UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL DE HUAMANGA

EL MODELO DE REGRESIÓN LINEAL CLÁSICO

Notas de clase del curso Econometría I

Material estrictamente académico

Prof. William Canales Molina

ESQUEMA

1. FUNCIÓN DE DENSIDAD NORMAL Y REGRESIÓN.
2. EL MODELO DE REGRESIÓN
3. EL TÉRMINO DE PERTURBACIÓN
4. LOS SUPUESTOS CLÁSICOS

1. FUNCIÓN DE DENSIDAD NORMAL

- Una variable aleatoria (X) sigue una distribución normal con parámetros (μ) real y finito, y σ > 0.

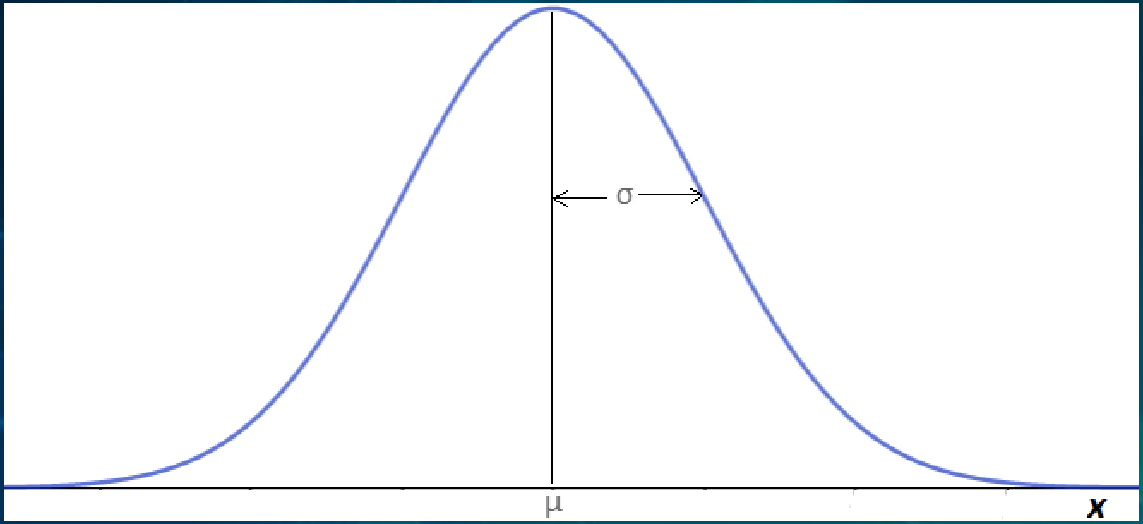
$$X \rightarrow N(\mu ; \sigma)$$

- La función de densidad de probabilidad se especifica como:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Función de densidad normal

- Su esperanza y su varianza son:
- $E(X) = \mu$ $Var(X) = \sigma^2$
- El gráfico de una distribución normal estándar con media 0 y varianza 1.



FUNCIÓN DE DENSIDAD NORMAL Y REGRESIÓN

- Sea la función de densidad Normal conjunta bivariada:

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

- Funciones de densidad de probabilidad marginal:

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma_x}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}$$

$$f(y)=\frac{1}{\sqrt{2\pi}\sigma_y}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}$$

Continuación…

- La función de densidad condicional de *Y dado X*:

$$f(y\,|\,x) = \frac{f(x,y)}{f(x)}$$

$$f(y\,|\,x) = \frac{1}{\sqrt{2\pi\sigma_y^2(1-\rho^2)}}\exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left[y-\left(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right)\right]^2\right\}$$

- Media y Varianza Condicionales:

$$E(Y\,|\,x) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x) = \left(\mu_y - \rho\frac{\sigma_y}{\sigma_x}\mu_x\right) + \rho\frac{\sigma_y}{\sigma_x}x = \alpha_{y|x} + \beta_{y|x}x$$

$$Var(Y\,|\,x) = (1-\rho^2)\sigma_y^2 \equiv \sigma_{Y|x}^2$$

2. MODELO DE REGRESIÓN

- El Modelo de Regresión simple y múltiple:

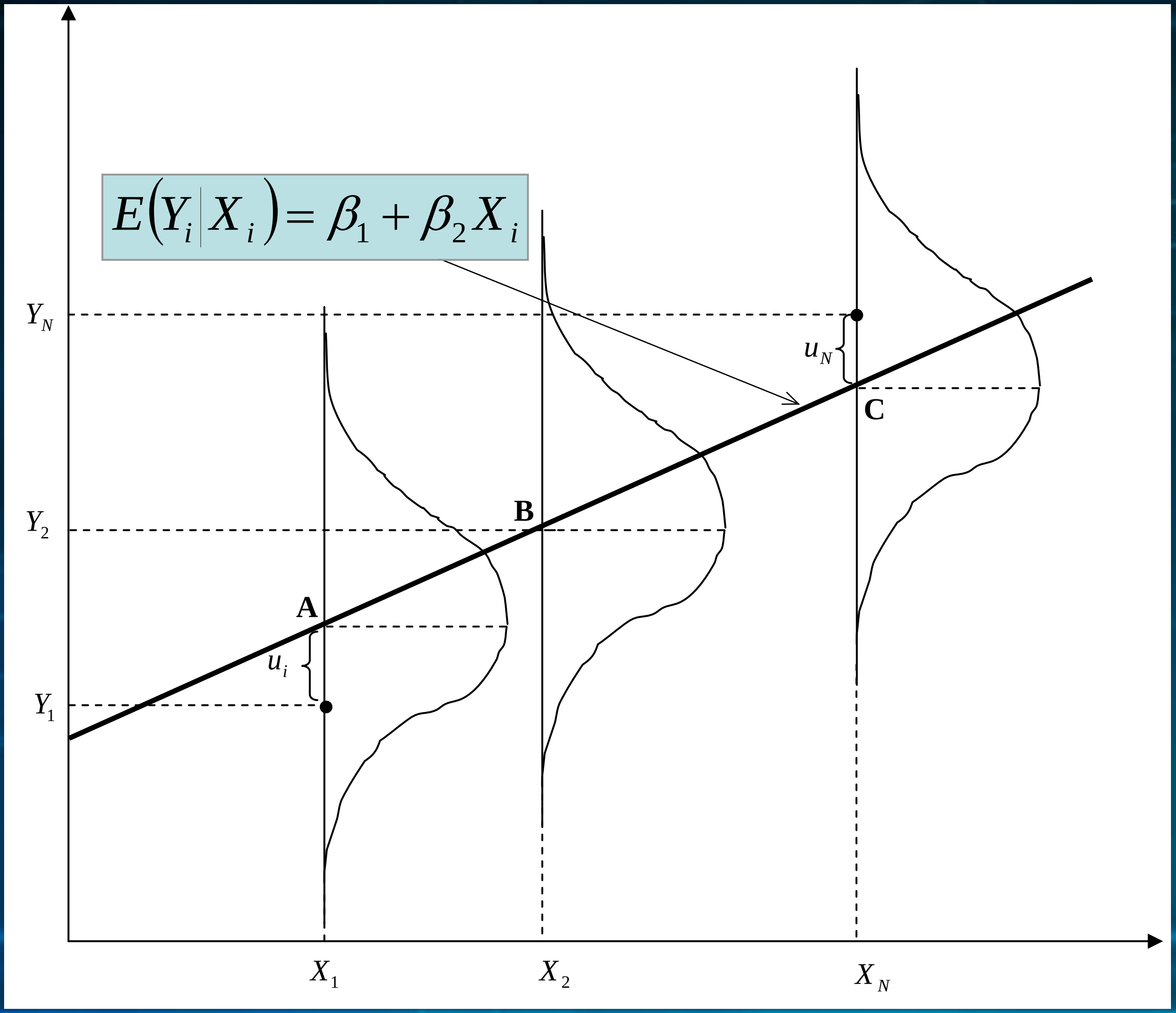
$$y = E(y \mid x) + u \qquad\qquad y = E(y \mid x_1, x_2, \ldots, x_k) + u$$

- – Y = variable respuesta, variable dependiente, regresando, explicada.

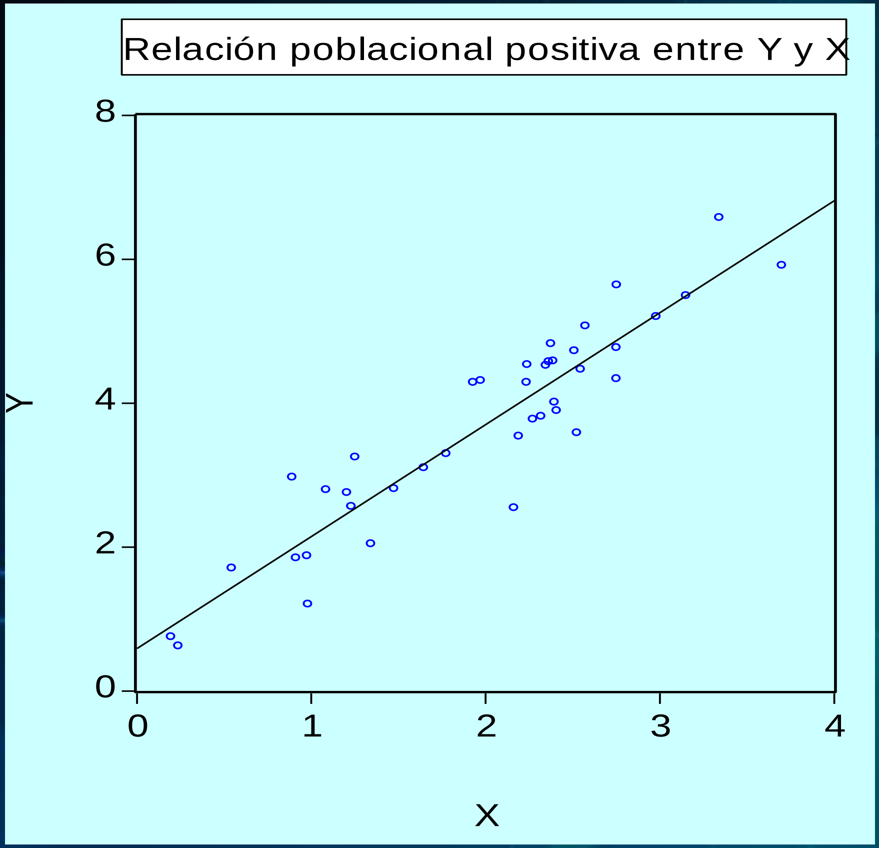
- – X_k = variables independientes, regresores, explicativas. *k* regresores.

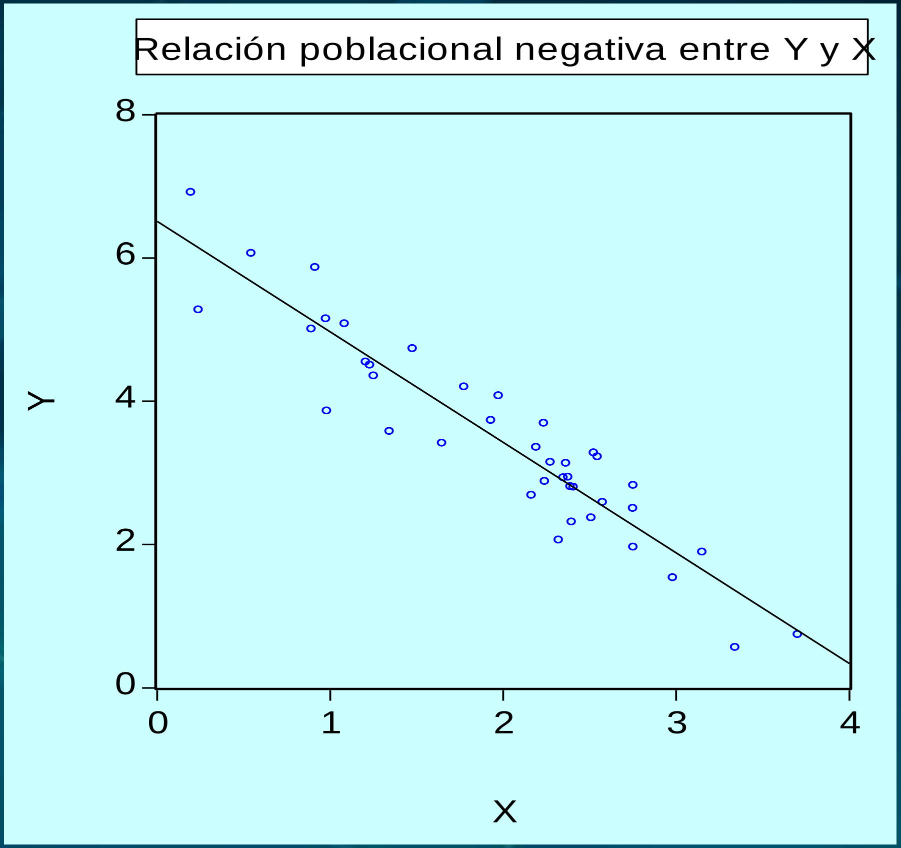
- Término de perturbación: $u \equiv y - E(y \mid x_1, x_2, \ldots, x_k)$

Continuación…



Continuación…





Causalidad

CAUSALIDAD EN EL MODELO DE REGRESIÓN LINEAL

- Es importante tener en cuenta que un modelo de regresión ***no implica la existencia de causalidad*** entre las variables.

$$y = E(y \mid x_1, \ldots, x_k) + u$$

$$x_i = E(x_i \mid y, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) + u$$

- La causalidad - si existiera - estará determinada por la teoría económica y reforzada por pruebas estadísticas adecuadas.

3. TÉRMINO DE PERTURBACIÓN O ERROR

Definición

- Denominado término estocástico.
- La palabra ***estocástico*** proviene del griego ***stokhos*** que significa ***objetivo*** o ***blanco de una ruleta***:
  - Una relación estocástica es una relación que ***no siempre da en el blanco***.
  - Así, el término de perturbación mide los ***errores*** o ***fallas*** de la relación determinística: $u = Y - \beta_1 - \beta_2 X$

Continuación…

- La presencia del término de perturbación se justifica por los siguientes argumentos (no mutuamente excluyentes):
  - Omisión de la influencia de eventos sistemáticos, muy importantes pero poco importantes para la relación.
  - Omisión de la influencia de innumerables eventos no sistemáticos, muy importantes pero poco importantes para la relación.
  - Error de medición de las variables utilizadas.
  - Aleatoriedad del comportamiento humano ante situaciones similares.

Continuación…

- Omisión de variables explicativas: se excluyen variables que no se pueden medir.
- Agregación de variables micro-económicas. Relaciones individuales pueden tener distintos parámetros.
- Incorrecta especificación del modelo en términos de su ***estructura***: común en datos de series de tiempo, la variable endógena puede depender de sus valores pasados.
- Incorrecta especificación ***funcional***: relaciones lineales vs. no lineales.

Especificación del término de perturbación

Suponga que una variable Y es una función lineal de otra variable X, con parámetros desconocidos β_1 **y** β_2 que deseamos estimar.

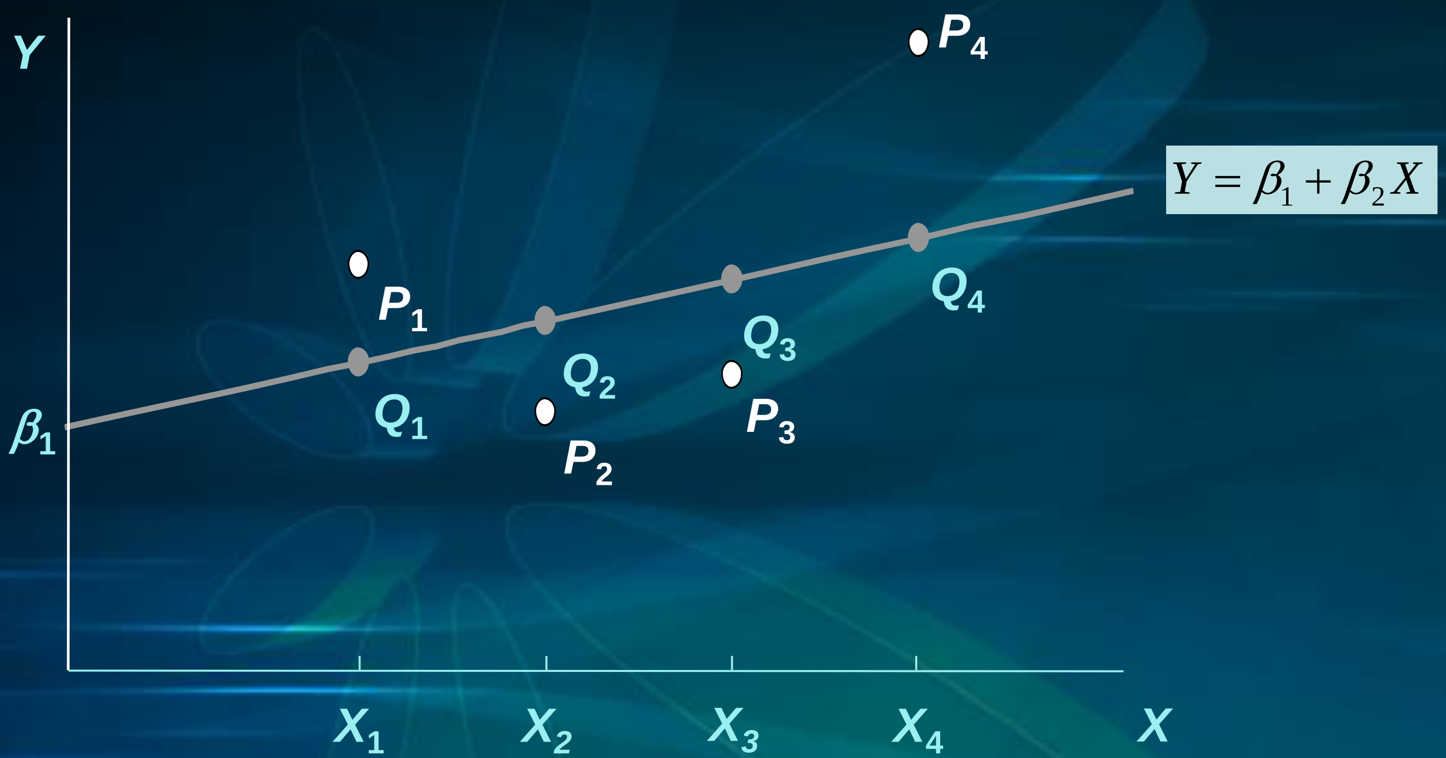
Continuación…

Suponga que se cuenta con una muestra de 4 observaciones para las variables X e Y.

Continuación…

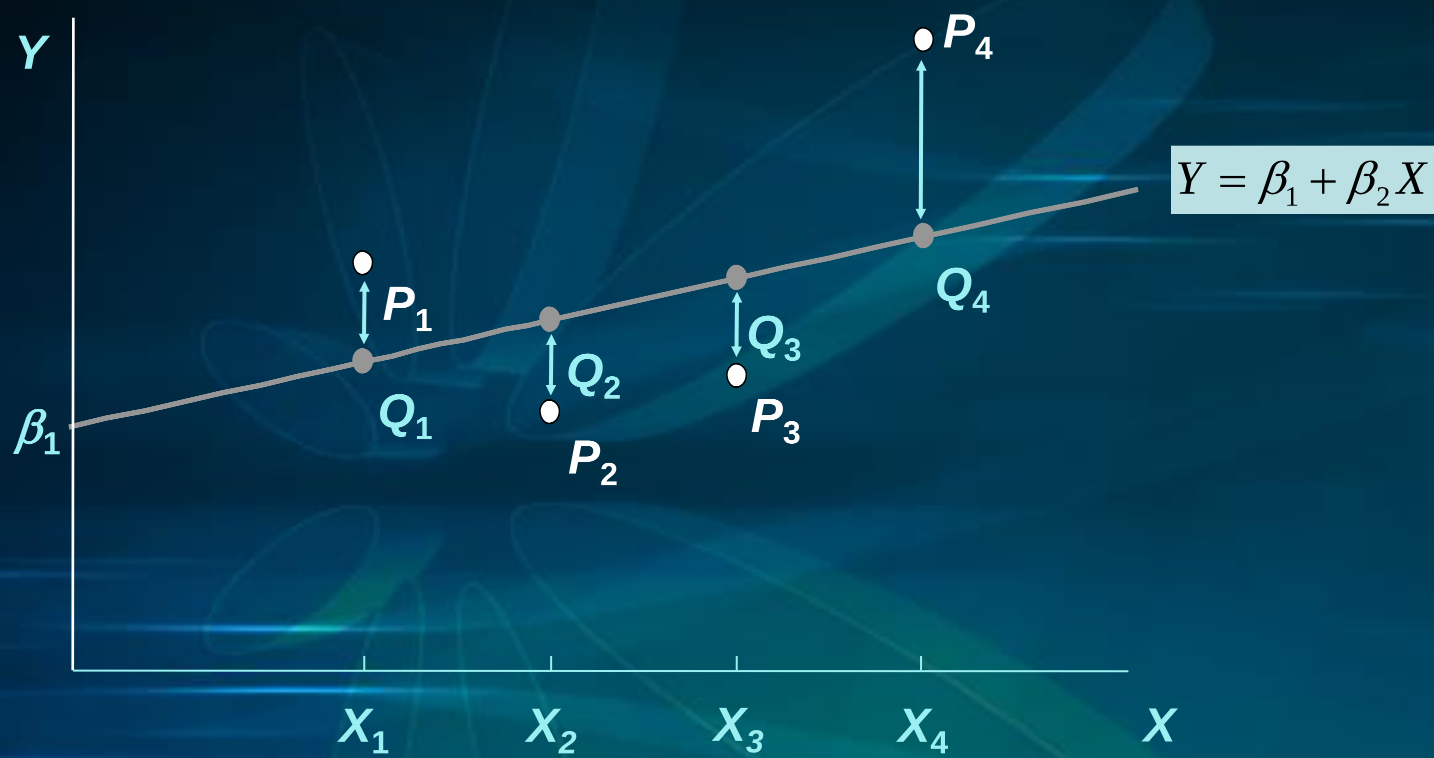
Si la relación entre X e Y fuera exacta, las observaciones estarían en la línea recta y no habría problema de obtener los valores exactos de los parámetros poblacionales β_1 y β_2.

Continuación…



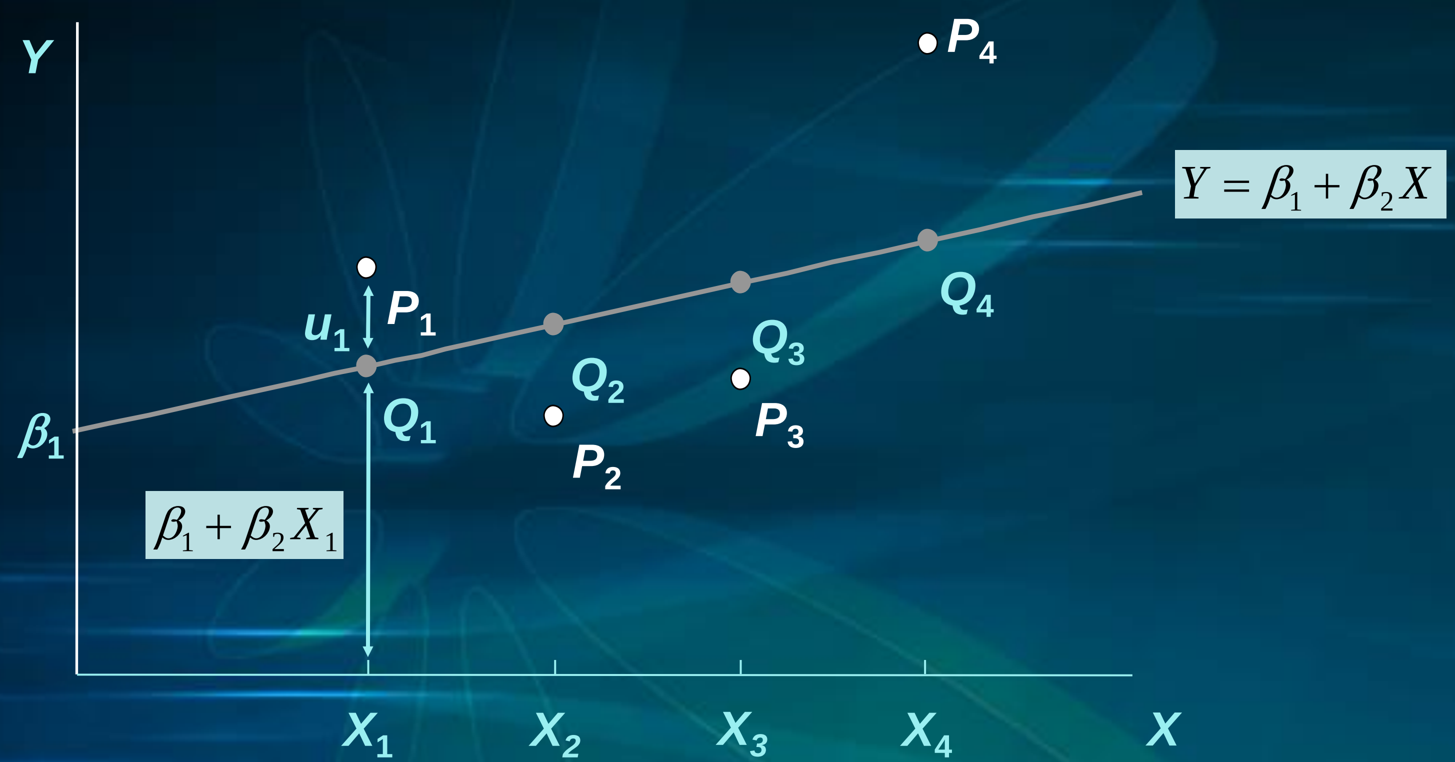
En la práctica, muchas relaciones no son exactas y los valores observados de Y son distintas de los valores que tomaría se estuvieran en la línea recta (P vs. Q)

Continuación…



Así, el término de perturbación permite justificar tal divergencia y por ello el modelo estadístico puede escribirse como $Y = \beta_1 + \beta_2 X + u$, donde ***u*** es el término de perturbación.

Continuación…



Cada valor de Y tiene un componente no estocástico, $\boldsymbol{\beta_1 + \beta_2 X}$**,** y un componente ***u*****.** Por ejemplo, la primera observación tiene estos dos componentes.

4. LOS SUPUESTOS CLÁSICOS

- **SC1**:
  - Linealidad de la esperanza condicional. ¿Término de perturbación aditivo? Sí.
  - Regresores Adecuados.
  - Parámetros Constantes.
- **SC2**: Supuesto de Regresión
- **SC3**: Rango Completo por columnas (no multicolinealidad).
- **SC4**: Ausencia de relación estadística entre X y perturbaciones.
- **SC5**: Perturbaciones esféricas: Homocedasticidad y No Autocorrelación.

Continuación…

SC1: LINEALIDAD DE LA ESPERANZA CONDICIONAL

1. Lineal en parámetros y variables:

$$
\begin{aligned}
y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \ldots + \beta_k x_{1k} + u_1 \\
y_2 &= \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \ldots + \beta_k x_{2k} + u_2 \\
&\vdots \\
y_n &= \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \ldots + \beta_k x_{nk} + u_n
\end{aligned}
$$

Notación matricial:

$$y = X\beta + u$$

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{(n\times 1)}
=
\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}_{(n\times k)}
\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k\times 1)}
+
\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}_{(n\times 1)}
$$

Interpretación…

- Coeficientes de la Regresión y Efectos Marginales
- Interpretación de los parámetros

<table>
<caption>Tabla 1: Interpretación de los coeficientes del modelo de regresión</caption>
<tr><th></th><th>X</th><th>Log(X)</th></tr>
<tr><td rowspan="2">Y</td><td>Efecto Marginal</td><td></td></tr>
<tr><td>Cambio en el nivel de Y ante un cambio en una unidad de X</td><td>Cambio en el nivel de Y ante un cambio porcentual de X
(Modelo Semilog)</td></tr>
<tr><td rowspan="2">Log(Y)</td><td>Semi-elasticidad de Y ante X</td><td>Elasticidad de Y ante X</td></tr>
<tr><td>Cambio porcentual de Y ante un cambio en una unidad de X
(Modelo Semilog)</td><td>Cambio porcentual de Y ante un cambio porcentual de X: ()
(Modelo Doble log)</td></tr>
</table>

Continuación…

2. Regresores adecuados: el modelo especificado es el “verdadero”

- No se omiten variables importantes.
- No se incluyen variables redundantes.

3. Los parámetros son constantes:

- Para la muestra analizada: individuos o tiempo.
- Al menos que fluctúen (poco) alrededor de un valor constante.
- No hay cambio estructural o de régimen (series de tiempo), cualidades (corte transversal).

Continuación…

SC2: SUPUESTO DE REGRESIÓN:

- *MEDIA INCONDICIONAL DEL TÉRMINO DE PERTURBACIÓN IGUAL A CERO:* $E(u_i)=0, \quad i=1,\ldots,n$ $\quad$ $E(u)=0$
  - Regresores son fijos en muestreo repetido.
  - Regresores son variables aleatorias y con distribución totalmente independiente del término de perturbación.
- *MEDIA CONDICIONAL DEL TÉRMINO DE PERTURBACIÓN DADO X ES IGUAL A CERO:* $E(u_i \mid X)=0, \quad i=1,\ldots,n$
  - Regresores son variables aleatorias y con distribución independiente en media del término de perturbación.

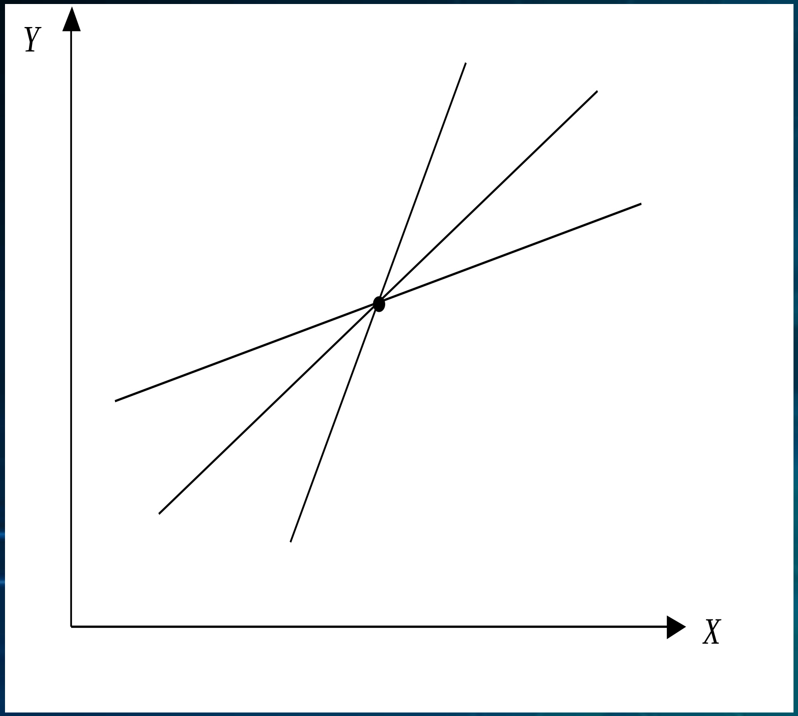
Continuación…

SC3: RANGO COMPLETO POR COLUMNAS DE X

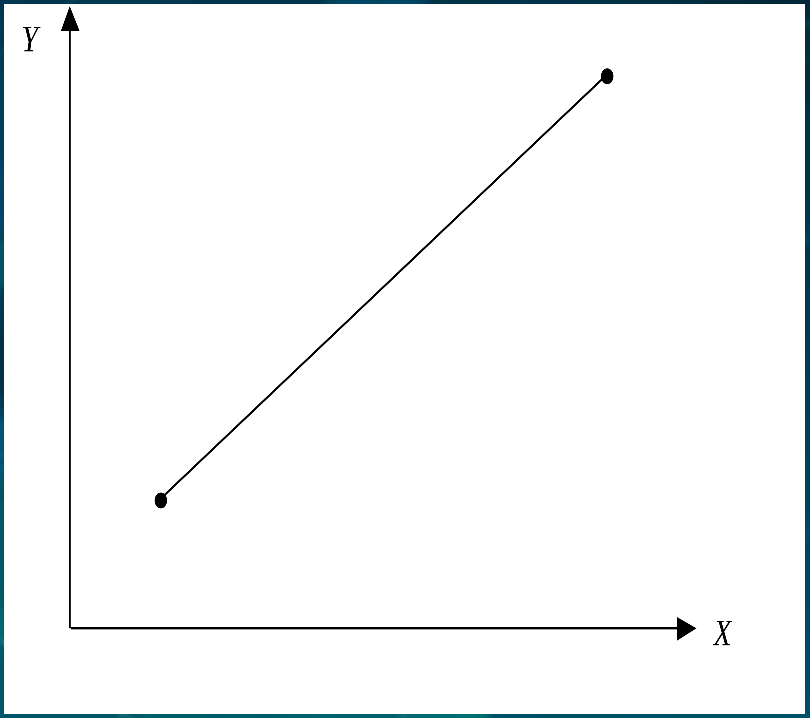
- No es posible que n<k
  - El número de observaciones es mayor al número de regresores: $n > k$ (variación de los regresores).
- Columnas linealmente independientes
  - *No existen relaciones lineales exactas entre regresores: Ausencia de Colinealidad o Multicolinealidad.*
- *Implicancias:*
  - *X'X es positivo definida*
  - *la inversa de (X'X) existe!*

Relaciones

Diversas relaciones posibles



Una única relación posible



Continuación…

SC4: AUSENCIA DE RELACIÓN ESTADÍSTICA ENTRE REGRESORES Y PERTURBACIONES:

Se presentan dos casos:

- Regresores Fijos en muestras repetidas (no estocásticos): **SC4f**
- Regresores Estocásticos:
  - Independencia total. **SC4ait**
  - Independencia en media. **SC4aim**
  - Ausencia de relación lineal contemporánea. **SC4acu**

Continuación…

- Independencia Total de las perturbaciones y regresores.

$$f\left(u_i, X_{ij}\right) = f\left(u_i\right) f\left(X_{ij}\right) \qquad i = 1, \ldots n \qquad \forall j = 1, \ldots, K$$

- Independencia en media de las perturbaciones.

$$E\left(u_i \mid X_{ij}\right) = E\left(u_i\right) \qquad i = 1, \ldots n \qquad \forall j = 1, \ldots, K$$

Si se cumple SC2 $E\left(u_i\right) = 0$, entonces : $E\left(u_i \mid X_{ij}\right) = 0$

Continuación…

- Ausencia de relación lineal contemporánea entre perturbaciones y regresores.

$$Cov(X_{ij}, u_i) = 0 \qquad i = 1, \ldots n \qquad \forall j = 1, \ldots, K$$

Si $E(u_i) = 0$, entonces :

$$Cov(X_{ik}, u_i) = E(X_{ik} u_i) = 0 \qquad i = 1, \ldots, n \qquad \forall k = 1, \ldots, K$$

Continuación…

SC5: PERTURBACIONES ESFÉRICAS

– *Homocedasticidad*: $E[u_i^2 \mid X] = \sigma^2 \quad \forall i = 1,\ldots,n$

- Supuesto sobre el segundo momento condicional.
- Si se cumple SC2 y SC4 (al menos independencia en media):

$$Var(u_i \mid X) = E[(u_i - E[u_i \mid X])^2 \mid X] = E[u_i^2 \mid X] = \sigma^2 \quad \forall i = 1,\ldots,n$$

Continuación…

– *No autocorrelación*: $E[u_i u_j \mid X] = 0 \quad \forall i \neq j$

- Si se cumple SC2 y SC3:

$$Cov[u_i, u_j] \mid X = E([u_i - E(u_i)][u_j - E(u_j)] \mid X)$$
$$Cov[u_i, u_j \mid X] = E(u_i u_j \mid X) = 0 \quad \forall i \neq j$$

- En series de tiempo: ausencia de correlación serial.

Continuación…

– *Perturbaciones Esféricas: Notación matricial*

$$E\,[\,uu'\,|\,X\,] = \sigma^2 I_n$$

- La matriz de segundos momentos es proporcional a la identidad.
- Si se cumple SC2 y SC4 (al menos independencia en media):

$$VCov(u) = E\,[uu'\,|\,X\,] = \sigma^2 I_n$$

Especificación de las perturbaciones