JUNTA DE PLANIFICACIÓN

ECONOMETRÍA

INTRODUCCIÓN

Aplicación de estadística matemática a información económica y social para dar soporte empírico a modelos construidos y obtener resultados numéricos.

Ciencia social en la cual las herramientas de la teoría económica y social, las matemáticas y la inferencia estadística son aplicadas al análisis de fenómenos económicos y sociales.

ECONOMETRÍA INTRODUCCIÓN

La teoría explica relación entre variables pero no la cuantifica

La econometría cuantifica la relación entre variables

- Da contenido empírico a la teoría
- Verificación empírica de la teoría

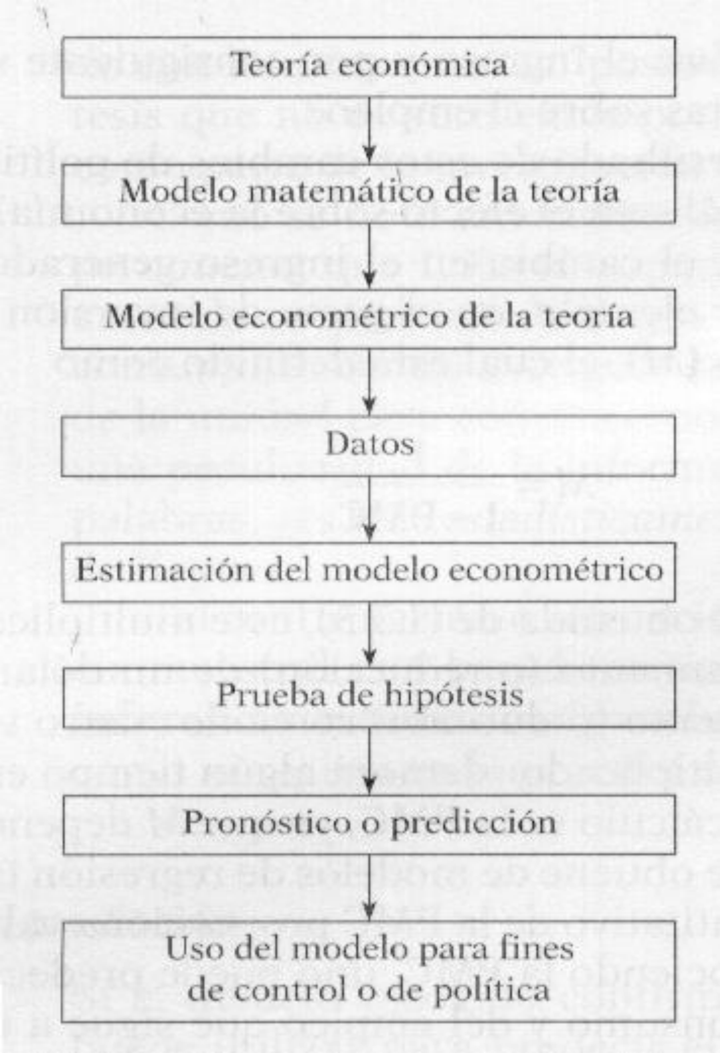
ECONOMETRÍA

INTRODUCCIÓN

Datos utilizados no se generan como resultado de un experimento controlado

- Errores de medición
- No hay control sobre las variables y la interacción entre variables

ECONOMETRÍA METODOLOGÍA CLÁSICA



ECONOMETRÍA
METODOLOGÍA CLÁSICA

1 Planteamiento de la teoría o hipótesis
2 Especificación de modelo matemático de la teoría
3 Especificación de modelo econométrico de la teoría

ECONOMETRÍA
METODOLOGÍA CLÁSICA

4 Obtención de datos
5 Estimación de los parámetros del modelo econométrico
6 Prueba de hipótesis
7 Pronóstico o predicción
8 Utilización de modelo para fines de control o de política

ECONOMETRÍA
METODOLOGÍA CLÁSICA

1 Planteamiento de la teoría o hipótesis

Keynes postula propensión marginal a consumir

Tasa de cambio del consumo generado por una unidad de cambio en ingreso > 0 y <1

JUNTA DE PLANIFICACIÓN

ECONOMETRÍA
METODOLOGÍA CLÁSICA

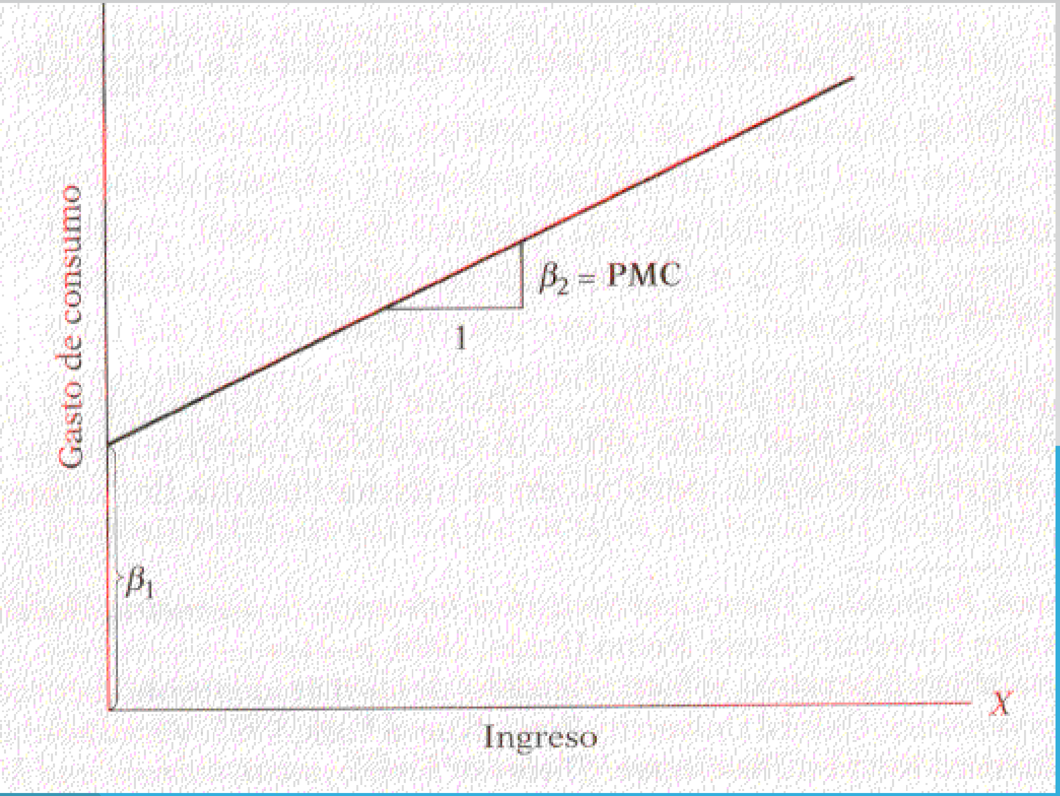
2 Especificación de modelo matemático de la teoría

$Y = \beta_1 + \beta_2 x \qquad 0 < \beta_2 < 1$

- Y = gasto de consumo
- X = ingreso
- β_1 , β_2 (PMC) son parámetros

ECONOMETRÍA
METODOLOGÍA CLÁSICA

2. Especificación de modelo matemático de la teoría



ECONOMETRÍA
METODOLOGÍA CLÁSICA

3 Especificación de modelo econométrico de la teoría

Modelo anterior supone relación exacta

Otras variables afectan el gasto de consumo

Tamaño de familia, edades de miembros, religión, etc.

ECONOMETRÍA
METODOLOGÍA CLÁSICA

3 Especificación de modelo econométrico de la teoría

$$Y = \beta_1 + \beta_2 x + u$$

Donde u es le término de perturbación o error

Variable estocástica o aleatoria

Tiene propiedades probabilísticas

Relación entre variables no es exacta

Ejemplo en programa E-views

EViews

File Edit Object View Proc Quick Options Window Help

Equation: EQ_CONSUMO Workfile: REGRESION2::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: CONSUMO
Method: Least Squares
Date: 08/28/13 Time: 17:38
Sample: 1982 1996
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-184.0780	46.26198	-3.979034	0.0016
INGRESO	0.706408	0.007827	90.24707	0.0000

R-squared	0.998406	Mean dependent var	3964.087
Adjusted R-squared	0.998284	S.D. dependent var	489.6614
S.E. of regression	20.28525	Akaike info criterion	8.981231
Sum squared resid	5349.390	Schwarz criterion	9.075638
Log likelihood	-65.35923	Hannan-Quinn criter.	8.980226
F-statistic	8144.534	Durbin-Watson stat	2.081830
Prob(F-statistic)	0.000000		

JUNTA DE PLANIFICACIÓN

ECONOMETRÍA METODOLOGÍA CLÁSICA

4 Obtención de datos

Para obtener valores de β_1 , β_2 , necesitamos datos

TABLA I.1 INFORMACIÓN SOBRE *Y* (GASTO DE CONSUMO PERSONAL) Y *X* (PRODUCTO INTERNO BRUTO), 1982-1996, EN MILES DE MILLONES DE DÓLARES DE 1992

Año	*Y*	*X*
1982	3 081.5	4 620.3
1983	3 240.6	4 803.7
1984	3 407.6	5 140.1
1985	3 566.5	5 323.5
1986	3 708.7	5 487.7
1987	3 822.3	5 649.5
1988	3 972.7	5 865.2
1989	4 064.6	6 062.0
1990	4 132.2	6 136.3
1991	4 105.8	6 079.4
1992	4 219.8	6 244.4
1993	4 343.6	6 389.6
1994	4 486.0	6 610.7
1995	4 595.3	6 742.1
1996	4 714.1	6 928.4

Fuente: Economic Report of the President, 1998, tabla B-2, p.282.

JUNTA DE PLANIFICACIÓN

ECONOMETRÍA
METODOLOGÍA CLÁSICA

5 Estimación de los parámetros del modelo econométrico

- Estimación numérica de contenido empírico
- Utilizamos análisis de regresión

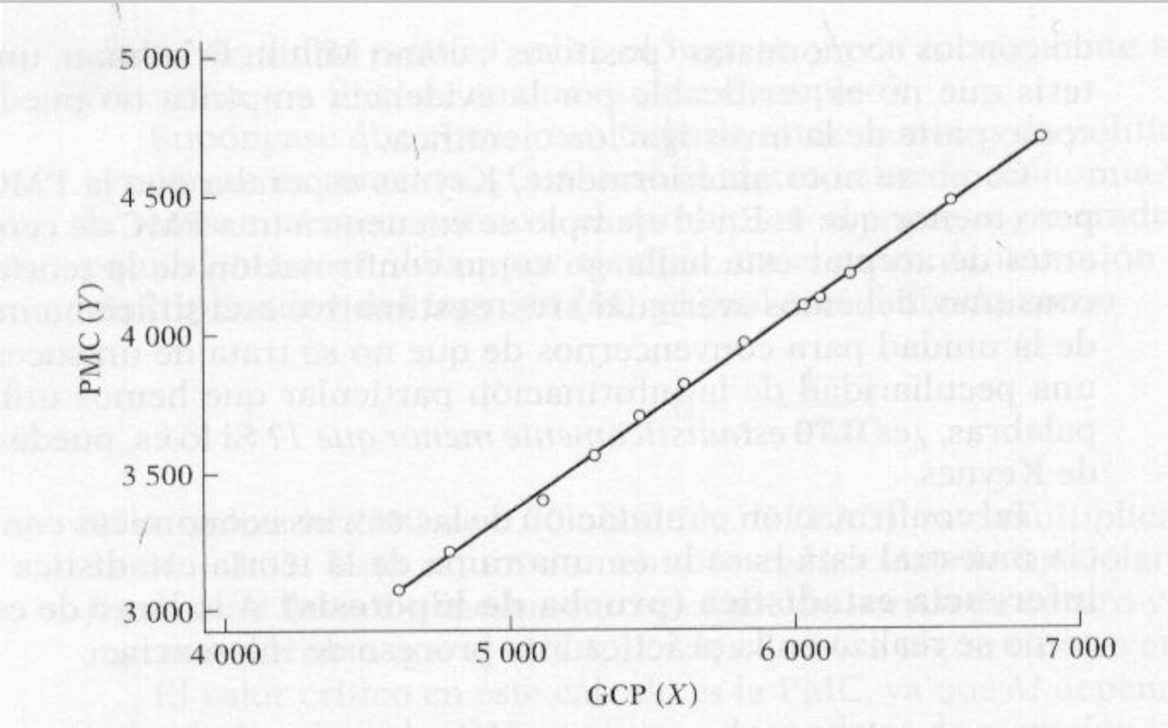
$Y = -184.1 + 0.7064X$

- Para el periodo muestral, un incremento de $1 en ingreso lleva, en promedio, un incremento de $0.72 en consumo real.

JUNTA DE PLANIFICACIÓN

ECONOMETRÍA
METODOLOGÍA CLÁSICA

5 Estimación de los parámetros del modelo econométrico



ECONOMETRÍA
METODOLOGÍA CLÁSICA

6 Prueba de hipótesis (Inferencia Estadística)

Confirmación o refutación de la teoría con base en evidencia muestral

- Encontrar si los valores estimados obtenidos concuerdan con expectativas de la teoría
  0.72 se encuentra entre 0 y 1
  ¿Es 0.72 estadísticamente menor que 1?
- Asegurar que resultado no se debe al azar o peculiaridad de la data obtenida

ECONOMETRÍA
METODOLOGÍA CLÁSICA

7 Pronóstico o predicción

- Si el modelo confirma la teoría o hipótesis, se puede utilizar para predecir valores futuros de Y, con base en el valor futuro conocido de X

ECONOMETRÍA
METODOLOGÍA CLÁSICA

8 Utilización de modelo para fines de control o de política

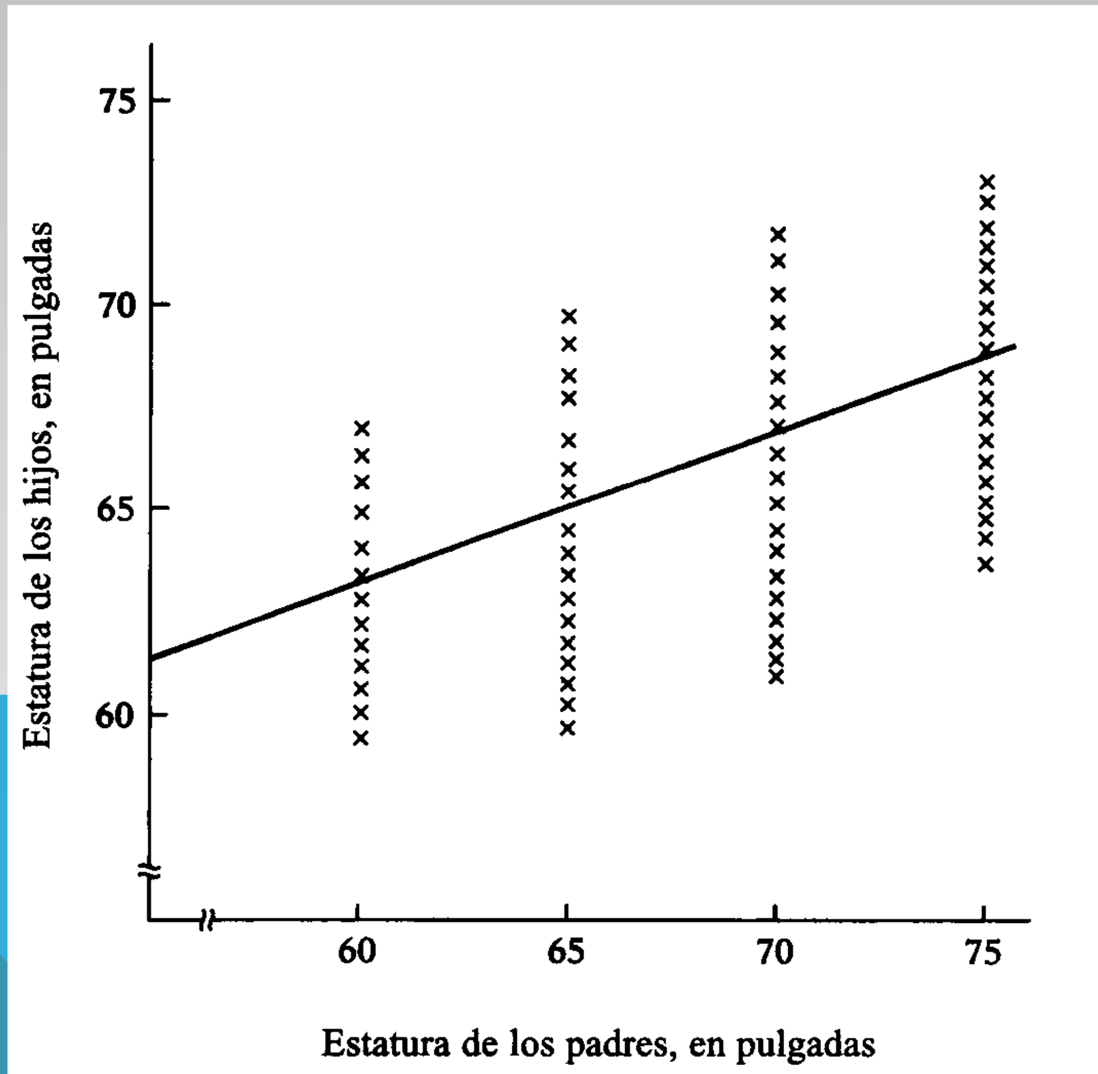
- El gobierno puede manejar la variable X para producir el nivel deseado enla variable objetivo Y
  - Nivel de gasto $4000 miles de millones
  - ¿Cuál nivel de ingreso garantiza ese nivel de gasto?
  - 3,970.9 = -184.1 + 0.7064 X
  - X = 5882 ajustar política fiscal y monetaria

NATURALEZA DE ANÁLISIS DE REGRESIÓN

Análisis de regresión trata del estudio de la dependencia de la variable dependiente en una o más variables (explicativas), con el objetivo de estimar el valor promedio poblacional de la primera en términos de los valores fijos de las últimas.

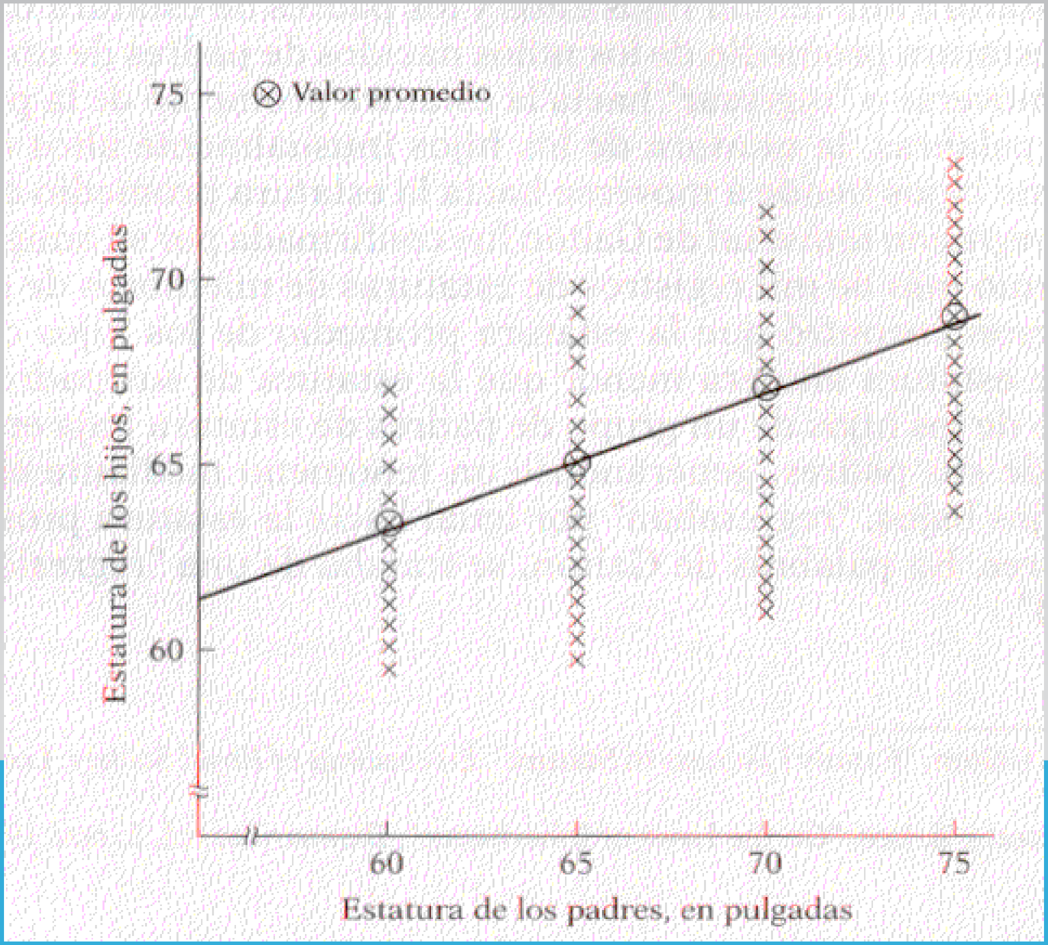
JUNTA DE PLANIFICACIÓN

EJEMPLO INICIAL



EJEMPLO INICIAL

Interés: predecir estatura promedio de los hijos conociendo estatura de los padres



RELACIONES ESTADÍSTICAS VS. RELACIONES DETERMINÍSTICAS

Relaciones Estadísticas

- Variables estocásticas – tienen distribuciones de probabilidad

Relación Funcional o Determinística

- Variables fijas

TERMINOLOGÍA Y NOTACIÓN

- Análisis de regresión simple vs. múltiple
- Terminología de variables

Variable dependiente	Variable explicativa
Variable explicada	Variable independiente
⇕	⇕
Predicha	Predictora
⇕	⇕
Regresada	**Regresora**
⇕	⇕
Respuesta	Estímulo
⇕	⇕
Endógena	Exógena
⇕	⇕
Resultado	Covariante
⇕	⇕
Variable controlada	Variable de control

TIPOS DE INFORMACIÓN

Series de tiempo

- Conjunto de observaciones sobre los valores que toma una variable en diferentes momentos del tiempo
- Variables cuantitativas, cualitativas (dicótomas y categóricas)

TIPOS DE INFORMACIÓN

Información de corte transversal

- Datos recogidos en el mismo momento en el tiempo

Información combinada

- Tiene elementos de series de tiempo y de corte transversal

Información de panel o longitudinal

- La misma unidad de corte transversal es encuestada a través del tiempo

PRECISIÓN DE LA INFORMACIÓN

Errores de observación

Errores de medición

Sesgo de selectividad muestral

Variación en métodos de muestreo

Agregación de datos

El resultado de la investigación solamente será tan bueno como lo sea la calidad de los datos

ANÁLISIS DE REGRESIÓN: IDEAS BÁSICAS

Ejemplo

- Población total de 60 familias
- Estudio de relación entre el gasto de consumo familiar semanal y el ingreso familiar semanal
- Predecir el nivel de la media poblacional de gasto de consumo conociendo el ingreso semanal

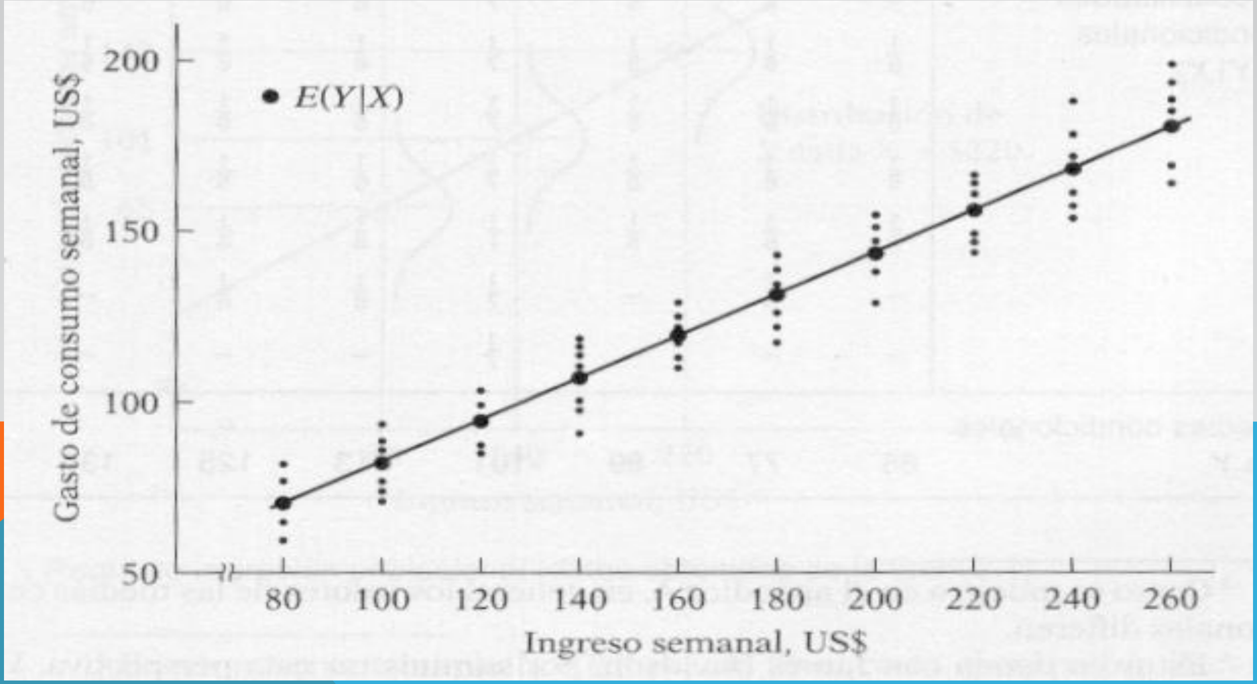
JUNTA DE PLANIFICACIÓN

EJEMPLO: INGRESO FAMILIAR SEMANAL

INGRESO FAMILIAR SEMANAL, X, \$

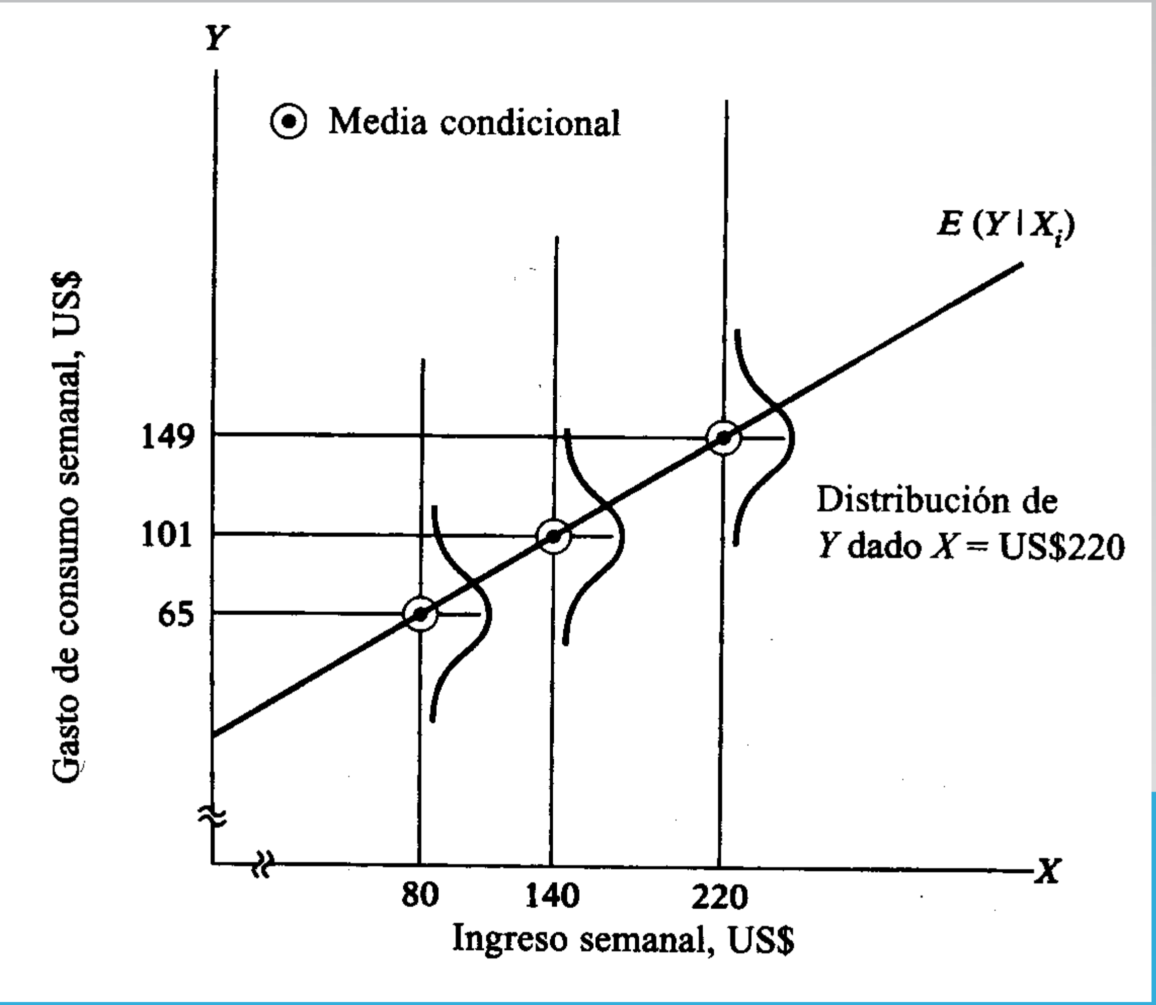
$Y\downarrow$ \ $X\rightarrow$	80	100	120	140	160	180	200	220	240	260
Gasto de consumo familiar semanal Y, \$	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Total	325	462	445	707	678	750	685	1 043	966	1 211
Medias condicionales de Y, $E(Y\mid X)$	65	77	89	101	113	125	137	149	161	173

Una curva de regresión poblacional es el lugar geométrico de las medias condicionales de la variable dependiente para los valores fijos de las variables independientes



Función de Regresión Poblacional

$E(Y/X_i)=f(X_i)=\beta_1 + \beta_2 X_i$



SIGNIFICADO DEL TÉRMINO LINEAL

Linealidad en las variables

Linealidad en los parámetros

Regresión lineal se refiere a:

- Linealidad en parámetros

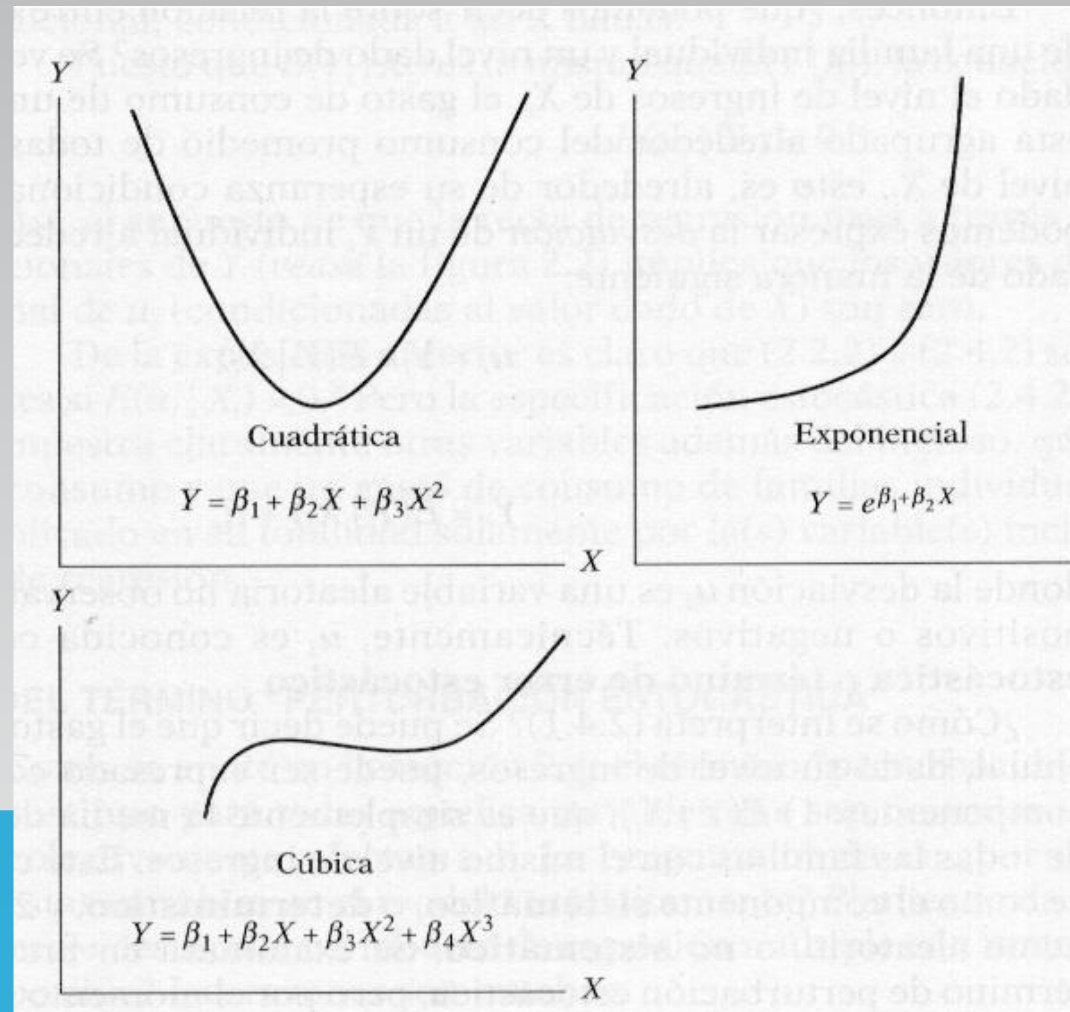
MODELOS DE REGRESIÓN LINEAL

¿Modelo lineal en los parámetros?	¿Modelo lineal en las variables?	
	Sí	No
Sí	MRL	MRL
No	MRNL	MRNL

Nota: MRL = modelo de regresión lineal.
MRNL = modelo de regresión no lineal.

JUNTA DE PLANIFICACIÓN

FUNCIONES LINEALES EN PARÁMETROS



ESPECIFICACIÓN ESTOCÁSTICA DE LA FRP

Qué sucede con el gasto de consumo de una familia individual con relación a su nivel fijo de ingreso?

- $U_i = Y_i - E(Y/X_i)$
- $Y_i = E(Y/X_i) + u_i = ß_1 + ß_2 X_i + Ui$
  - U_i es el término de error estocástico
  - $E(u_i/X_i) = 0$ (supuesto que linea de regresión pasa por medias condicionales de Y implica esto)

ESPECIFICACIÓN ESTOCÁSTICA DE LA FRP

El gasto de consumo de familias individuales no puede ser explicado en su totalidad por las variables incluidas en el modelo de regresión

U_i es sustituto para todas las variables omitidas

¿POR QUÉ NO SE INTRODUCEN TODAS LAS VARIABLES EN EL MODELO?

- Vaguedad de la teoría
- No disponibilidad de información
- Variables centrales vs. periféricas
- Aleatoriedad intrínseca en el comportamiento humano
- Variables próximas inadecuadas

¿POR QUÉ NO SE INTRODUCEN TODAS LAS VARIABLES EN EL MODELO?

Principio de parsimonia

- Principio de la cuchilla de afeitar de Occam: “Las descripciones deben mantenerse lo más simples posibles hasta el momento que se demuestre que resultan inadecuadas”
- Forma funcional incorrecta

JUNTA DE PLANIFICACIÓN

Dos muestras

FUNCIÓN DE REGRESIÓN MUESTRAL

TABLA 2.4
Muestra aleatoria tomada de la población de la tabla 2.1

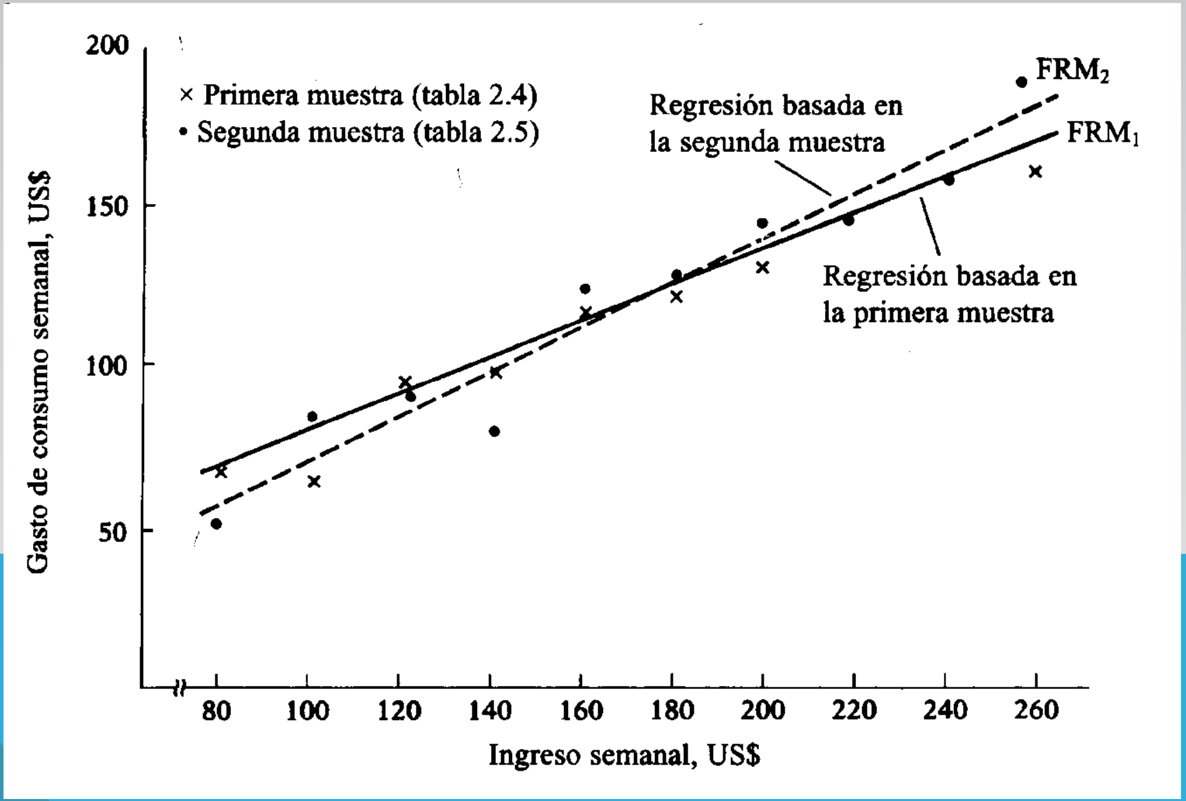
Y	*X*
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

TABLA 2.5
Otra muestra aleatoria tomada de la población de la tabla 2.1

Y	*X*
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260

FUNCIÓN DE REGRESIÓN MUESTRAL

Líneas de regresión basadas en dos muestras diferentes



JUNTA DE PLANIFICACIÓN

FUNCIÓN DE REGRESIÓN MUESTRAL

Un estimador es un método que dice cómo estimar el parámetro poblacional a partir de la información suministrada por la muestra

Un estimado es el valor númerico obtenido por el estimador

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

donde $\hat{Y}$ se lee «Y-gorro» o «Y-sombrero»

$\hat{Y}_i$ = estimador de $E(Y \mid X_i)$

$\hat{\beta}_1$ = estimador de β_1

$\hat{\beta}_2$ = estimador de β_2

FUNCIÓN DE REGRESIÓN MUESTRAL

La función de regresión muestral se puede expresar como:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

Para estimar $Y_i = ß_1 + ß_1 X_i + u_i$, podemos utilizar

$$Y_i = \hat{Y}_i + \hat{u}_i$$

y en términos de la FRP, puede ser expresada como

$$Y_i = E(Y \mid X_i) + u_i$$

MODELO DE REGRESIÓN: PROBLEMA DE ESTIMACIÓN

Estimar la función de regresión poblacional (FRP) con base en la función de regresión muestral (FRM)

Estudiaremos un método para calcular la función de regresión muestral:

- Método de mínimos cuadrados ordinarios

MODELO DE REGRESIÓN: MÉTODO DE MÍNIMOS CUADRADOS ORDINARIOS

Recuérdese la FRP de dos variables:

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

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned}$$

$$\begin{aligned} \hat{u}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \end{aligned}$$

MODELO DE REGRESIÓN: MÉTODO DE MÍNIMOS CUADRADOS ORDINARIOS

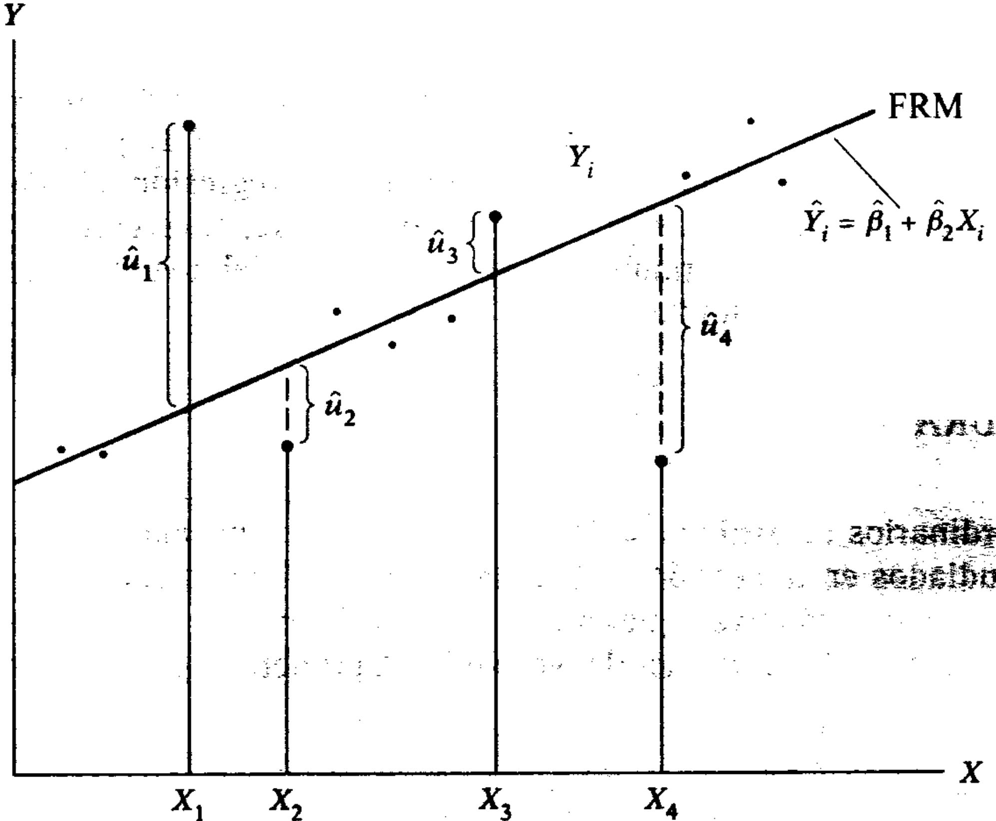


FIGURA 3.1
Criterio de mínimos cuadrados.

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2$$
$$= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

$$\sum \hat{u}_i^2 = f(\hat{\beta}_1, \hat{\beta}_2)$$

MODELO DE REGRESIÓN: MÉTODO DE MÍNIMOS CUADRADOS ORDINARIOS

TABLA 3.1
Determinación experimental de la FRM

	Y_i (1)	X_t (2)	$\hat{Y}_{1i}$ (3)	$\hat{u}_{1i}$ (4)	$\hat{u}^2_{1i}$ (5)	$\hat{Y}_{2i}$ (6)	$\hat{u}_{2i}$ (7)	$\hat{u}^2_{2i}$ (8)
	4	1	2.929	1.071	1.147	4	0	0
	5	4	7.000	−2.000	4.000	7	−2	4
	7	5	8.357	−1.357	1.841	8	−1	1
	12	6	9.714	2.286	5.226	9	3	9
Suma:	28	16		0.0	12.214		0	14

Notas: $\hat{Y}_{1i} = 1.572 + 1.357X_i$(i.e., $\hat{\beta}_1 = 1.572$ y $\hat{\beta}_2 = 1.357$)
$\hat{Y}_{2i} = 3.0 + 1.0X_i$(i.e., $\hat{\beta}_1 = 3$ y $\hat{\beta}_2 = 1.0$)
$\hat{u}_{1i} = (Y_i - \hat{Y}_{1i})$
$\hat{u}_{2i} = (Y_i - \hat{Y}_{2i})$

MODELO DE REGRESIÓN: MÉTODO DE MÍNIMOS CUADRADOS ORDINARIOS

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

JUNTA DE PLANIFICACIÓN

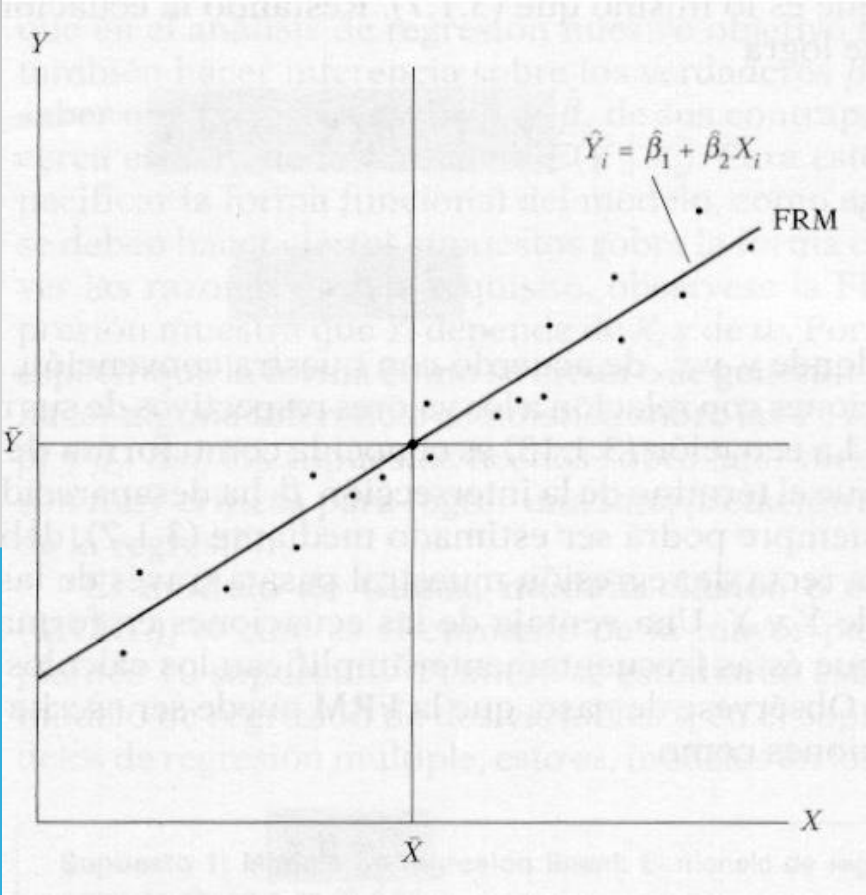
MODELO DE REGRESIÓN: MÉTODO DE MÍNIMOS CUADRADOS ORDINARIOS

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$
$$= \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$
$$= \frac{\sum x_i Y_i}{\sum X_i^2 - n\bar{X}^2}$$
$$= \frac{\sum X_i y_i}{\sum X_i^2 - n\bar{X}^2}$$

MODELO DE REGRESIÓN: MÉTODO DE MÍNIMOS CUADRADOS ORDINARIOS

$$\begin{aligned}\hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i \\ &= \bar{Y} + \hat{\beta}_2 (X_i - \bar{X})\end{aligned}$$



MODELO DE REGRESIÓN: SUPUESTOS DETRÁS DEL MÉTODO DE MÍNIMOS CUADRADOS

Supuesto 1: Modelo de regresión lineal. El modelo de regresión es **lineal en los parámetros**, como se observa en (2.4.2)

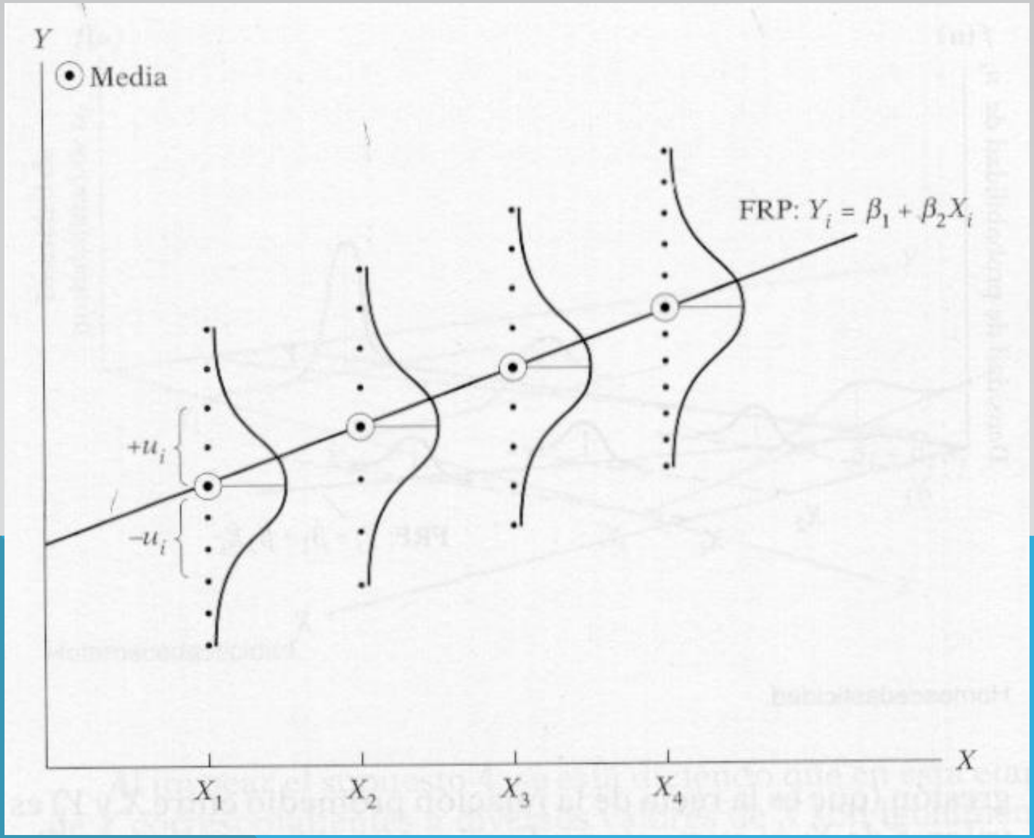
$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (2.4.2)$$

Supuesto 2: Los valores de X son fijos en muestreo repetido. Los valores que toma el regresor X son considerados fijos en muestreo repetido. Más técnicamente, se supone *no estocástica.*

Supuesto 3: El valor medio de la perturbación u_i es igual a cero. Dado el valor de X, la media, o el valor esperado del término aleatorio de perturbación u_i es cero. Técnicamente, el valor de la media condicional de u_i es cero. Simbólicamente, se tiene

$$E(u_i \mid X_i) = 0 \qquad (3.2.1)$$

MODELO DE REGRESIÓN:
SUPUESTOS DETRÁS DEL MÉTODO MÍNIMOS CUADRADOS



JUNTA DE PLANIFICACIÓN

MODELO DE REGRESIÓN: SUPUESTOS DETRÁS DEL MÉTODO DE MÍNIMOS CUADRADOS

Supuesto 4: Homoscedasticidad o igual varianza de u_i. Dado el valor de X, la varianza de u_i es la misma para todas las observaciones. Esto es, las varianzas condicionales de u_i son idénticas. Simbólicamente, se tiene que

$$\begin{aligned} \textbf{var}(u_i \mid X_i) &= E[u_i - E(u_i) \mid X_i]^2 \\ &= E(u_i^2 \mid X_i) \quad \text{por el supuesto 3} \\ &= \sigma^2 \end{aligned} \tag{3.2.2}$$

donde **var** significa varianza.

Modelo de Regresión: Supuestos detrás del Método de Mínimos Cuadrados

MODELO DE REGRESIÓN: SUPUESTOS DETRÁS DEL MÉTODO DE MÍNIMOS CUADRADOS

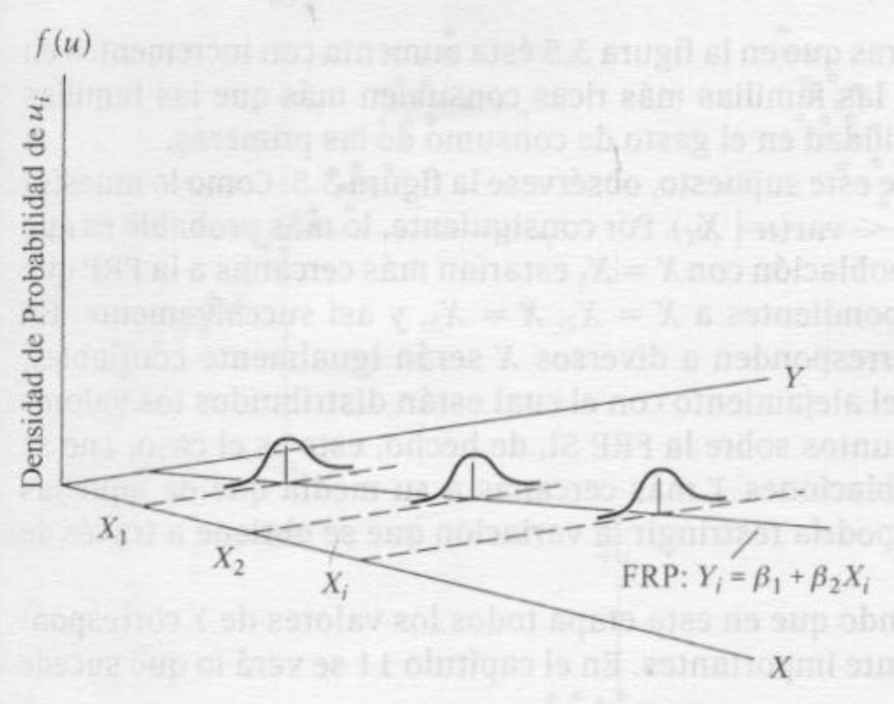


FIGURA 3.4
Homoscedasticidad.

MODELO DE REGRESIÓN: SUPUESTOS DETRÁS DEL MÉTODO DE MÍNIMOS CUADRADOS

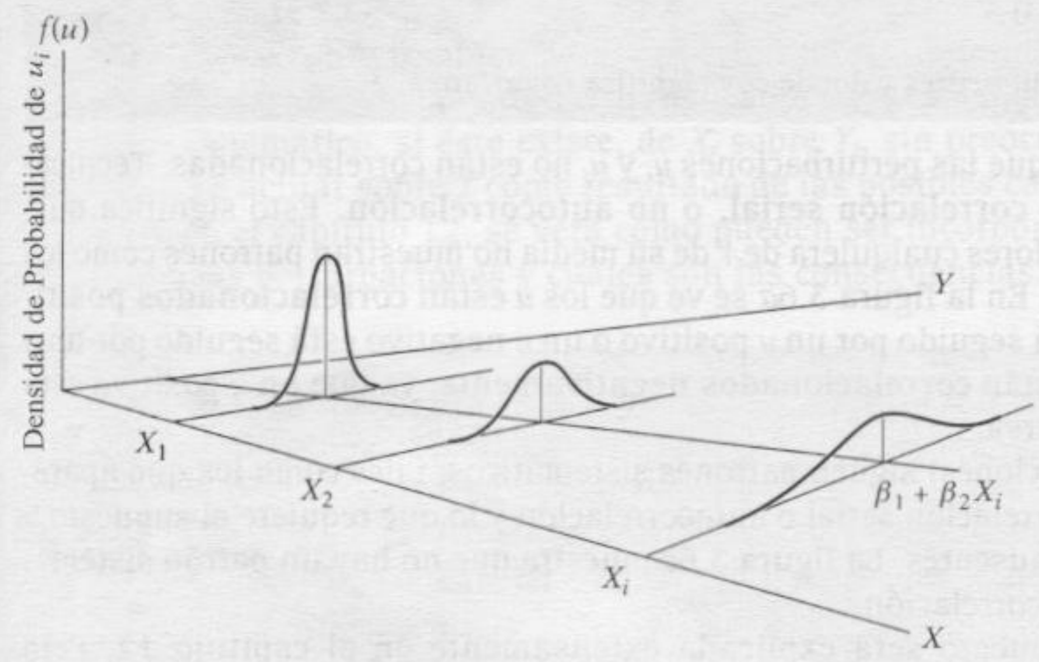


FIGURA 3.5
Heteroscedasticidad.

MODELO DE REGRESIÓN: SUPUESTOS DETRÁS DEL MÉTODO DE MÍNIMOS CUADRADOS

Supuesto 5: No autocorrelación entre las perturbaciones. Dados dos valores cualquiera de X, X_i y X_j ($i \neq j$), la correlación entre dos u_i y u_j cualquiera ($i \neq j$) es cero. Simbólicamente,

$$\begin{aligned} \mathbf{cov}(u_i, u_j \mid X_i, X_j) &= E[u_i - E(u_i) \mid X_i][u_j - E(u_j) \mid X_j] \\ &= E(u_i \mid X_i)(u_j \mid X_j) \qquad \text{¿por qué?} \\ &= 0 \qquad\qquad (3.2.5) \end{aligned}$$

donde i y j son dos observaciones diferentes y donde **cov** significa **covarianza**.

MODELO DE REGRESIÓN: SUPUESTOS DETRÁS DEL MÉTODO DE MÍNIMOS CUADRADOS

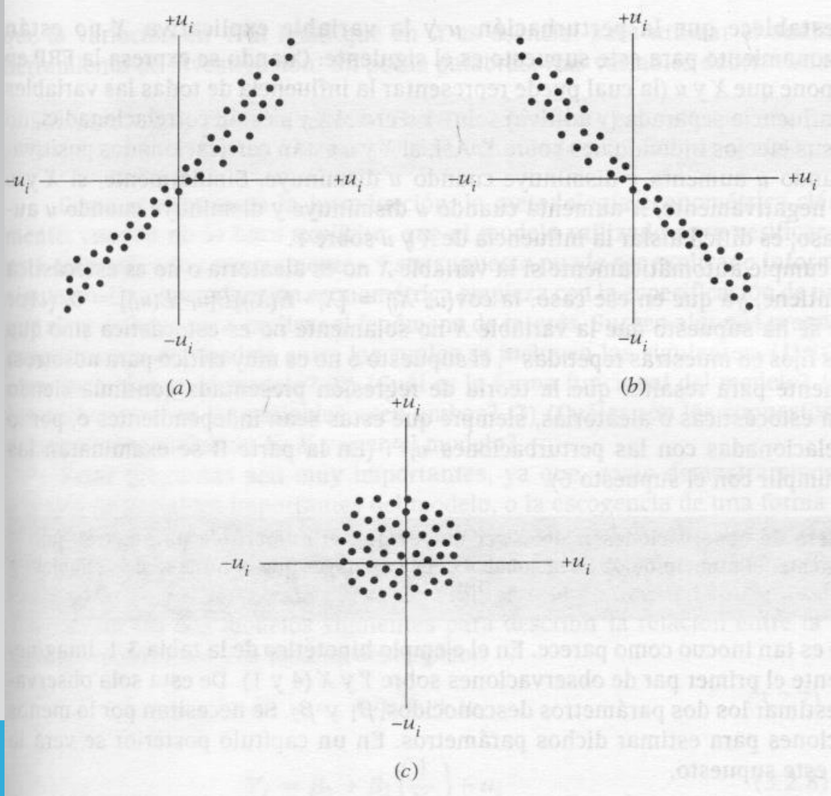


FIGURA 3.6
Patrones de correlación entre perturbaciones. (*a*) correlación serial positiva; (*b*) correlación serial negativa; (*c*) correlación cero.

MODELO DE REGRESIÓN: SUPUESTOS DETRÁS DEL MÉTODO DE MÍNIMOS CUADRADOS

Supuesto 6: La covarianza entre u_i y X_i es cero, o $E(u_i X_i) = 0$. Formalmente,

$$
\begin{aligned}
\mathbf{cov}(u_i, X_i) &= E[u_i - E(u_i)][X_i - E(X_i)] \\
&= E[u_i(X_i - E(X_i))], \qquad \text{puesto que } E(u_i) = 0 \\
&= E(u_i X_i) - E(X_i)E(u_i), \text{ puesto que } E(X_i) \text{ es no estocástica} \\
&= E(u_i X_i), \quad \text{puesto que } E(u_i) = 0 \\
&= 0, \qquad \text{por supuesto}
\end{aligned}
\tag{3.2.6}
$$

MODELO DE REGRESIÓN: SUPUESTOS DETRÁS DEL MÉTODO DE MÍNIMOS CUADRADOS

Supuesto 7: El número de observaciones *n* debe ser mayor que el número de parámetros por estimar. Alternativamente, el número de observaciones *n* debe ser mayor que el número de variables explicativas.

Supuesto 8: Variabilidad en los valores de *X*. No todos los valores de *X* en una muestra dada deben ser iguales. Técnicamente, var(*X*) debe ser un número positivo finito[14].

Supuesto 9: El modelo de regresión está correctamente especificado. Alternativamente, no hay un **sesgo de especificación o error** en el modelo utilizado en el análisis empírico.

Supuesto 10: No hay multicolinealidad perfecta. Es decir, *no hay relaciones perfectamente lineales entre las variables explicativas.*

MODELO DE REGRESIÓN: COEFICIENTE DE DETERMINACIÓN R^2

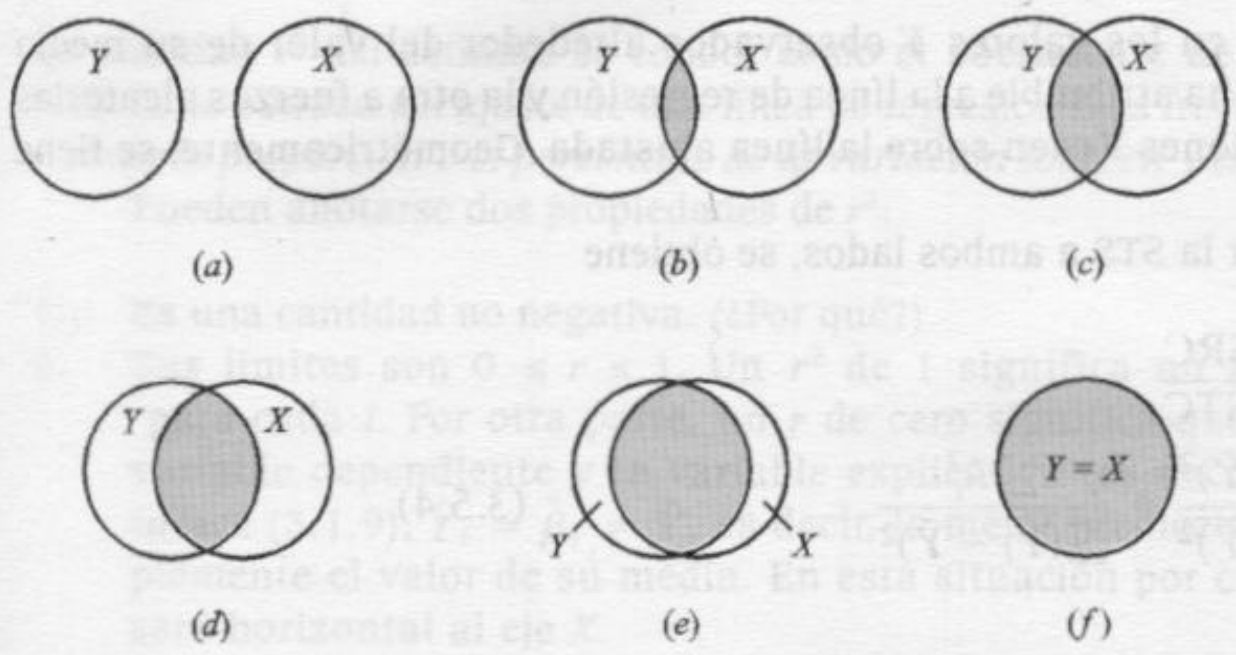


FIGURA 3.9
Enfoque Ballentine de r^2 :(a) $r^2 = 0$; (f) $r^2 = 1$.

MODELO DE REGRESIÓN: COEFICIENTE DE DETERMINACIÓN R²

$$1 = \frac{\text{SEC}}{\text{STC}} + \frac{\text{SRC}}{\text{STC}}$$
$$= \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} + \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2}$$

$$r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\text{SEC}}{\text{STC}} \qquad (3.5.5)$$

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2}$$
$$= 1 - \frac{\text{SRC}}{\text{STC}}$$

MODELO DE REGRESIÓN: COEFICIENTE DE DETERMINACIÓN R^2

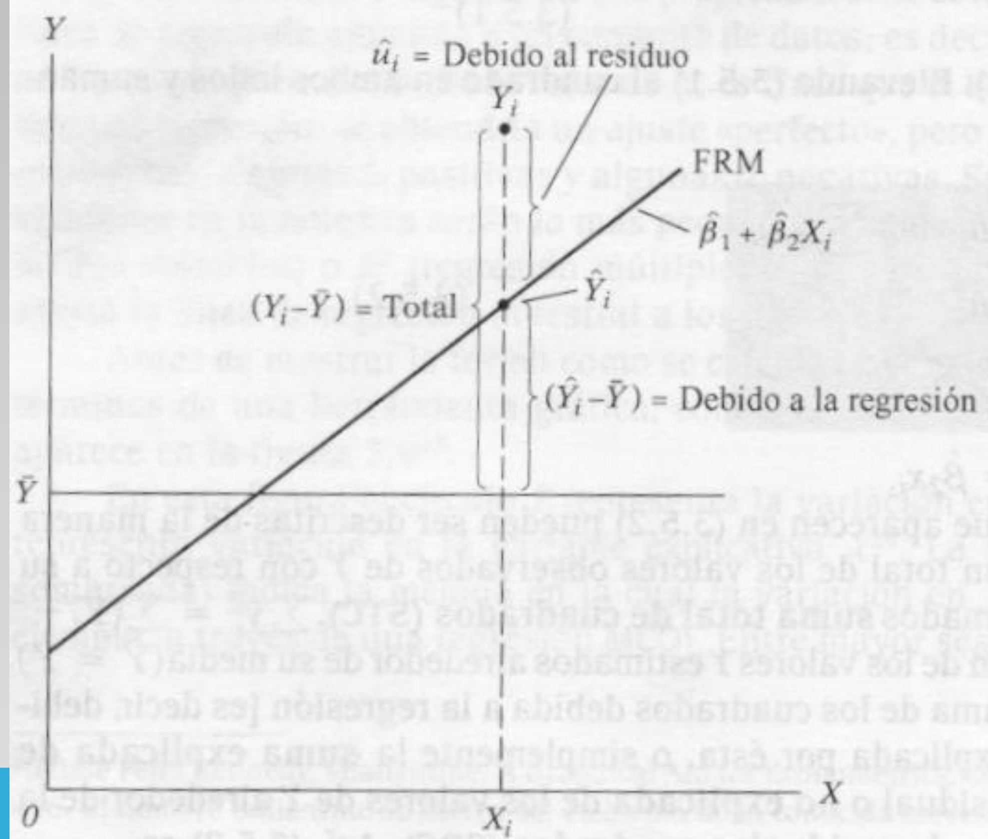


FIGURA 3.10
Partición de la variación de Y_i en dos componentes.

JUNTA DE PLANIFICACIÓN

MODELO DE REGRESIÓN: EJEMPLO

TABLA 3.2
Datos hipotéticos sobre el gasto de consumo familiar semanal Y y el ingreso familiar semanal *X*

Y($)	*X*($)
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

Modelo de Regresión: Ejemplo

MODELO DE REGRESIÓN: EJEMPLO NUMÉRICO

TABLA 3.3
Datos primarios basados en la Tabla 3.2

	Y_i (1)	X_i (2)	Y_iX_i (3)	X_i^2 (4)	$x_i = X_i - \bar{X}$ (5)	$y_i = Y_i - \bar{Y}$ (6)	x_i^2 (7)	x_iy_i (8)	$\hat{Y}_i$ (9)	$\hat{u}_i = Y_i - \hat{Y}_i$ (10)	$\hat{Y}_i\hat{u}_i$ (11)
	70	80	5600	6400	−90	−41	8100	3690	65.1818	4.8181	314.0524
	65	100	6500	10000	−70	−46	4900	3220	75.3636	−10.3636	−781.0382
	90	120	10800	14400	−50	−21	2500	1050	85.5454	4.4545	381.0620
	95	140	13300	19600	−30	−16	900	480	95.7272	−0.7272	−69.6128
	110	160	17600	25600	−10	−1	100	10	105.9090	4.0909	433.2631
	115	180	20700	32400	10	4	100	40	116.0909	−1.0909	−126.6434
	120	200	24000	40000	30	9	900	270	125.2727	−6.2727	−792.0708
	140	220	30800	48400	50	29	2500	1450	136.4545	3.5454	483.7858
	155	240	37200	57600	70	44	4900	3080	145.6363	8.3636	1226.4073
	150	260	39000	67600	90	39	8100	3510	156.8181	−6.8181	−1069.2014
Suma	1110	1700	205500	322000	0	0	33000	16800	1109.9995	0	0.0040
									≑ 1110.0		≑ 0.0
Media	111	170	nc	nc	0	0	nc	nc	110	0	0

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = 16{,}800/33{,}000 = 0.5091$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X} = 111 - 0.5091(170) = 24.4545$$

Notas: ≑ simboliza «aproximadamente igual a»; nc significa «no calculado».

MODELO DE REGRESIÓN: EJEMPLO

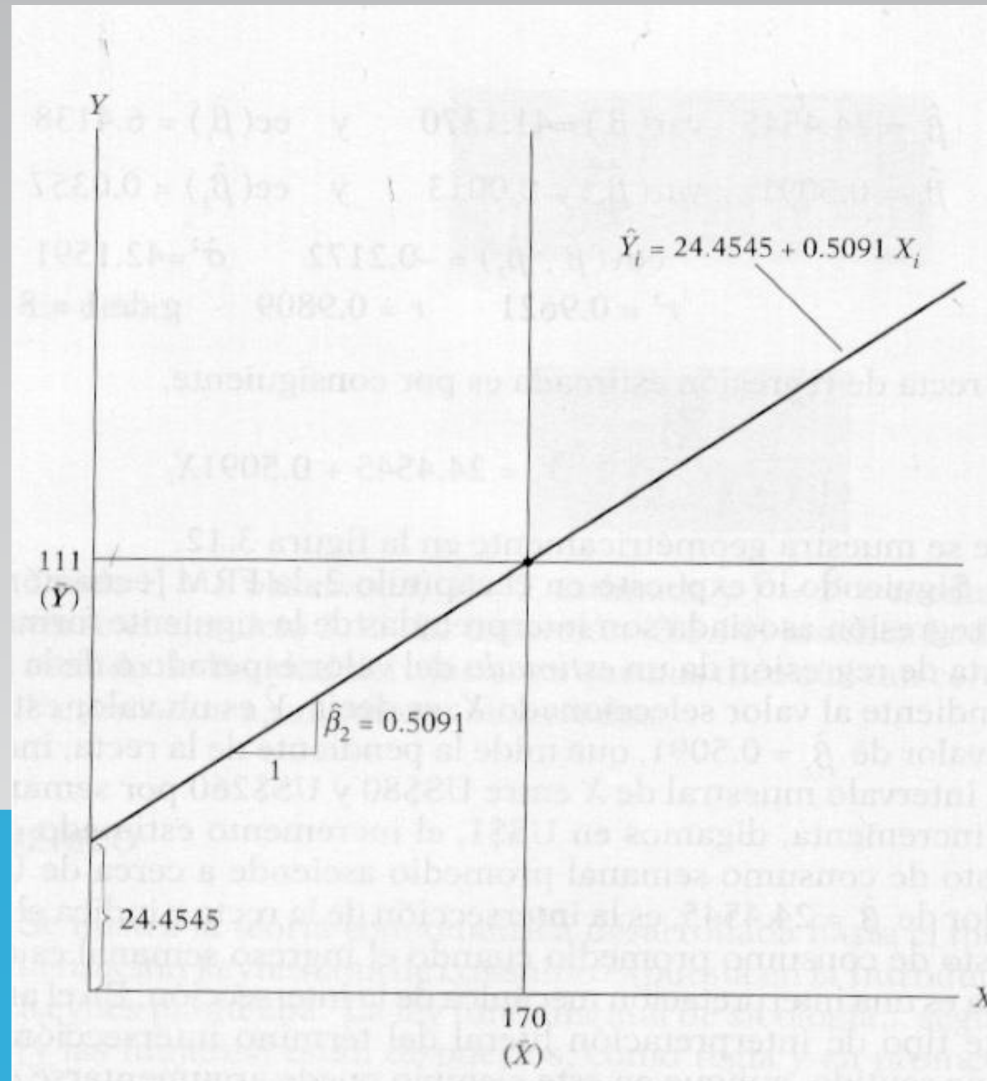
$$
\begin{aligned}
\hat{\beta}_1 &= 24.4545 \quad & \operatorname{var}(\hat{\beta}_1) &= 41.1370 \quad & \text{y} \quad & \operatorname{ee}(\hat{\beta}_1) = 6.4138 \\
\hat{\beta}_2 &= 0.5091 \quad & \operatorname{var}(\hat{\beta}_2) &= 0.0013 \quad & \text{y} \quad & \operatorname{ee}(\hat{\beta}_2) = 0.0357
\end{aligned}
\tag{3.6.1}
$$

$$
\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) = -0.2172 \qquad \hat{\sigma}^2 = 42.1591
$$

$$
r^2 = 0.9621 \qquad r = 0.9809 \qquad \text{g de l} = 8
$$

$$
\hat{Y}_i = 24.4545 + 0.5091X_i
$$

MODELO DE REGRESIÓN: EJEMPLO



ENFOQUE DE PRUEBA DE SIGNIFICANCIA

Procedimiento para verificar falsedad de hipótesis nula

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{ee}(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}}$$

$$\Pr\left[-t_{\alpha/2} \le \frac{\hat{\beta}_2 - \beta_2^*}{\text{ee}(\hat{\beta}_2)} \le t_{\alpha/2}\right] = 1 - \alpha$$

$$\Pr(0.2177 \le \hat{\beta}_2 \le 0.3823) = 0.95$$

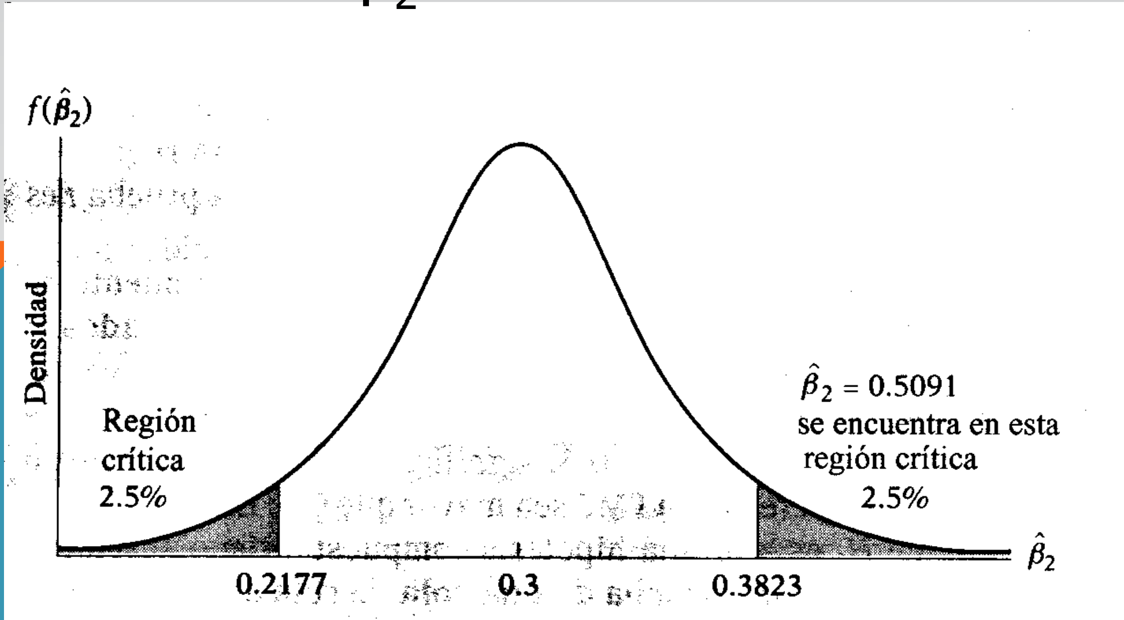
$$t = \frac{0.5091 - 0.3}{0.0357} = 5.86$$

ENFOQUE DE PRUEBA DE SIGNIFICANCIA

Intervalo de confianza para β_2

$$\Pr[\beta_2^* - t_{\alpha/2}\,\mathrm{ee}(\hat{\beta}_2) \leq \hat{\beta}_2 \leq \beta_2^* + t_{\alpha/2}\,\mathrm{ee}(\hat{\beta}_2)] = 1 - \alpha$$

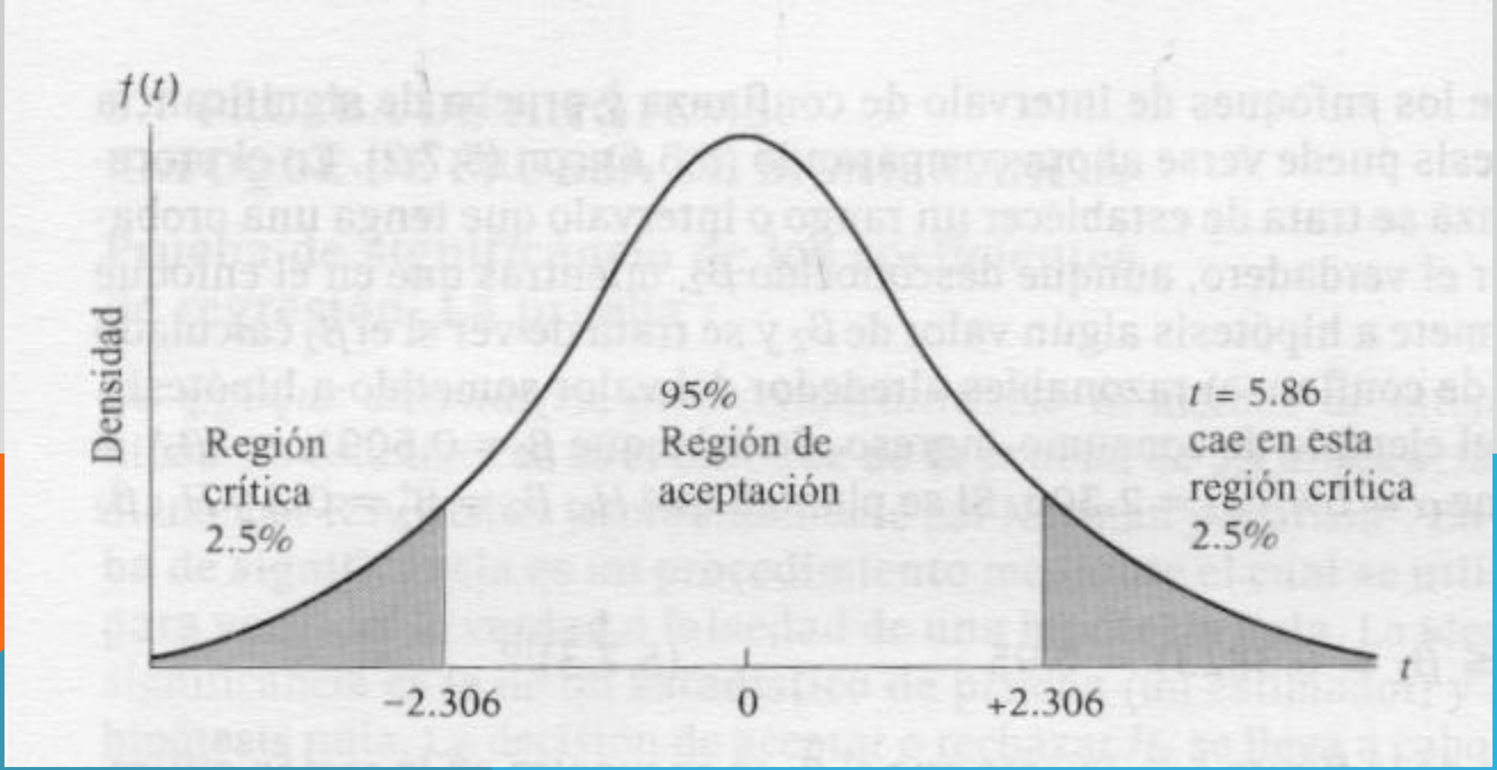
Distribución β_2



JUNTA DE PLANIFICACIÓN

ENFOQUE DE PRUEBA DE SIGNIFICANCIA

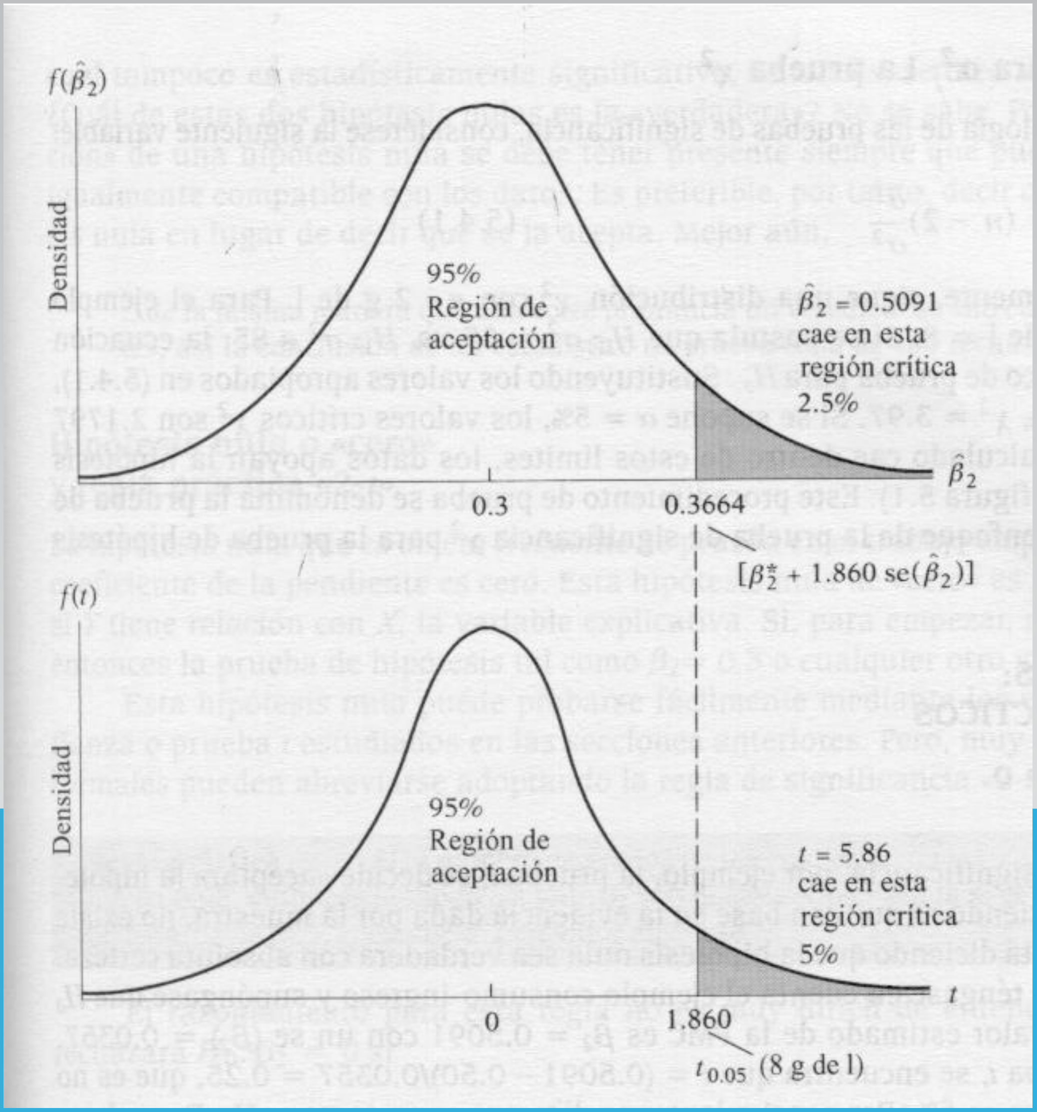
Distribución t



JUNTA DE PLANIFICACIÓN

ENFOQUE DE PRUEBA DE SIGNIFICANCIA

Diferencia entre distribución $\hat{\beta}_2$ y distribución t



JUNTA DE PLANIFICACIÓN

ENFOQUE DE PRUEBA DE SIGNIFICANCIA

Cómo expresar H_o y H_A y reglas de decisión

TABLA 5.1
La prueba *t* de significancia: Reglas de decisión

Tipo de hipótesis	H_0: La hipótesis nula	H_1: La hipótesis alterna	Regla de decisión: Rechazar H_0 si
Dos colas	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\|t\| > t_{\alpha/2,\, g\ de\ l}$
Cola derecha	$\beta_2 \leq \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_{\alpha,\, g\ de\ l}$
Cola izquierda	$\beta_2 \geq \beta_2^*$	$\beta_2 < \beta_2^*$	$t < -t_{\alpha,\, g\ de\ l}$

HIPÓTESIS NULA O "CERO" Y REGLA PRÁCTICA 2-T

Regla práctica «2-*t*». Si el número de grados de libertad es 20 o más y si α, el nivel de significancia, se fija en 0.05, entonces la hipótesis nula $\beta_2 = 0$ puede ser rechazada si el valor de t [$= \hat{\beta}_2/\text{se}(\hat{\beta}_2)$] calculado a partir de (5.3.2) excede a 2 en valor absoluto.

$$t = \hat{\beta}_2/\text{ee}(\hat{\beta}_2) > t_{\alpha/2} \quad \text{cuando } \hat{\beta}_2 > 0$$

$$t = \hat{\beta}_2/\text{ee}(\hat{\beta}_2) < -t_{\alpha/2} \quad \text{cuando } \hat{\beta}_2 < 0$$

$$|t| = \left|\frac{\hat{\beta}_2}{\text{ee}(\hat{\beta}_2)}\right| > t_{\alpha/2}$$

$$|t| = \left|\frac{\hat{\beta}_2}{\text{ee}(\hat{\beta}_2)}\right| > t_{\alpha}$$