Corr(\hat{y}, e) = 0$

Características muestrales de MCO

- Para el modelo de regresión clásico de k variables:

  - La ecuación de regresión o hiperplano de regresión pasa por los ***puntos medios*** del espacio “k” dimensional.

$$\bar{Y} = b_1 + b_2\bar{X}_2 + b_3\bar{X}_3 + \cdots + b_3\bar{X}_k \quad \Rightarrow \quad \bar{y} = \bar{X}b$$

  - El promedio observado es igual al promedio estimado: $\boldsymbol{i}'y = \boldsymbol{i}'\hat{y}$

  - La correlación muestral de los regresores con los residuos es cero. $\mathbf{X}'e = \mathbf{0}$

  - Los valores estimados de la variable dependiente no están correlacionados con los residuos. $\hat{\mathbf{y}}'e = \mathbf{0}$

Estimación de MCO como proyección

- ***M*** es la Matriz Generadora de Residuos: $My = e$
  - ***M*** (n*xn*), simétrica, idempotente, dim(M)=tr(M).
  - ***M*** es la Matriz Aniquiladora: la proyección de X sobre X es perfecta y el residuo es cero. $MX = 0$
- ***P*** es la ***Matriz de Proyección*** sobre el espacio columna ***X***

$$\hat{y} = Xb = Py$$

  - ***P*** (*nxn*), simétrica, idempotente, dim(P)=tr(P).
  - La proyección de X sobre X es igual a X: $PX = X$

Continuación…

- Además, ***PM = MP = 0***.
- Entonces, la estimación por MCO particiona ***y*** en dos componentes ortogonales:

$$y = \hat{y} + e$$
$$y = Py + My$$

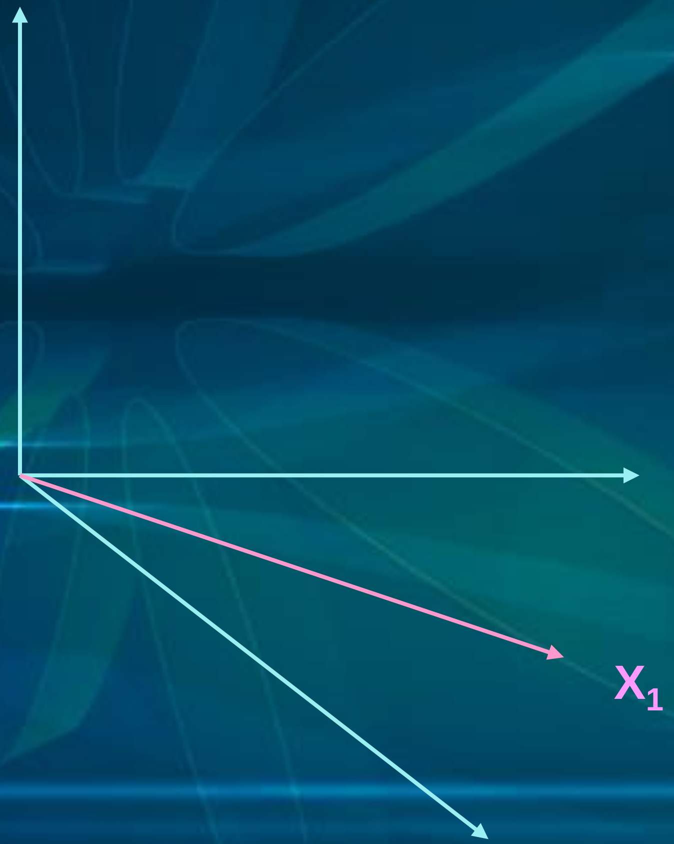
- ***Teorema de Pitágoras: Suma de Cuadrados***

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

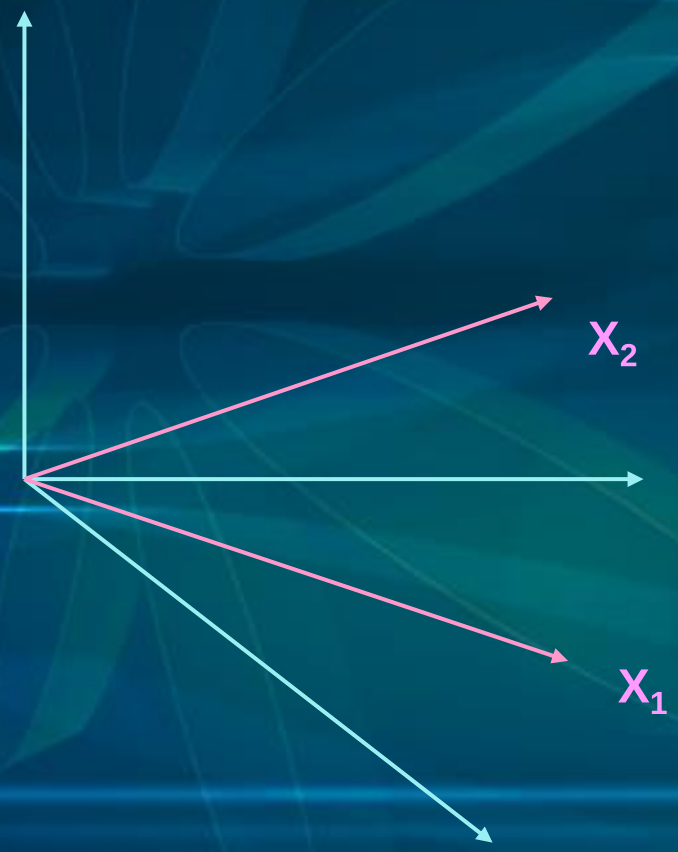
Geometría de MCO

- Geometría de MCO:

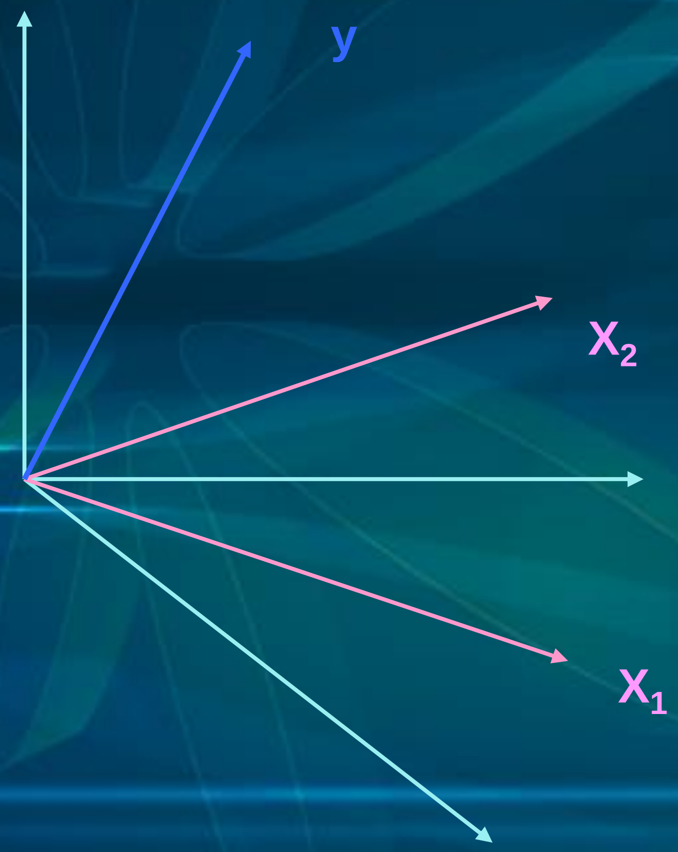
- Geometría de MCO:



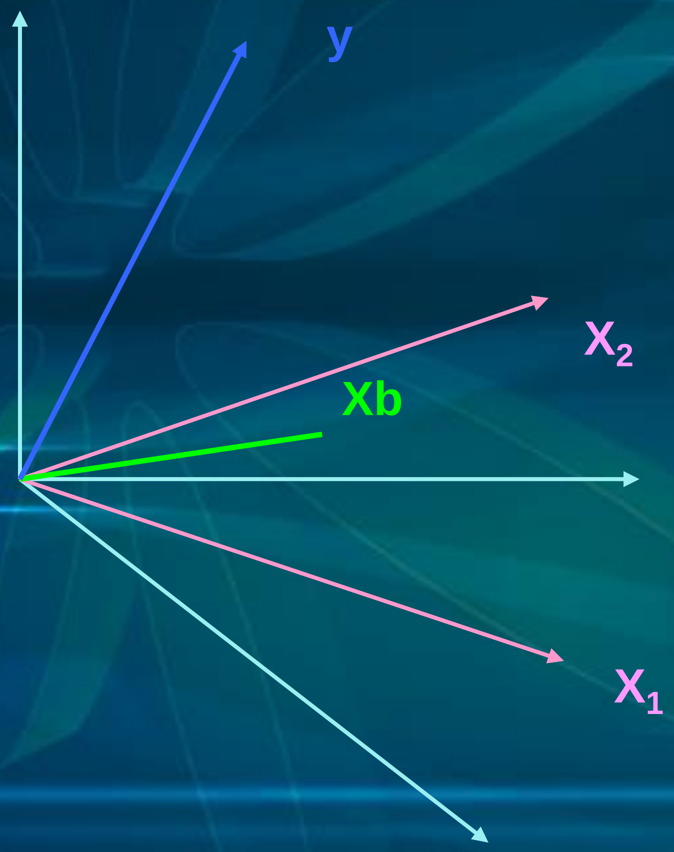
- Geometría de MCO:



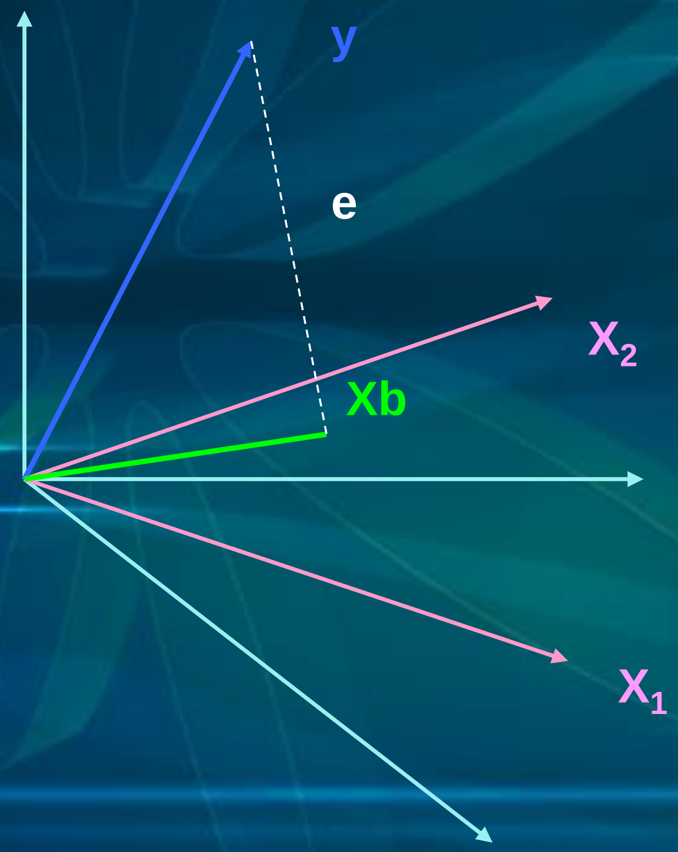
- Geometría de MCO:



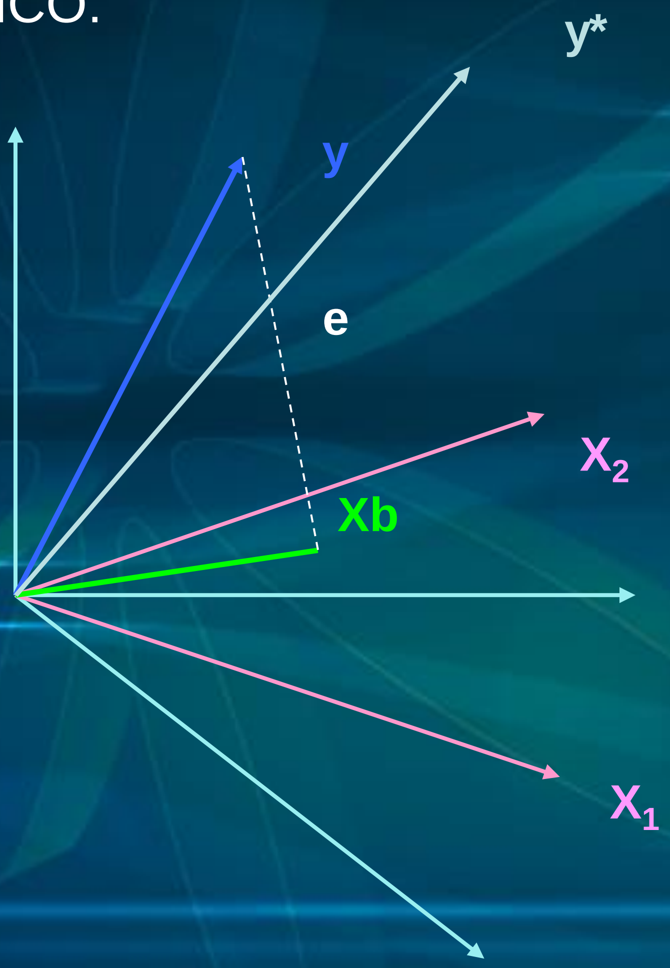
- Geometría de MCO:



- Geometría de MCO:

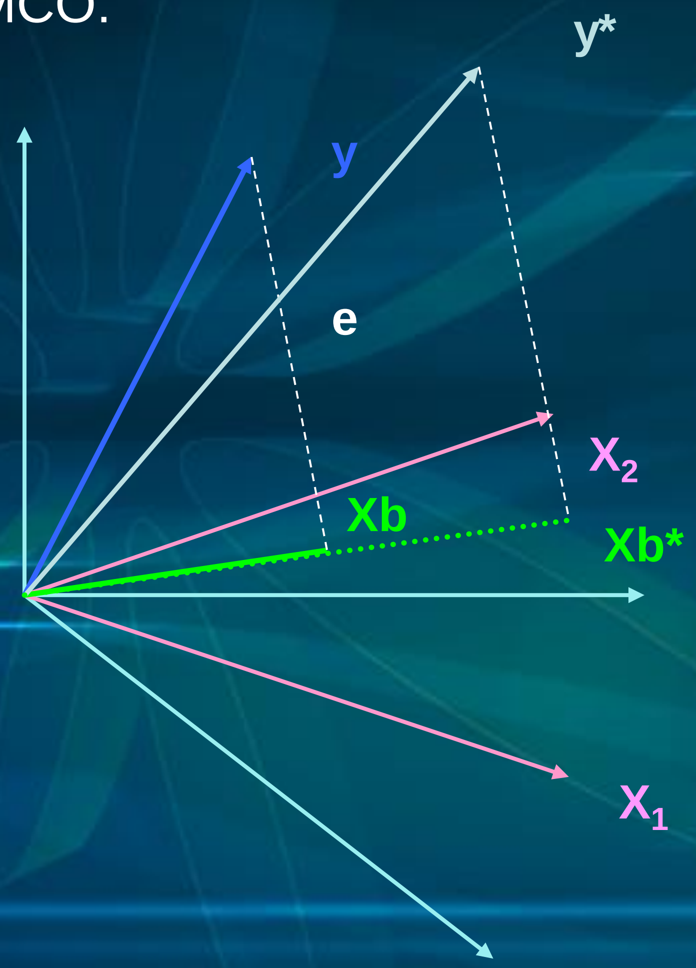


- Geometría de MCO:



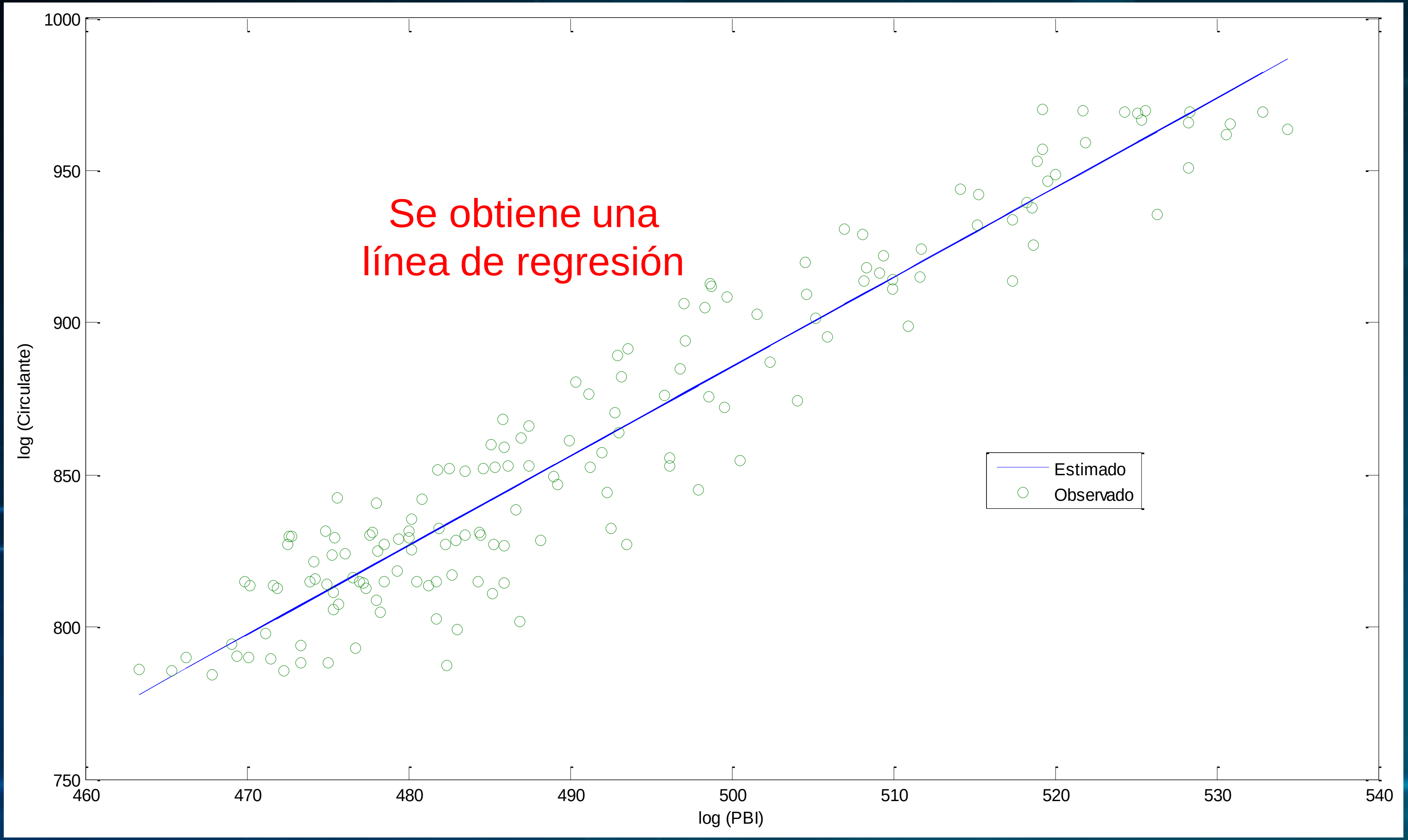
Estimación como proyección

- Geometría de MCO:



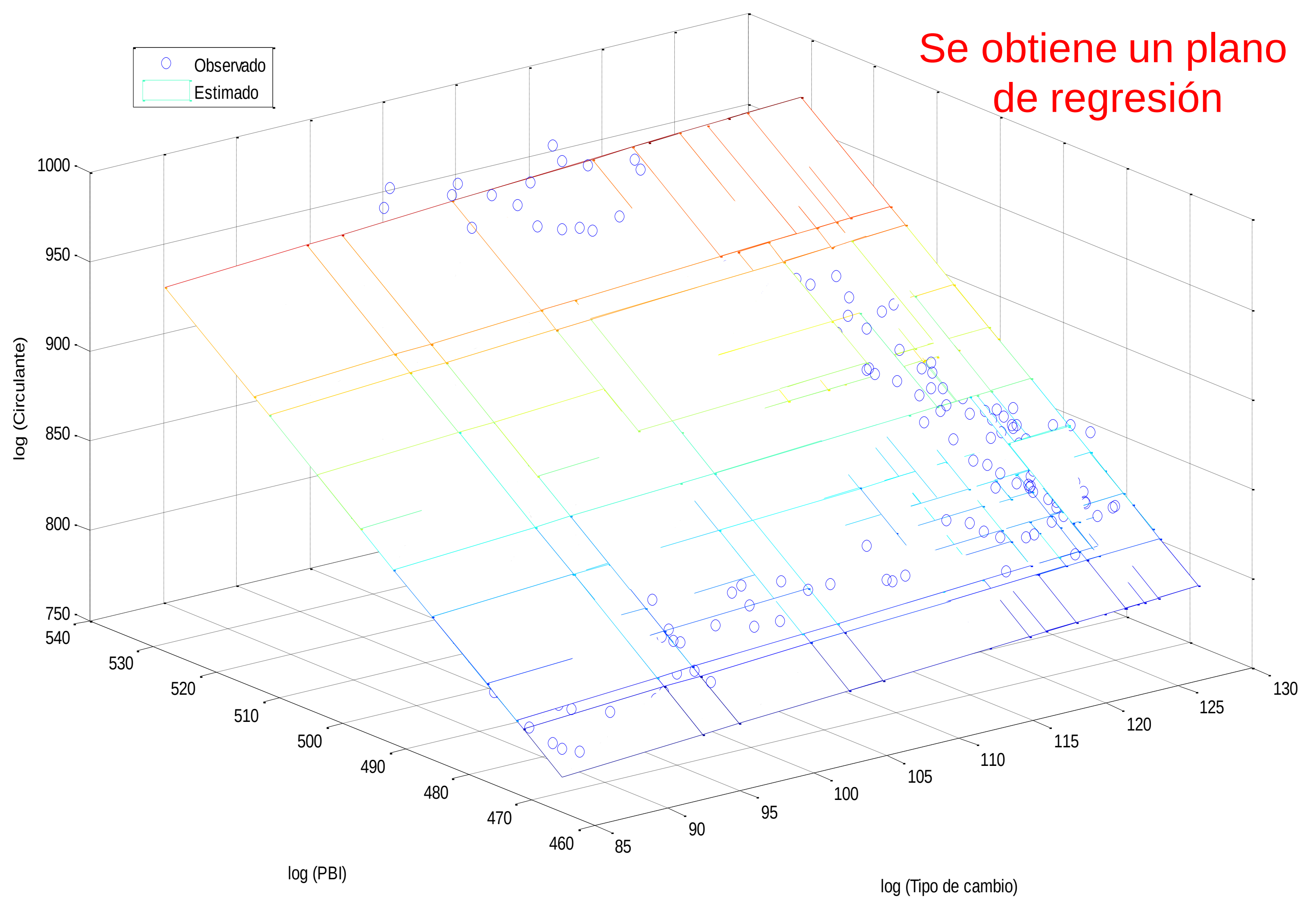
Continuación…

- A través del uso de matrices : caso de dos variables

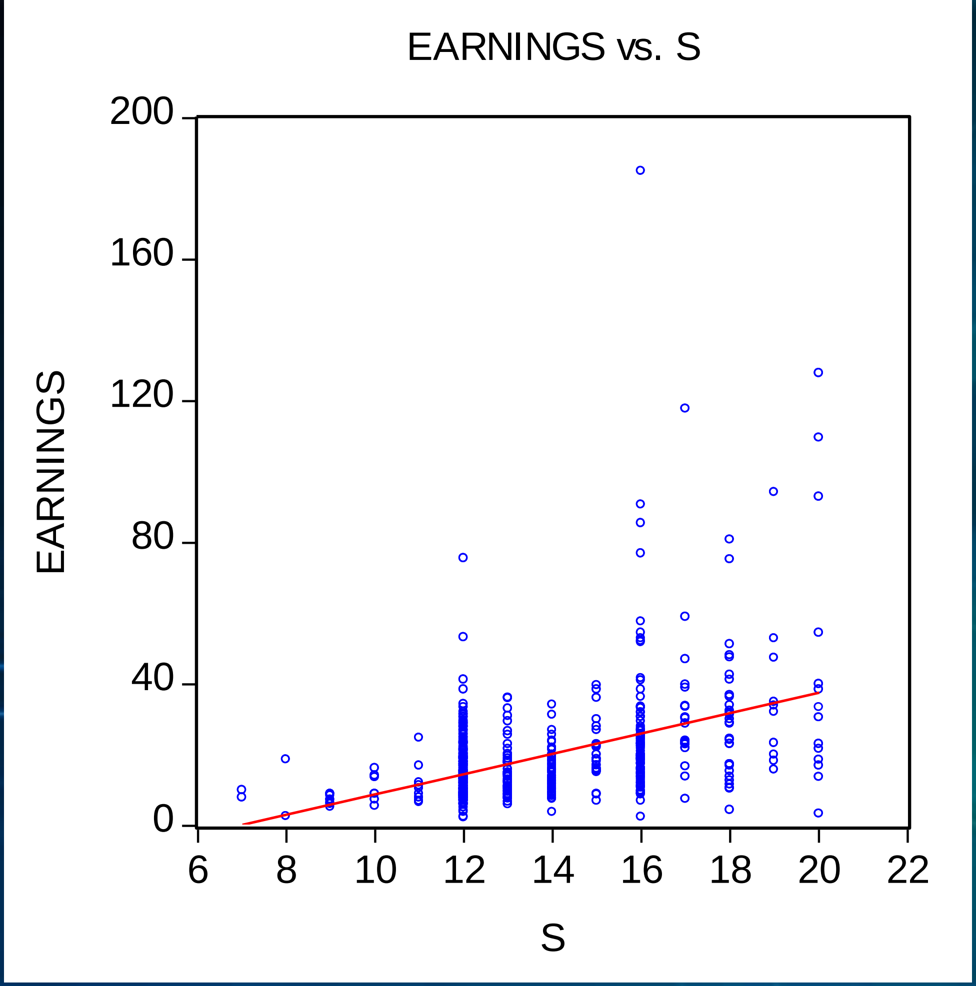


Continuación…

- A través del uso de matrices : caso de tres variables



APLICACIÓN: Estimación del salario por hora



Earnings: salario por hora
S: años de educación

Mayores observaciones para eduación básica y secundaria. Pocas observaciones para mayores años de estudio en la muestra analizada.

APLICACIÓN: Estimación del salario por hora

Dependent Variable: EARNINGS
Method: Least Squares
Date: 07/31/11 Time: 22:59
Sample: 1 540
Included observations: 540

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-19.79196	3.585928	-5.519341	0.0000
S	2.870943	0.255034	11.25710	0.0000

R-squared	0.190639	Mean dependent var	19.93339
Adjusted R-squared	0.189135	S.D. dependent var	16.43569
S.E. of regression	14.80002	Akaike info criterion	8.230831
Sum squared resid	117843.8	Schwarz criterion	8.246725
Log likelihood	-2220.324	F-statistic	126.7222
Durbin-Watson stat	1.885753	Prob(F-statistic)	0.000000

Continuación…

Dependent Variable: EARNINGS
Method: Least Squares
Date: 07/31/11 Time: 22:59
Sample: 1 540
Included observations: 540

$$EARNINGS = -19.79 + 2.87S$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-19.79196	3.585928	-5.519341	0.0000
S	2.870943	0.255034	11.25710	0.0000

R-squared	0.190639	**Mean dependent var**	19.93339
Adjusted R-squared	0.189135	S.D. dependent var	16.43569
S.E. of regression	14.80002	Akaike info criterion	8.230831
Sum squared resid	117843.8	Schwarz criterion	8.246725
Log likelihood	-2220.324	F-statistic	126.7222
Durbin-Watson stat	1.885753	Prob(F-statistic)	0.000000