UNIVERSIDAD TORCUATO DI TELLA
ESCUELA DE NEGOCIOS

Probabilidad y Estadística

Prof. Sebastián Auguste
sauguste@utdt.edu

Capítulo 6

Regresión Lineal

Referencia: Defusco, cap 3

Análisis de Correlación

- Un análisis de correlación expresa la relación entre dos series de datos utilizando un único número.
- El coeficiente de correlación es una medida de cuán estrechamente relacionadas con dos series de datos.
- El coeficiente de correlación mide la asociación lineal entre dos variables.

Coeficiente de Correlación Muestral

- Covarianza muestral

$$S_{XY} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)}$$

- Coeficiente de Correlación Muestral

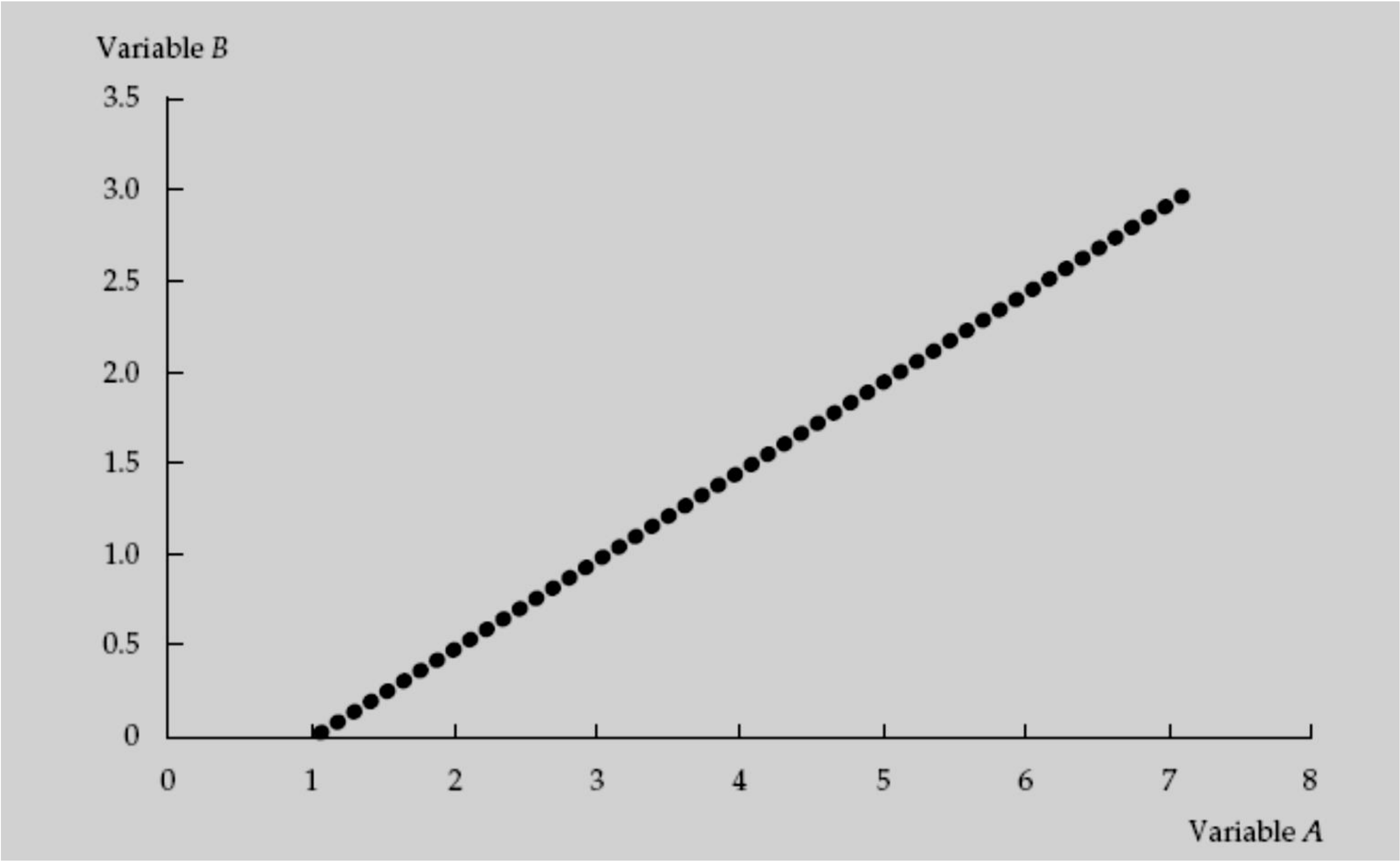
$$r = \frac{s_{XY}}{s_X s_Y}$$

- Coeficiente de Correlación Poblacional

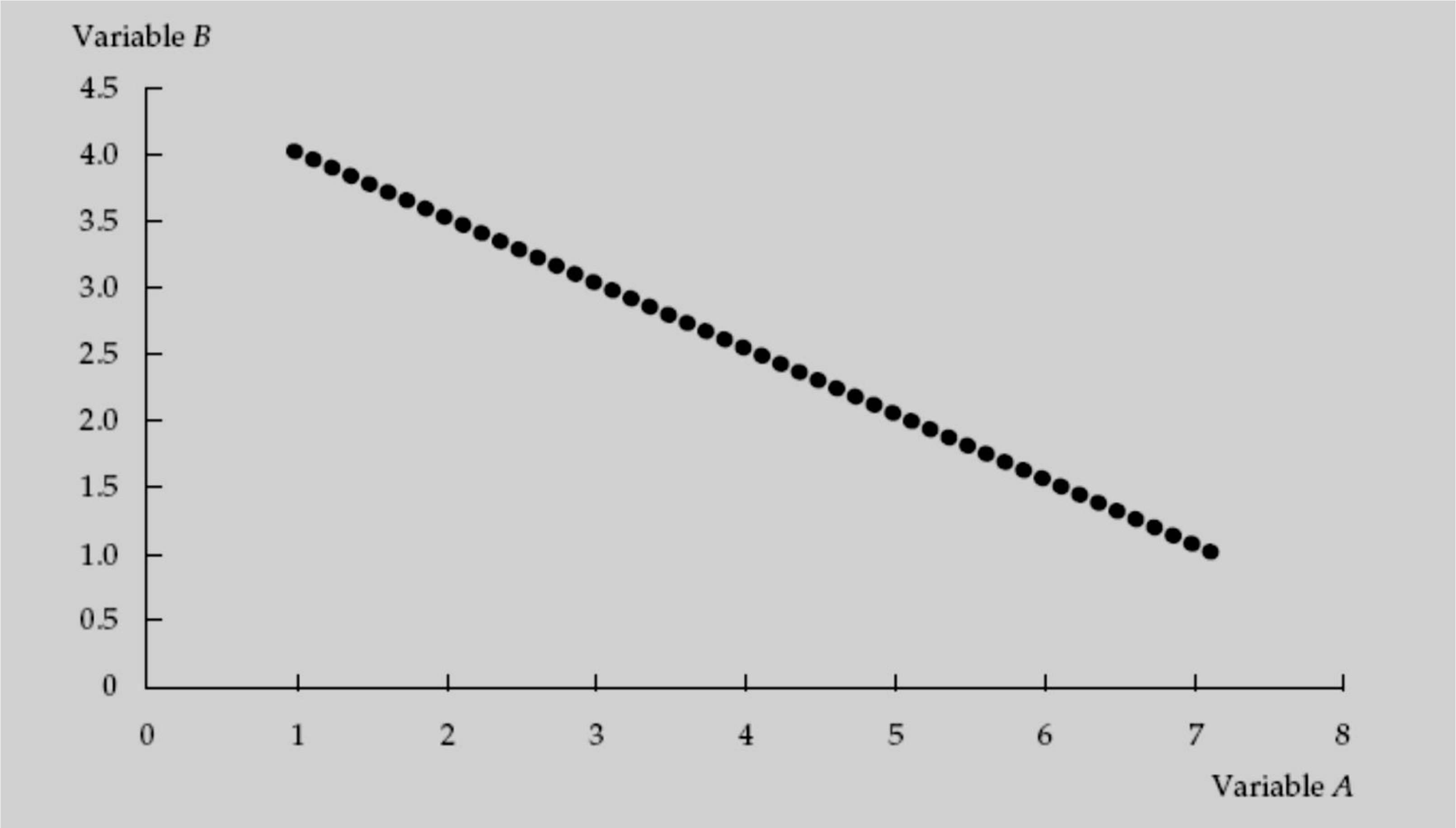
$$\rho = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y}$$

- ρ= 1 implica todos los puntos sobre una línea con pendiente +
- ρ= -1 idem pero pendiente negativa
- ρ=0 no existe una correlación lineal entre las variables

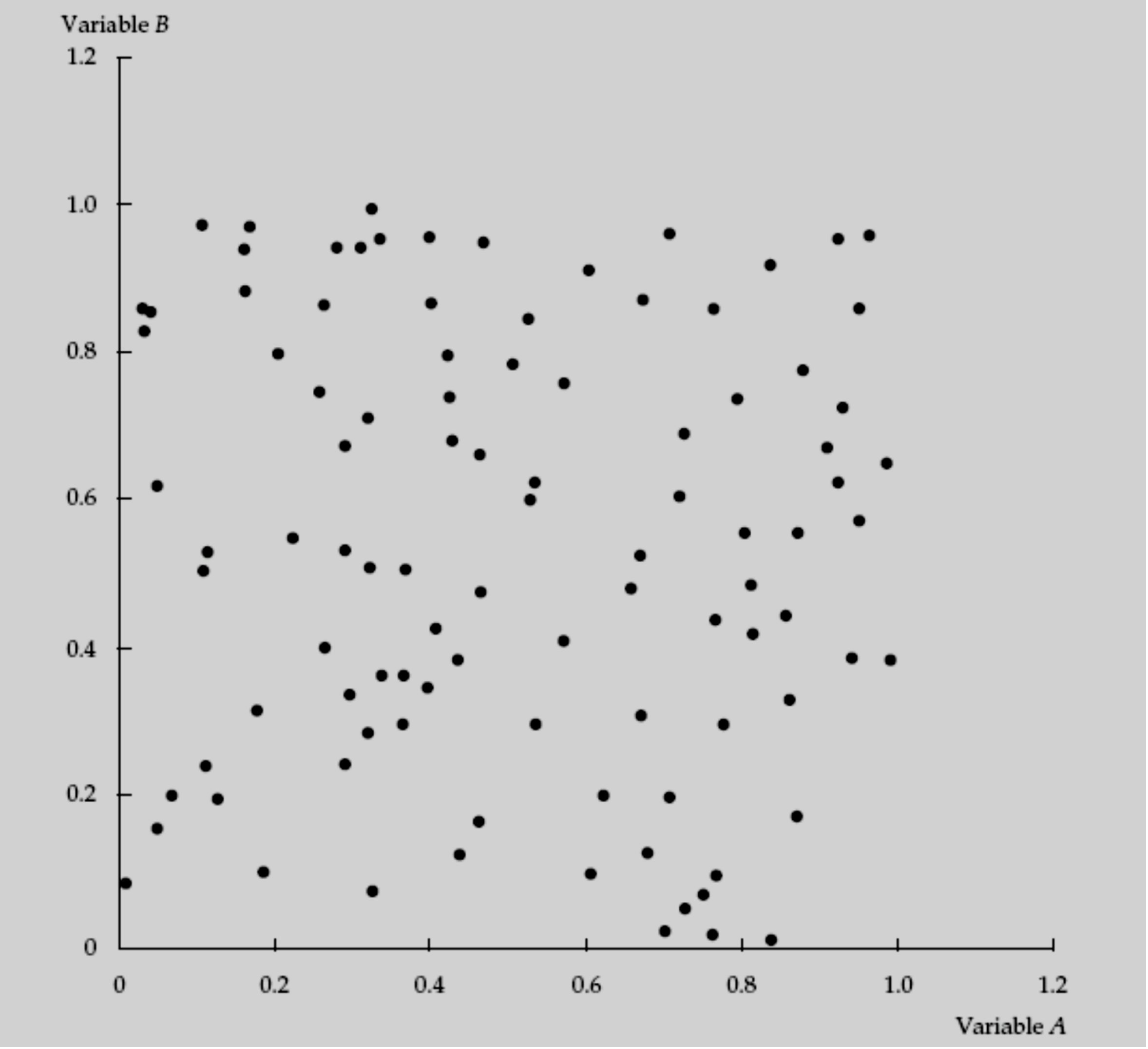
Variables con Correlación Positiva Perfecta



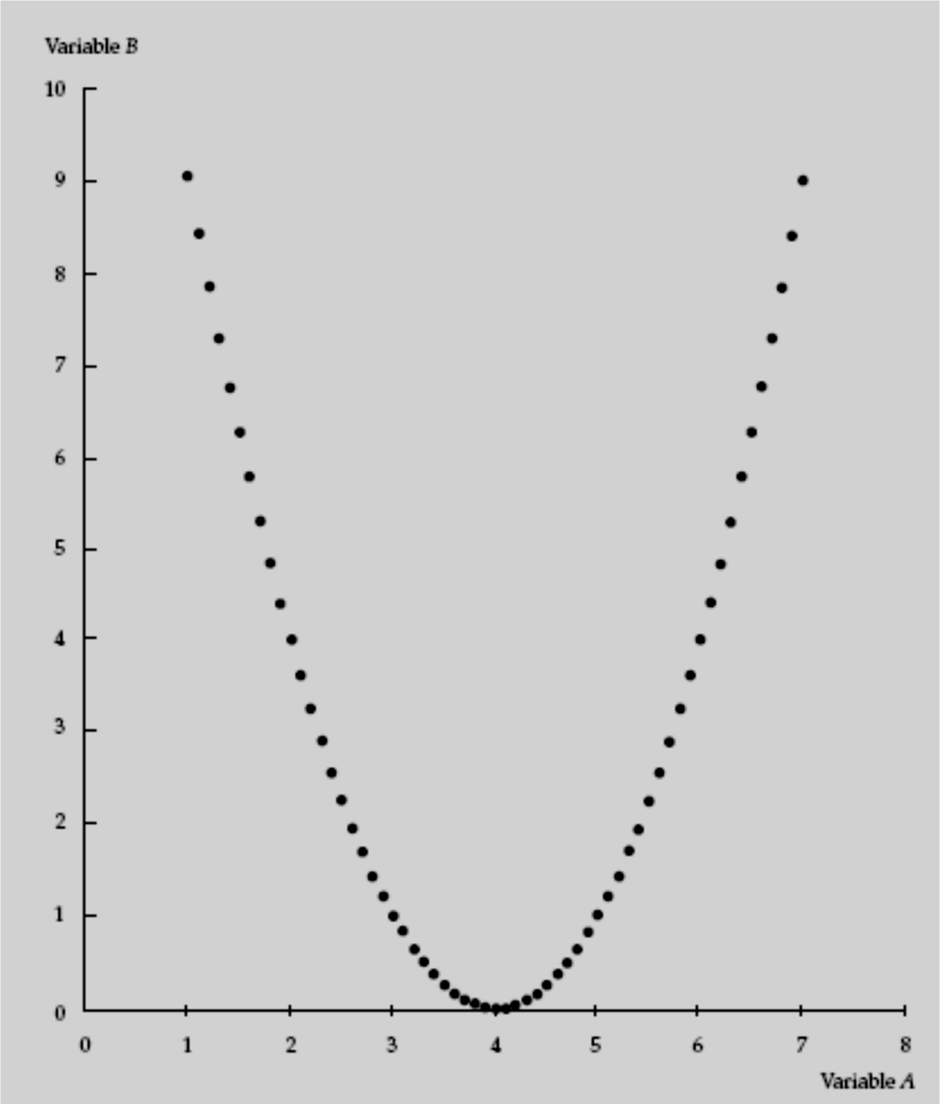
Variables con Correlación Negativa Perfecta



Variables con Correlación Cero



Variables con una fuerte asociación no lineal



DESCRIPTIVE STATISTICS Monthly Returns 1-1990 a 3-2011					
	MERVAL $	MERVAL USD	S&P500 USD	BOVESPA R$	BOVESPA USD
Mean	3.4%	1.6%	0.6%	8.1%	2.6%
Median	1.6%	1.7%	1.1%	4.2%	2.2%
Maximum	279.6%	119.0%	11.2%	114.3%	79.7%
Minimum	-39.1%	-42.3%	-16.8%	-56.2%	-69.5%
Std. Dev.	0.225	0.141	0.044	0.209	0.162
Skewness	7.973	2.271	-0.631	1.826	0.378
Kurtosis	93.735	22.513	4.091	8.654	7.016
Accumulated Return		557%	289%		2466%
Return on 1 USD in 21 years		6.6	3.9		25.7
Jarque-Bera	89821	4248	29	479	177
Probability	0	0	0	0	0
Sum	8.64	4.15	1.60	20.66	6.67
Sum Sq. Dev.	12.80	5.02	0.48	11.05	6.61
Observations	254	254	254	254	254

Correlaciones entre activos

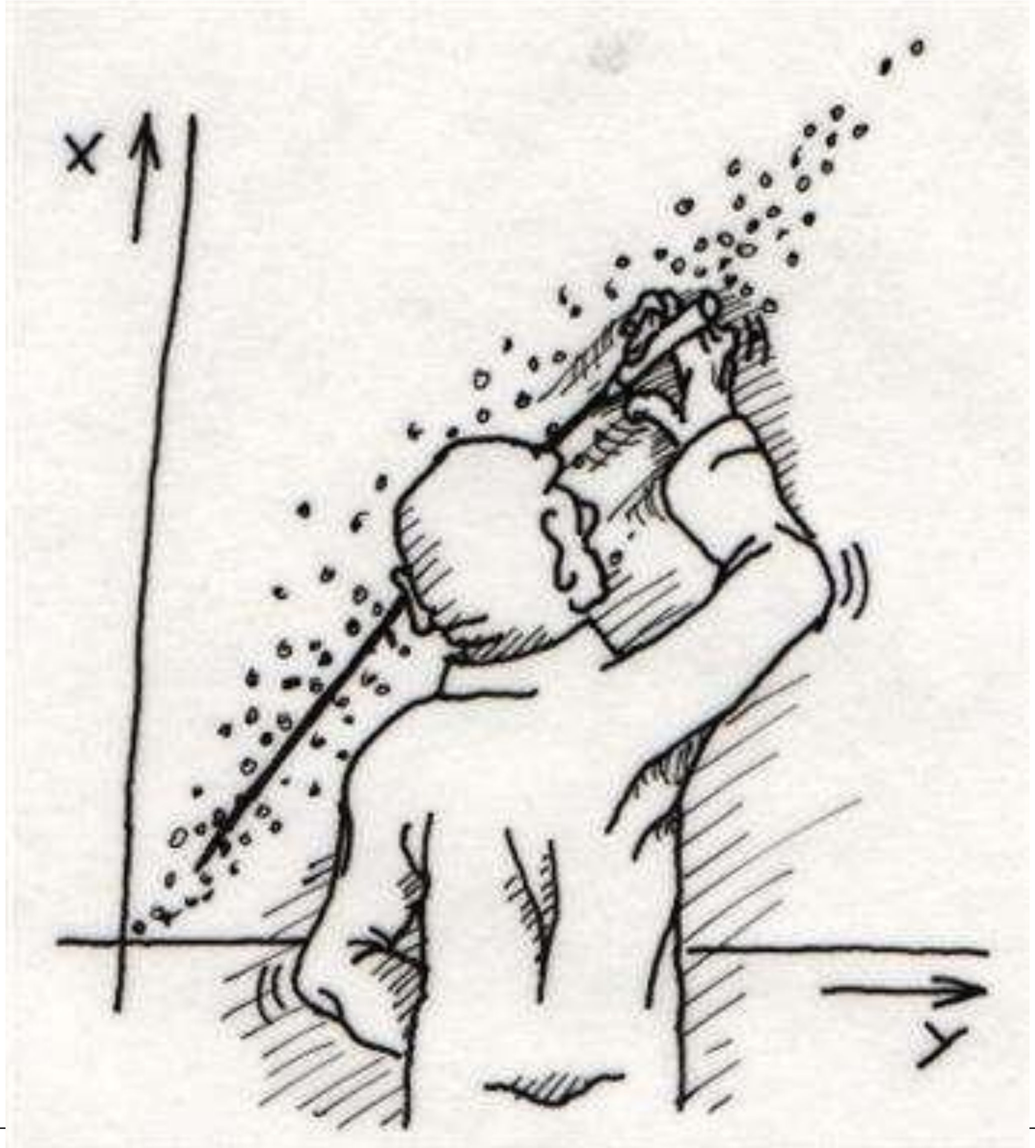
CORRELATION MATRIX				1-1990 a 3-2011	
	MERVAL $	MERVAL USD	S&P500 USD	BOVESPA R$	BOVESPA USD
MERVAL $	1.00	0.64	0.22	0.34	0.25
MERVAL USD	0.64	1.00	0.35	0.26	0.34
S&P500 USD	0.22	0.35	1.00	0.28	0.45
BOVESPA R$	0.34	0.26	0.28	1.00	0.76
BOVESPA USD	0.25	0.34	0.45	0.76	1.00

Test de Hipótesis

- Si la población es Normal, bajo Ho de cero correlación:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \approx t(n-2)$$

Fitting line



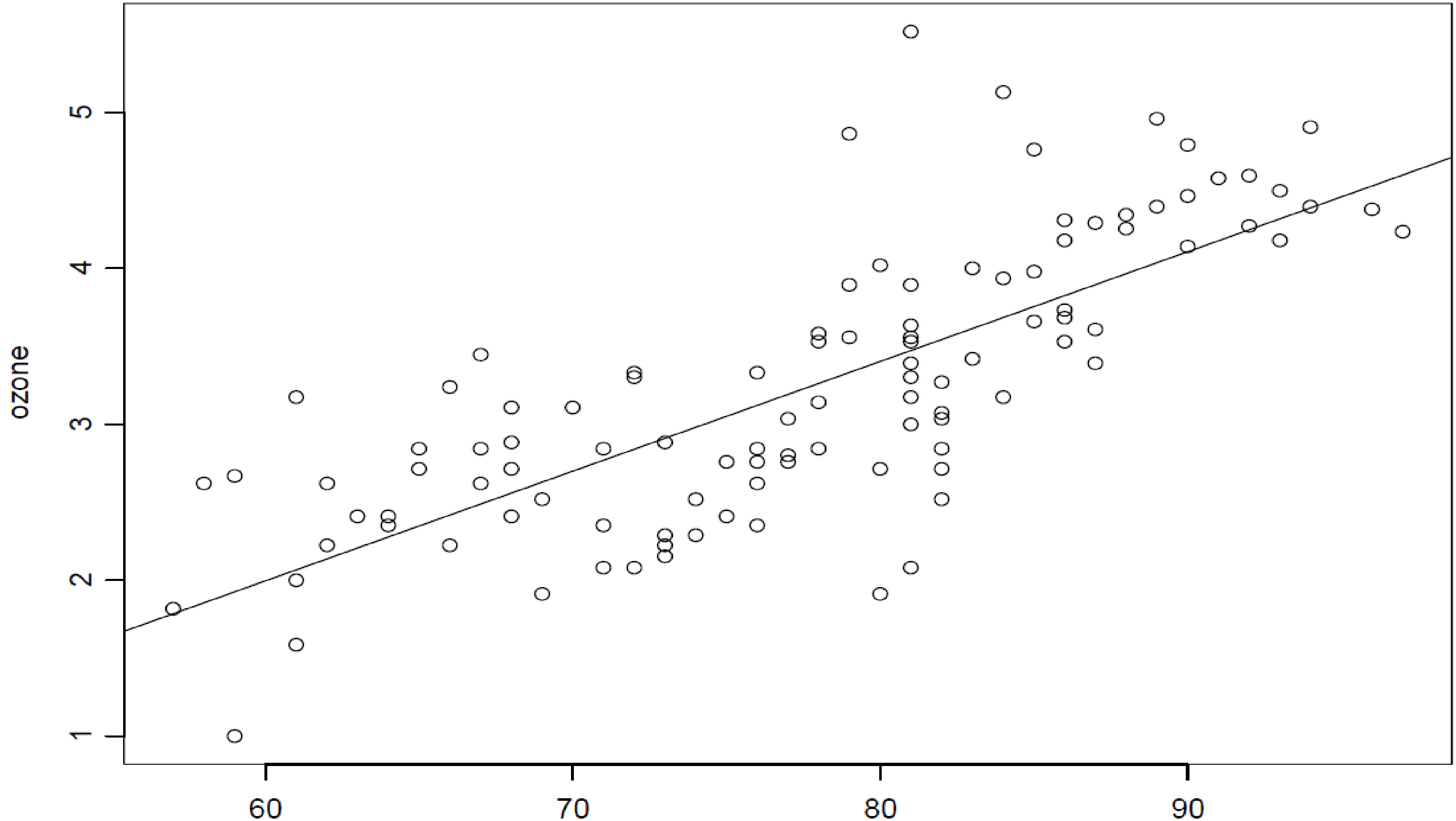
Algunas de las muchas aplicaciones

- Evaluar tendencias y hacer predicciones (e.g. ventas)
- Analizar el impacto de cambio en precios
- Identificar valor de los atributos de un producto
- Analizar el riesgo
- Detectar problemas de calidad
- Auditoría y Benchmarking
- Encontrar evidencia de cheating
- Uso Legal: antitrust, daños y litigios

Algunas de las muchas aplicaciones en academia

- Permite
  - realizar análisis multivariado (entender como muchos factores afectan a una variable).
  - probar teorías alternativas (que por ejemplo tienen distinta incidencia en cómo una variable afecta a otra)
  - encontrar relaciones empíricas sobre las cuales construir teorías
  - Entender mejor los datos

Fitting line (regresión lineal)



El coeficiente de correlación nos dice si los puntos están todos cerca de una relación lineal, pero no nos dice cual es esa relación

Regresión lineal

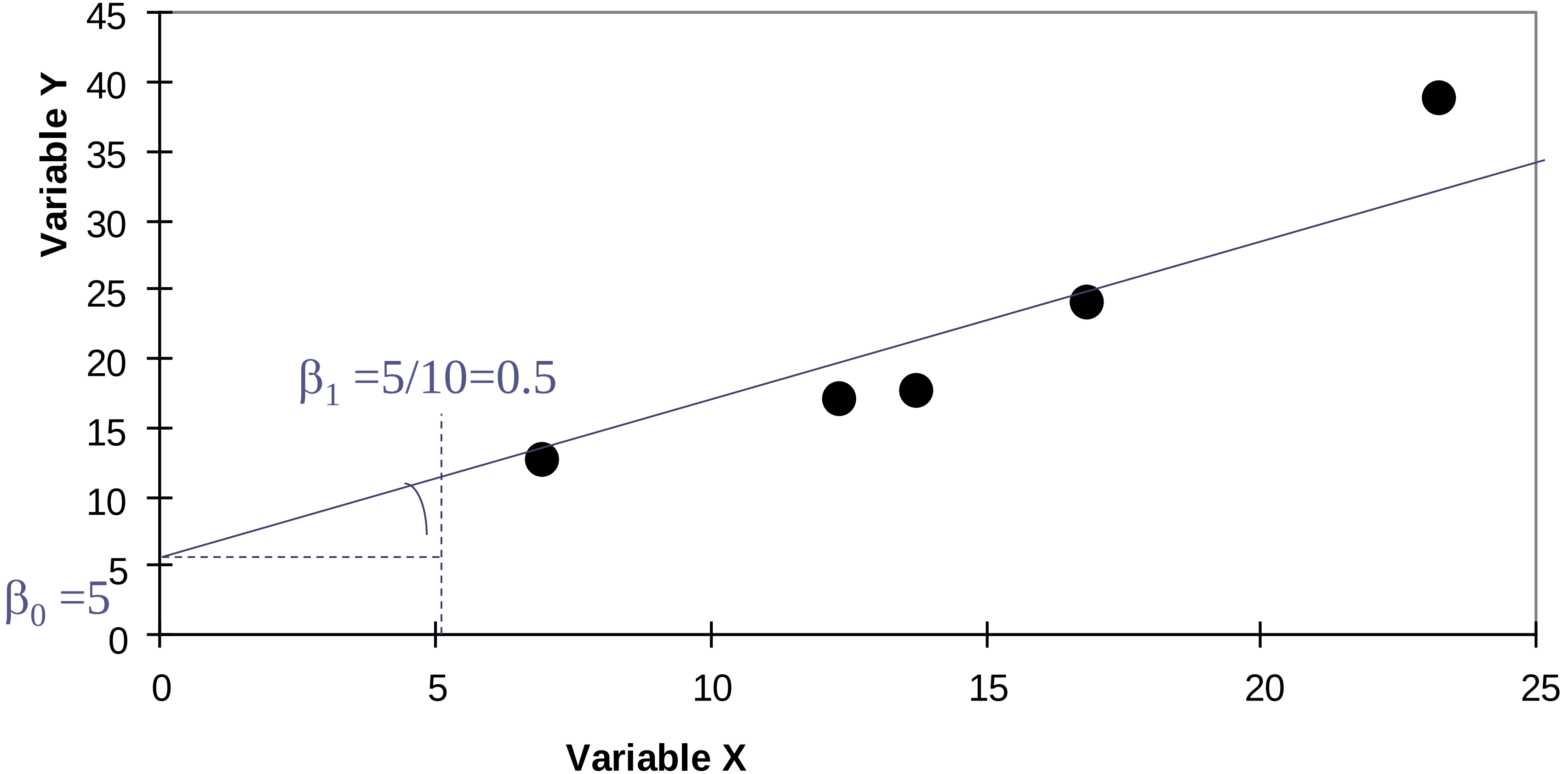
- El método más común es ajustar una recta a la nueve de puntos usando el método de Minimos Cuadrados Clásicos.

- Técnicamente

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

✓ β_0 es la ordenada al origen de la recta
✓ β_1 la pendiente
✓ Subindide "i" indica la observación (persona, firma, etc).
✓ El sombrero indica "predicción"

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$



- El problema de estimaci☻n es encontrar valores para β_0 β_1 basado en los datos de una muestra.
- Muestra puede
  - Serie de Tiempo (T)
  - Cross Section (N)
  - Panel (combina ambas)
- Ejemplo
  - CAPM para Tenaris en base a Merval (time series)
  - Datos del último balance de las firmas en bolsa para ver como D/E afecta V (cross section)
  - Encuesta a personas para saber si les gusta mi producto (cross section)

El Método de Mínimos Cuadrados

(Otras siglas: OLS, MICO, MCC)

Llamemos a Cacho "o"

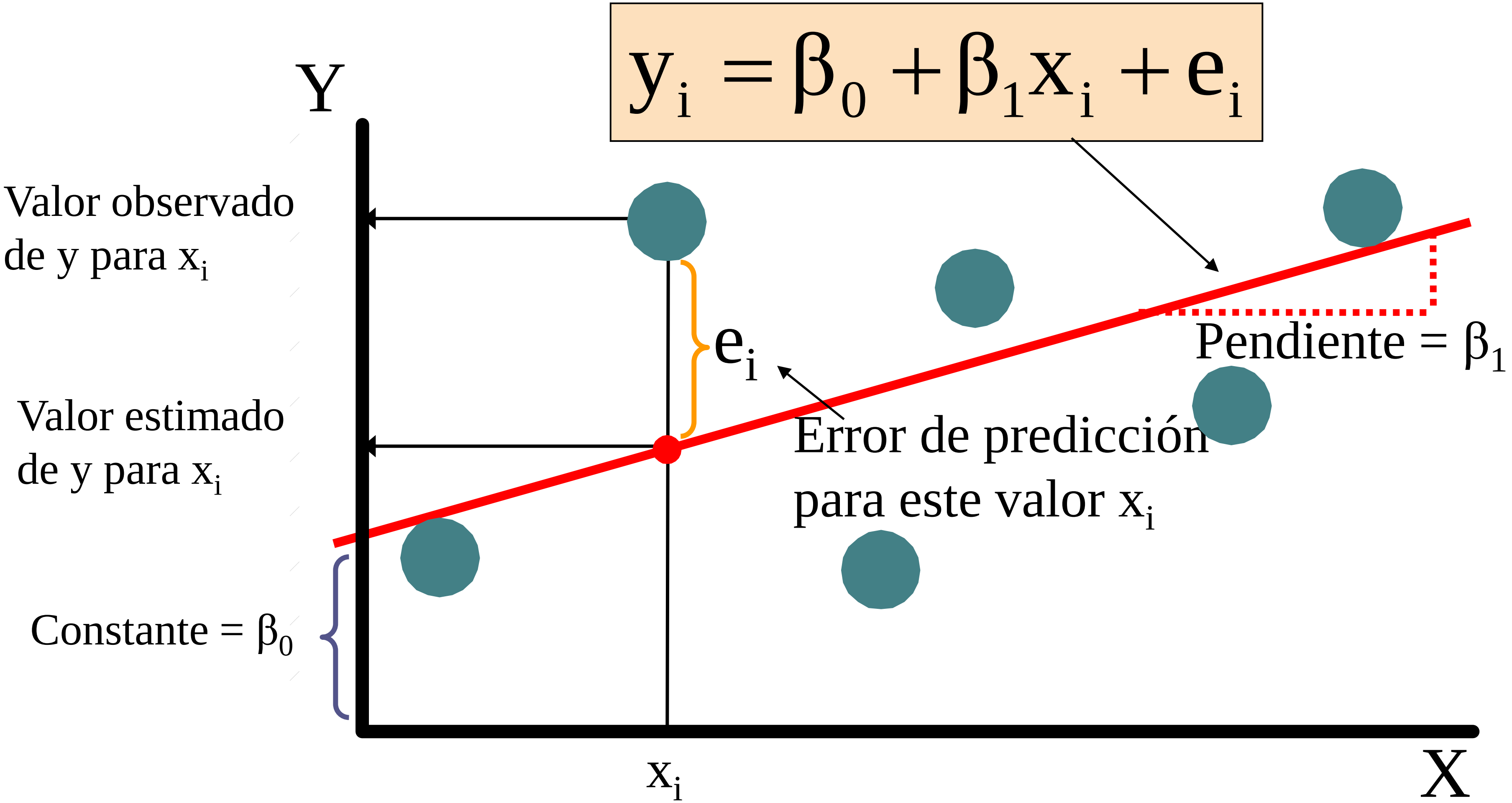
Predicción para "o" $\longrightarrow$ $\hat{y}_o = \beta_0 + \beta_1 x_o$

Observado para "o" $\longrightarrow$ y_o

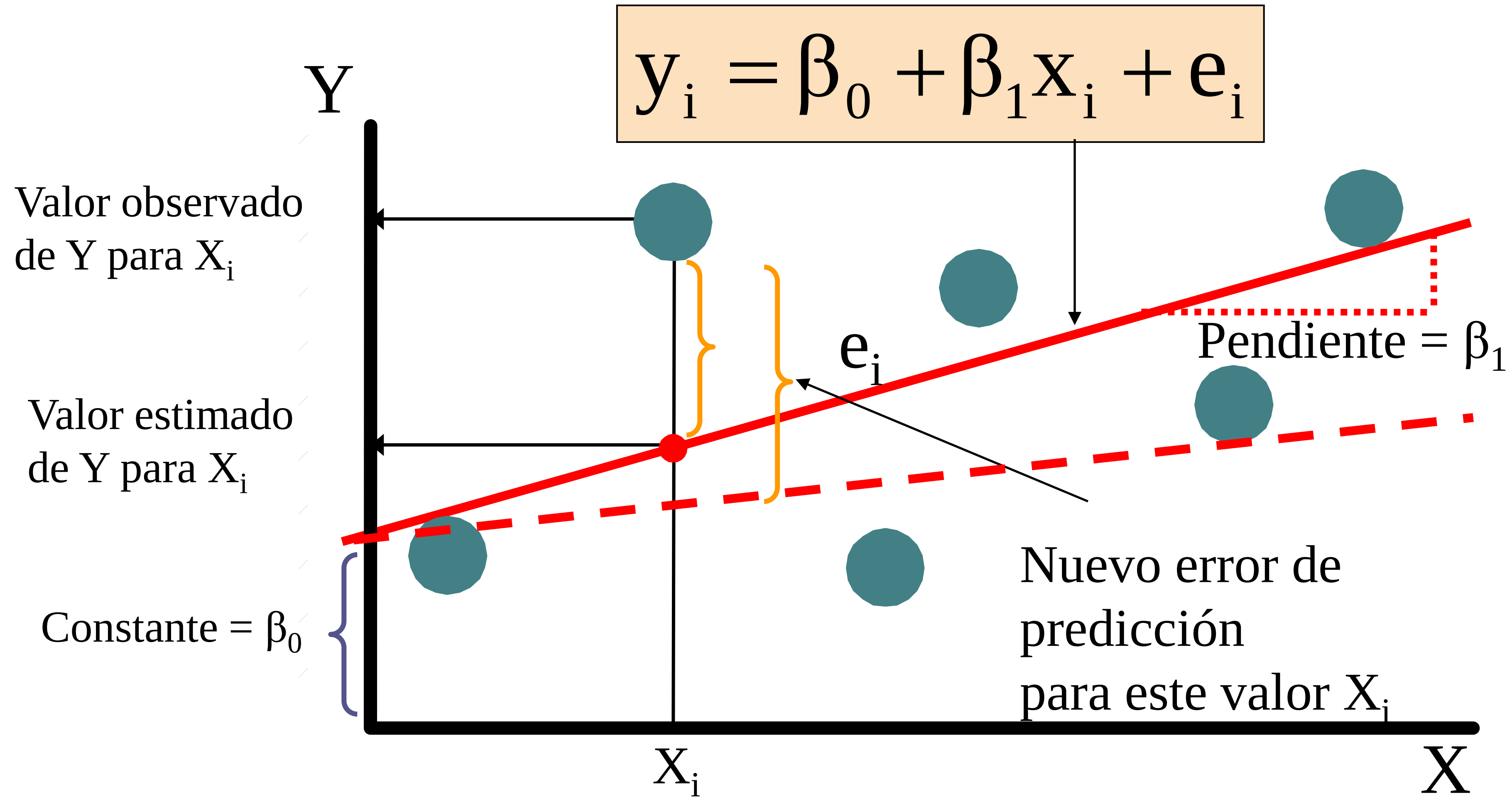
Error de predicción para "o" $\longrightarrow$ $y_o - \hat{y}_o = e_o$

Eliegiendo los betas altero el error de predicción, idea es hacerlo lo más chico posible!

Modelo de Regresión Lineal Simple



Cambiando la pendiente cambio el error



- Problemas
  - Tengo que tener una medida de cómo erro en conjunto y no sólo para una observación
  - Cuando achico un error, se me agrandan otros
  - Sumar todos los errores no es una buena idea, porque al sumar se cancelan los más y los menos
  - …..

  - Solución ¡MCC! Elegir los betas que hacen mínima la suma de los errores al cuadrado

- Si se cumple $E[e_i]=0$ (espero no tener error de predicción en promedio)

$$E[Y_i] = \beta_0 + \beta_1 X_i \quad (*)$$

- Por lo que la recta da el valor esperado para Y dado X
- Notar que

$$e_i = Y_i - E(Y)$$

tiene entonces interpretación como "error de predicción"

- Método de Mínimos Cuadrados Clásico: β_0 y β_1 son elegidos para minimizar (colectivamente) la distancia vertical entre las observaciones y la línea de tendencia

(*) Muchas veces se pone E(Y/X) para que quede claro que es el valor esperado de Y dado el X que observo (condicional en). Esta notación condicional es más apropiada

$$\min_{\beta_0,\beta_1}\left(e_1^2+\ldots+e_N^2\right)=$$

$$\min_{\beta_0,\beta_1}\left((y_1-\beta_0-\beta_1x_1)^2+\ldots+(y_N-\beta_0-\beta_1x_N)^2\right)$$

- O bien

$$\min_{\beta_0,\beta_1}\sum_{i=1}^{N}(y_i-\beta_0-\beta_1x_i)^2$$

- Se puede resolver analíticamente (derivadas) o numéricamente (búsqueda iterativa) – los software lo hacen por vos!

Remark

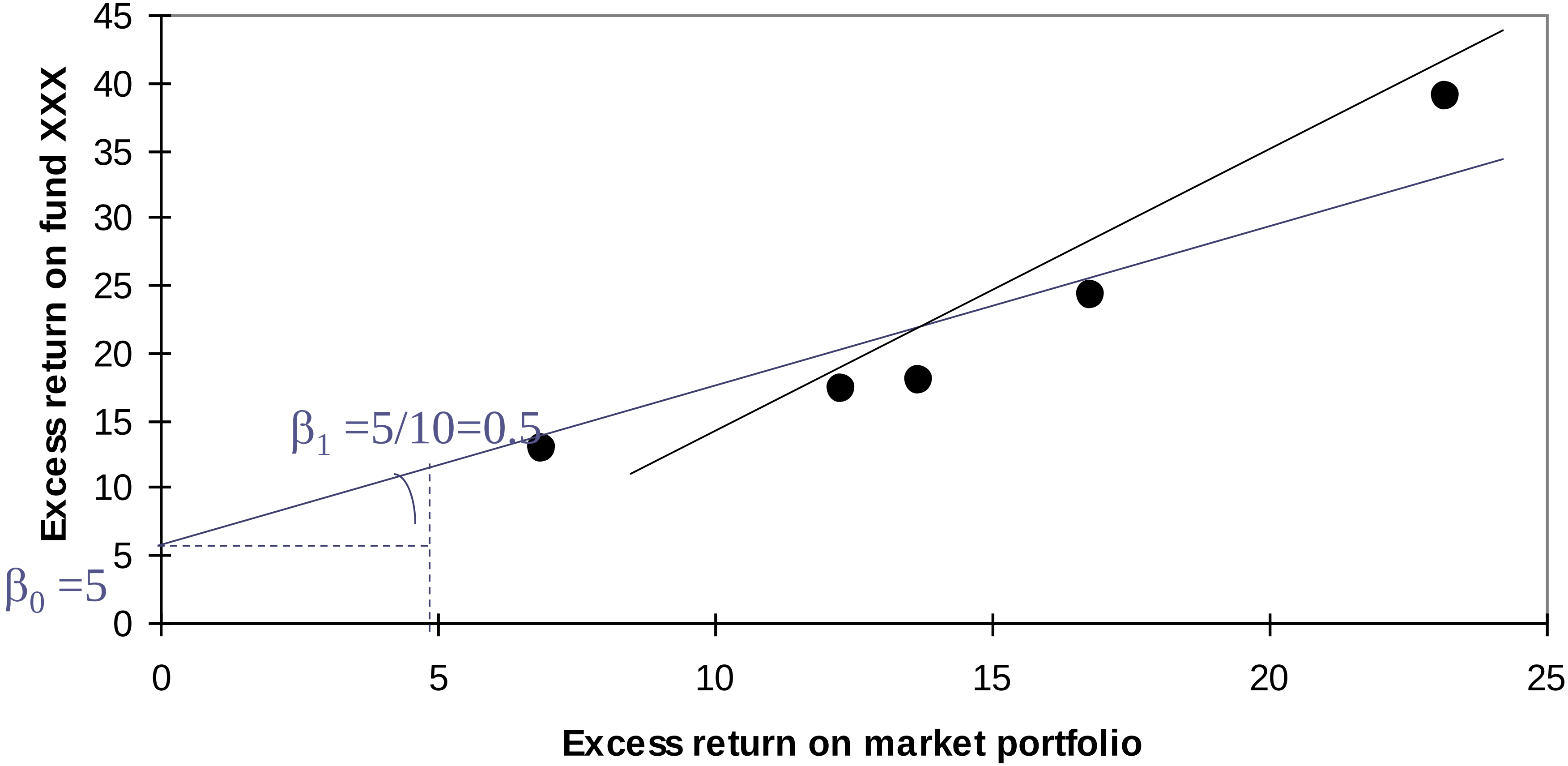
- y= *data* = *fit* + *residuo*
- *fit* = $a+bx$ *(en el caso lineal)*
- Pero no tiene porqué ser un modelo lineal, se pueden usar otros
- Mínimos Cuadrados siempre hace

$$\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

$$b = S_{xy} / S_x$$

$$a = \overline{y} - b\overline{x}.$$

β_0 es la ordenada al origen de la recta β_1 la pendiente



Derivación

- Estima β como el valor para los parámetros que minimiza la suma de los residuos al cuadrado

$$\min_{\beta} \sum_{i=1}^{N} e_i^2$$

- "i" indica las distintas observaciones (tiempo o cross-section)
- ¿Queremos elegir los betas para minimizar el error de predicción?
- Por qué minimizamos los errores al cuadrado?

Intuición Gráfica

- Una vez estimados podemos computar:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

- Donde $\hat{Y}_i$ es una estimación de E(Y | X_i) y

$$\hat{\beta}_0 \ y \ \hat{\beta}_1$$

son los estimadores de los parámetros poblacionales.

- Podemos también estimar:

$$\hat{e}_i = Y_i - E[Y / X_i] = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

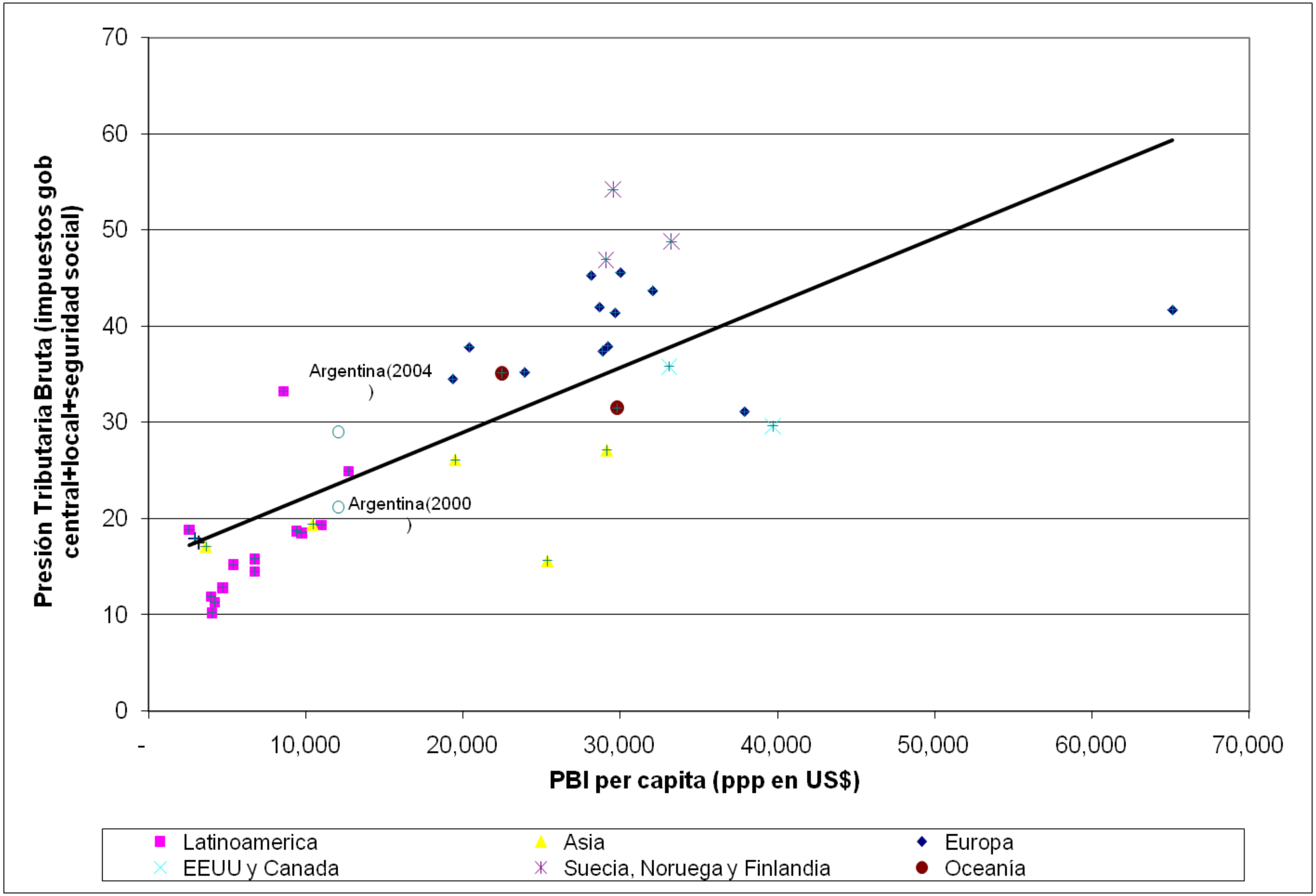
- El interés se centra en $\hat{\beta}_0$ y $\hat{\beta}_1$ que son estimadores de los veraderos betas (que desconocemos).
- Por ser estimadores, son variables aleatorias, y tienen una distribución.
- Nos interesa hacer test de hipótesis sobre estos estimadores.
- Por el TCL, se puede aproximar la distribución de los betas con una Normal (o una T de Student si tengo pocos datos)
- Nos interesa hacer intervalos de confianza (o test de hipótesis).

$$\hat{\beta}_1 \pm Z_{\alpha/2} s_{\hat{\beta}_1}$$

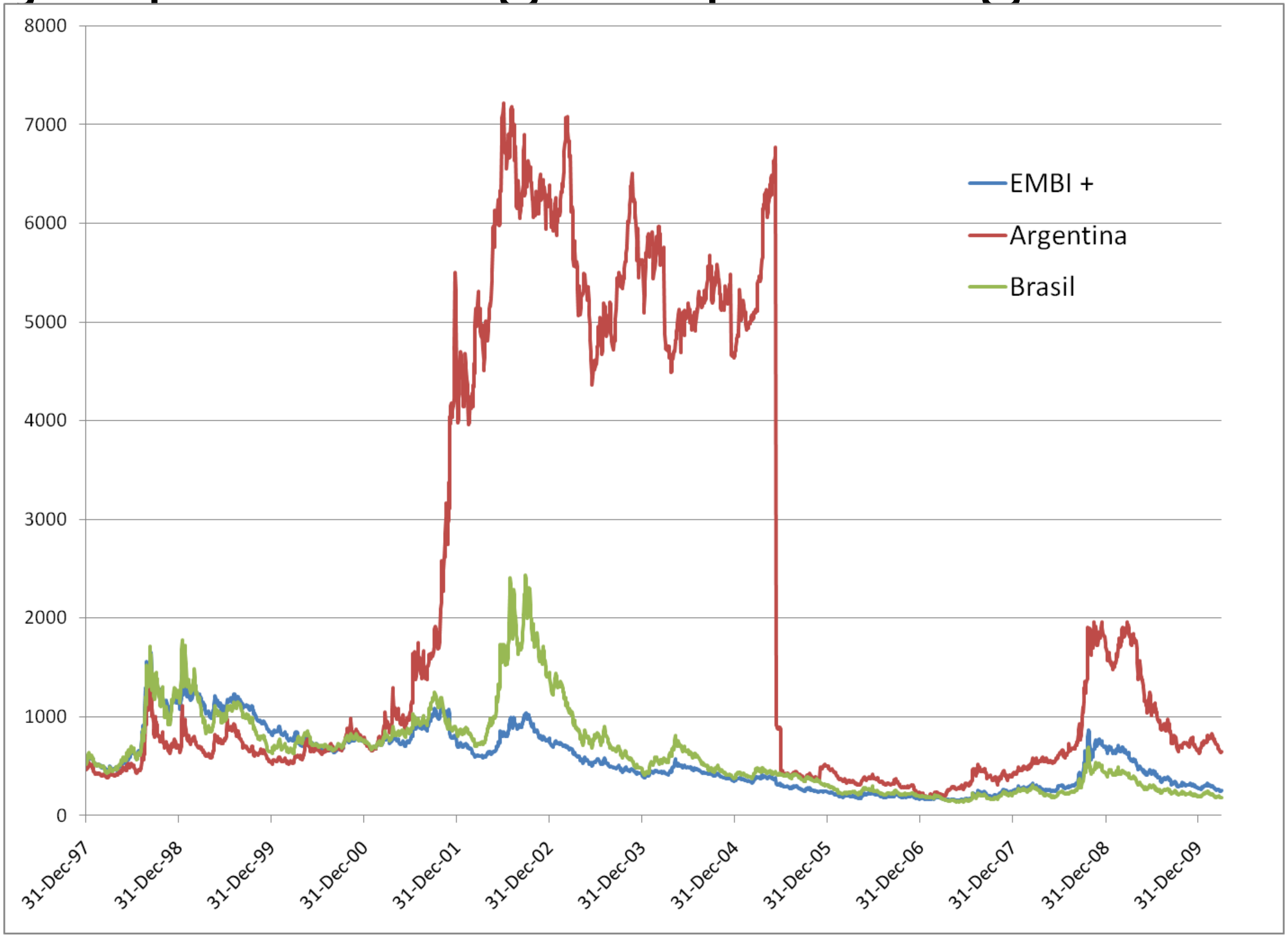
¿Para que sirve?

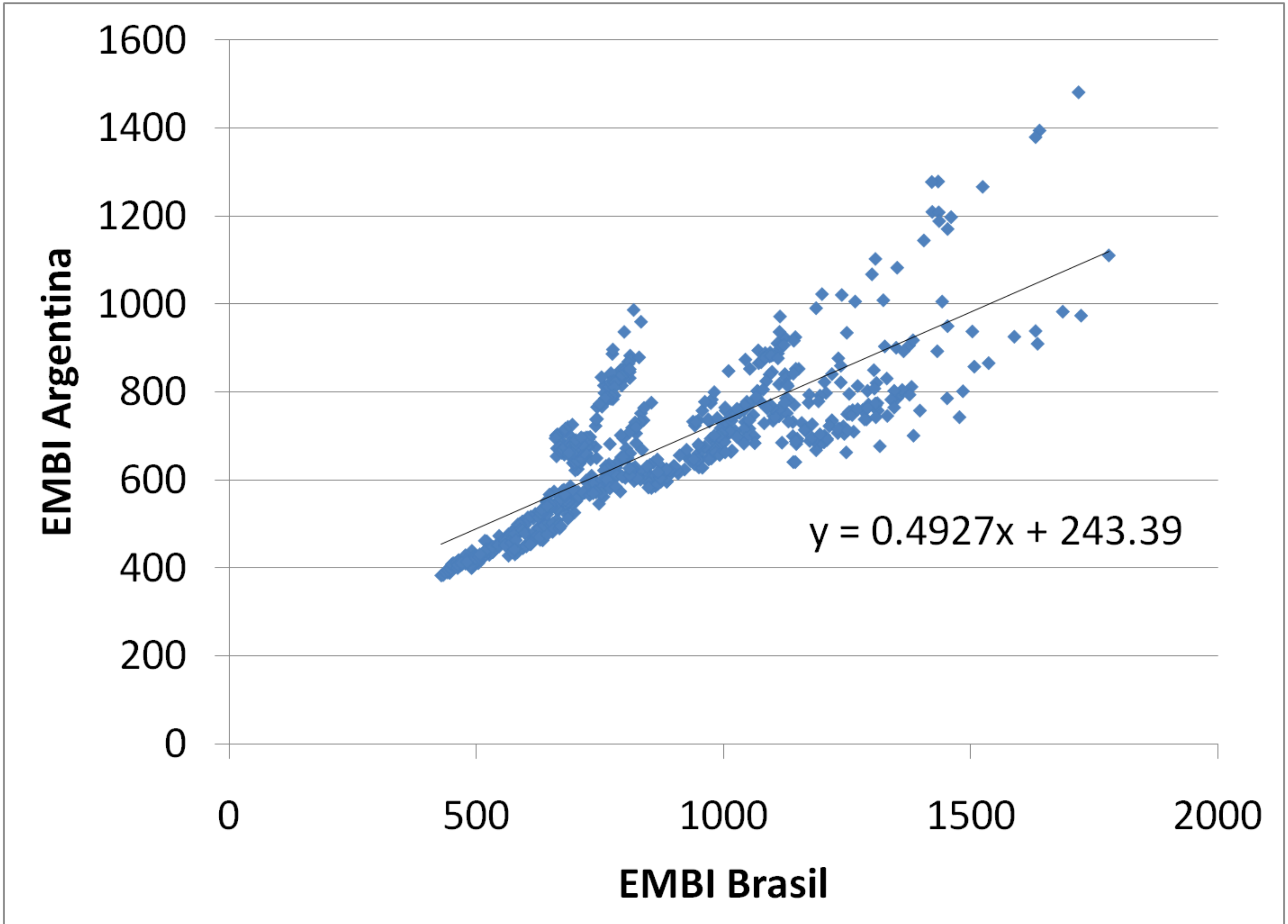
- Describir relaciones entre variables (no implica causalidad)
- Para estimar los parámetros (ejemplo beta de CAPM)
- Para testear modelos
- Para predecir.

Ejemplo 1. ¿Es la presión tributaria actual de Argentina Excesiva?



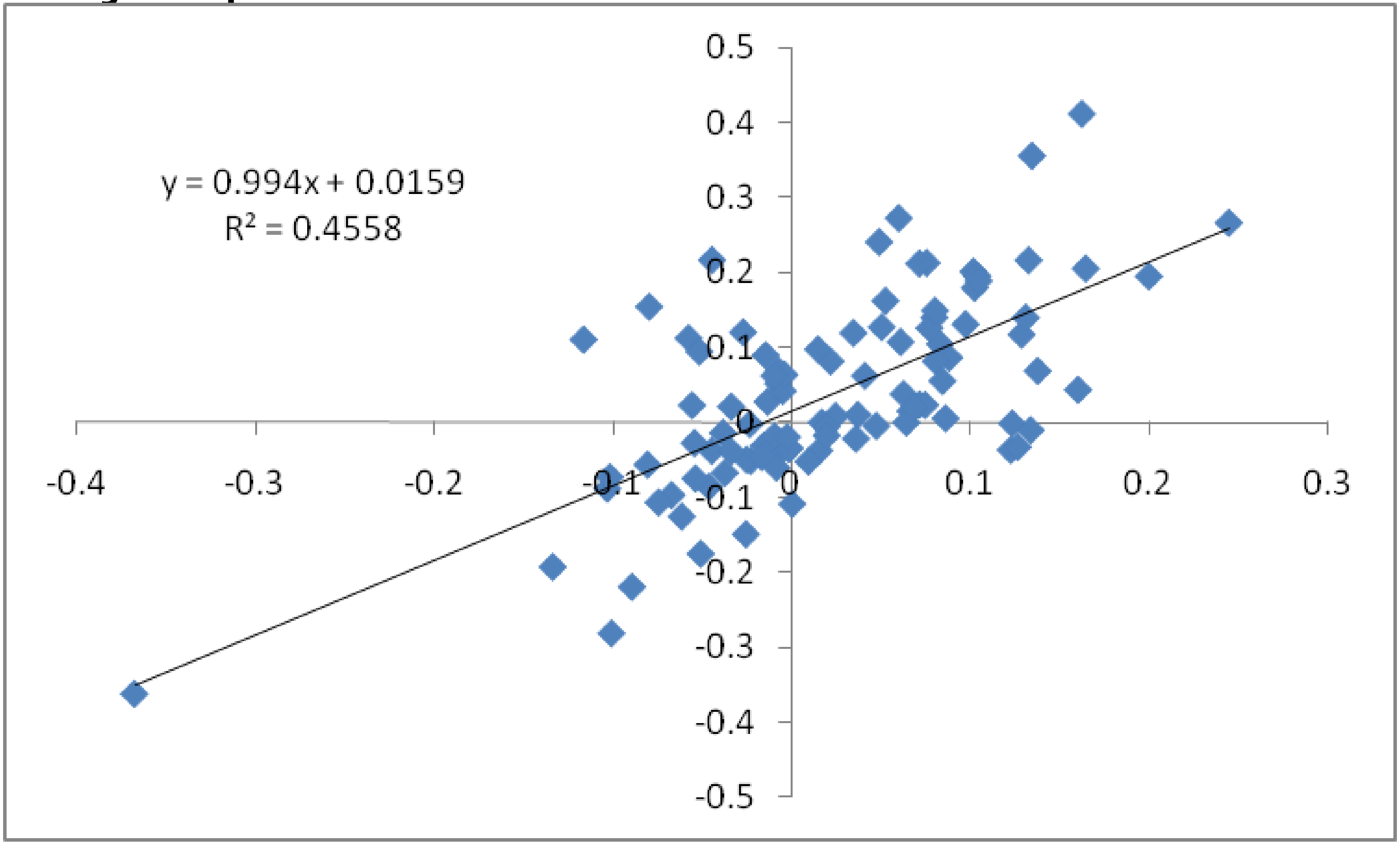
Ejemplo 2. Riesgo empresa Argentina





1600
1400
1200
1000
800
600
400
200
0
EMBI Argentina
y = 0.4927x + 243.39
0
500
1000
1500
2000
EMBI Brasil

Ejemplo 3. Testeando CAPM



Modelo General (más regresores)

- Asume:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + e_i \;, i=1, \ldots, n$$

Y: variable dependiente

Hay k variable explicativas

β_i (k+1) parámetros desconocidos

e error (lo que no está explicado por todas las Xs)

- El modelo es lineal en los β (coeficientes de regresión), que son parámetros desconocidos.

Derivación de Estimadores MCC (OLS)

$$Y_i = \beta_0 + \beta_1 X_1 + e_i \ , i=1, \ldots, n$$

– MMC resuelve

$$\min_{\beta} \sum_{i=1}^{N} e_i^2$$

$$\min_{\beta} \sum_{i=1}^{N} \left(y_i - \beta_0 - \beta_1 x_i\right)^2$$

Condiciones de Primer Orden

$$\frac{\partial S}{\partial \beta_0} = \sum_{i=1}^{N} 2\left(y_i - \beta_0 - \beta_1 x_i\right)(-1)$$

$$\frac{\partial S}{\partial \beta_1} = \sum_{i=1}^{N} 2\left(y_i - \beta_0 - \beta_1 x_i\right)(-x_i)$$

- Cuando se igualan estas derivadas parciales a cero se obtienen los estimadores

$$\frac{\partial S}{\partial \beta_0} = \sum_{i=1}^{N} 2\left(y_i - \beta_0 - \beta_1 x_i\right)(-1) = 0 \quad \Rightarrow \quad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

- Reemplazando en la otra CPO obtenemos para β1

$$\frac{\partial S}{\partial \beta_1} = \sum_{i=1}^{N} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-x_i) = 0$$

$$= \sum_{i=1}^{N} \left(y_i - \left(\overline{y} - \hat{\beta}_1 \overline{x}\right) - \hat{\beta}_1 x_i\right)(-x_i) = 0$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S^2{}_x}$$

$$\sum_{i=1}^{N} \left(\left(y_i - \overline{y}\right)x_i\right) = \sum_{i=1}^{N} \hat{\beta}_1 \left(\left(x_i - \overline{x}\right)x_i\right)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} \left(\left(y_i - \overline{y}\right)x_i\right)}{\sum_{i=1}^{N} \left(\left(x_i - \overline{x}\right)x_i\right)} = \frac{\sum_{i=1}^{N} \left(\left(y_i x_i - \overline{y} x_i\right)\right)}{\sum_{i=1}^{N} \left(\left(x^2{}_i - \overline{x} x_i\right)\right)} = \frac{\sum_{i=1}^{N} y_i x_i - N \overline{y}\,\overline{x}}{\sum_{i=1}^{N} x^2{}_i - N\left(\overline{x}\right)^2}$$

Derivación Formal de Estimadores MCO con MATRICES

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i, \quad$ para $i=1,\ldots,n$

En términos matriciales (más compacto): $Y = X\beta + U$

MCO resuelve:

$$\min_{\beta} \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{1i} - \ldots - \beta_k X_{ki} \right)^2$$

En términos matriciales:

$$\min_{\beta} U'U = \left(Y - X\beta\right)'\left(Y - X\beta\right)$$

Con matrices

Condiciones de Primer Orden:

$$\frac{\partial UU'}{\partial \beta'} = -2X'Y + 2X'X\beta = 0$$

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

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} n & \sum X_{1i} & \cdots & \sum X_{ki} \\ & \sum X_{1i}^2 & & \sum X_{1i}X_{ki} \\ & & \ddots & \vdots \\ & & & \sum X_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum Y_i \\ \sum X_{1i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix}$$

Ejemplo Excel.

- Usar comando regresión en Data Analysis

- Base de housing:
  - Price: Precio por pie cuadrado
  - Location: 4 barrios
  - Lot size: tamaño del lote (en pies cuadrados)

- ¿Es importante la edad de la casa para su precio? ¿Cuán importante?

- Estime cual sería el precio esperado de una casa de 20 años

Propiedades de MCC

Propiedad 1

La regresión lineal pasa por la media muestral de X e Y

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

Propiedad 2

La media de una predicción es igual a la media muestral de Y

$$\sum_{i=1}^{N} \hat{Y}_i / N = \bar{Y}$$

Propiedad 3

La media y la suma de los erroes de predicción es cero (si el modelo tiene constante)

$$\sum_{i=1}^{N} e_i = \sum_{i=1}^{N} e_i / N = 0$$

Propiedad 4

Los errors de predicción tienen correlación nula con predicted Y values

$$\sum_{i=1}^{N} e_i \hat{Y}_i = 0$$

Propiedad 5

Los errors de predicción tiene correlación nula con los valores de X

$$\sum_{i=1}^{N} e_i X_i = 0$$

Propiedad 6

La estimación OLS no es más que un promedio ponderado de los valores de Y

¿Qué cosas mirar de una regresión?

- El signo de los coeficientes.
- El R2 que me dice que porcentaje de la variación de Y está explicada por mi modelo.
- El tamaño de la muestra
- Una predicción (donde imputo valores para los x en la recta estimada y me tira los valores esperados para y)
- Significación individual
- Significación global

Poder Explicativo del Modelo

Una medida de cuán bueno es el ajuste del modelo es el Coeficiente de Determinación R^2:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Suma de cuadrados **total** $= SST = (Y_1 - \bar{Y})^2 + \cdots + (Y_n - \bar{Y})^2$

$=$ **"variabilidad total en los datos"**

Suma de cuadrados **regresión** $= SSR = (\hat{Y}_1 - \bar{Y})^2 + \cdots + (\hat{Y}_n - \bar{Y})^2$

$=$ **"variabilidad explicada"**

Suma de cuadrados **error** $= SSE = (Y_1 - \hat{Y}_1)^2 + \cdots + (Y_n - \hat{Y}_n)^2$

$=$ **"variabilidad no explicada"**

Propiedad de las variabilidades: $SST = SSR + SSE$

Total = Explicada + No explicada

Si el modelo tiene constante:

$$SST = SSR + SSE$$

por lo que :

$$0 \leq R^2 \leq 1$$

En este caso R^2 se lo interpreta como el porcentaje de la variación total de la variable y explicada por el modelo.

Nota Tecnica:

a) R^2 es una medida de cuán exitoso es el modelo siempre y cuando el modelo incluya una constante (si no incluye constante, no tiene porque estar entre 0 y 1).

b) por construcción el R^2 sube cuando agregamos más variables. Esta es una propiedad no deseada de este indicador.

- R2 múltiple no se usa mucho, es simplemente:

R Múltiple= $\sqrt{R^2}$ (en caso de una variable $= |\rho_{XY}|$)

- El R2 ajustado se usa para comparar modelos:

Generalmente se usa en lugar de R^2 el R^2 ajustado", definido como:

$$\overline{R^2} = 1 - \frac{SSE/(T-K)}{SST/(T-1)}$$

El $\overline{R^2}$ es una medida arbitraria, no tiene porqué estar entre 0 y 1.

Lo mejor es utilizar un test de significación para ver si el modelo con más variables explica mejor la variable de interés, cosa que veremos en breve.

Test de Hipótesis

- A menudo nos importan hipótesis.
- Por ejemplo, queremos saber si el parámetro para un regresor es cero

- Ejemplo:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i \;, i=1, \ldots, n$$

Y: ventas

X_1: publicidad en Nación

X_2: publicidad en Infobae

- Hipótesis: ¿podría ser β_2 cero?
- Respuesta. Hago un intervalo de confianza para β_2 y me fijo si cero cae dentro.

1. Test de significatividad individual

- Se pregunta si un beta en particular podría ser cero

$$H_0 : \beta_i = 0$$

- Hipótesis que se testea con el estadístico t

$$t = \frac{\hat{\beta}_i - \beta_{i,Ho}}{\hat{s}_{\hat{\beta}_i}} \approx t(N-(k+1))$$

- k cantidad de regresores, N tamaño muestra, t valor tabla T de Student, S error estándar del estimador.

- Esto es similar a lo visto en clase pasada para la media muestral, donde se sabía

$$\overline{X} \rightarrow N(\mu, \frac{\sigma^2}{N})$$

$$z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1) \; bajo \; H_0$$

$$t = \frac{\overline{X} - \mu_0}{S / \sqrt{n}} \sim \mathrm{t}_{n-1} \;\; bajo \; H_0$$

- Si $|t| > t_{0.025}(n-1)$ rechazo H_0 (al 95% de confianza)
- Acá se puede probar que bajo ciertos supuesto el estimador beta sigue una Normal

Test de hipótesis en criollo

- Intuición del Test: Sé que el estimador tiene distribución Normal pero no sé su media y su desvío. Vos me decis que la media es tanto (la hipótesis nula), entonces eso me determina una distribución de valor probables para la estimación, que quiere decir que cuando vaya y tome una muestra al azar, con alta probabilidad debería ver que la estimación que obtenga debe caer entre cierto rango de valores. Veo si la estimación me cayó en ese rango, si es cierto, OK podés tener razón (no rechazo la hipótesis nula) si cae fuera, digo no-way tenés razón, la probabilidad es demasiado baja de que la realidad me tire un valor como el que obtuve si H0 es cierto, con lo cual debe ser falso (rechazo H0)

Pasos de todo test de hipótesis

1. Planteo hipótesis nula y alternativa para algún parametro o relación entre parámetros
2. Busco un buen estimador del parámetro o su relación planteada en 1 (bueno en el sentido que lo estime bien, insesgado y consistente)
3. Averiguo cuál es la distribución del estimador para saber que valores puede tomar la estimación.
4. Asumo H0 es verdad, y veo cual es el rango de valores que debiera tirar una estimación con alta probabilidad (nivel de confianza, usualmente 99%, 95% o 90%)
5. Comparo realidad con 4.

P-value

- Alternativa del test, donde en vez de comparar los t comparo probabilidades
- P-value: computa cuál es la probabilidad de haber obtenido el valor del estimador que obtuvimos en la realidad si la hipótesis nula es cierta.
- Si esta probabilidad es muy chica, entonces lo que vemos no es consistente con la hipótesis y H0 no es cierta.
- El punto es cuánto es chico para el p-value, y eso es lo que se elige como alpha (nivel de significatividad)
- Si P-value<alpha, entonces la Ho es falsa

Formas de presentar la info

- Errores estándares
- Valores t (ya asmen un α)
- Intervalos de confianza (ya asmen un α)
- P-Valores
- Estrellitas (*** 1%, ** 5%, * 10%

Table 2

The enhancing impact of friendship networks on sales managers' performance, Braz. Adm. Rev. vol.10 no.2 Rio de Janeiro Apr./June 2013

Results of the Ordinary Least Square Regression

Variable	Total Sales			New Product Sales			Prospecting New Deals			Converting New Deals		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Network Enhancing Factor												
Professional Network * Friend Network $H_{1,2,3,4}$			.567			.485			.428			.448
			(2.83)**			(2.35)**			(1.94)*			(1.97)*
Network Variables												
Professional Network		.020	.291		.026	.291		.043	.191		.009	.253
		(.25)	(2.22)		(.34)	(2.15)*		(.53)	(1.32)		(.11)	(1.69)†
Friend Network		.120	.181		.025	.233		.027	.200		.175	.061
		(1.87)†	(1.46)		(.37)	(1.82)†		(.39)	(1.46)		(2.41)*	(.43)
Control Variables												
Experience (years)	.225	.199	.184	.186	.187	.174	.142	.127	.116	.150	.119	.107
	(3.26)**	(3.10)**	(2.91)**	(2.94)**	(2.84)**	(2.67)**	(2.01)*	(1.82)†	(1.66)†	(2.12)*	(1.64)	(1.48)
Education level	.106	.096	.107	.216	.217	.226	.216	.210	.219	.121	.110	.118
	(1.74)†	(1.57)	(1.78)†	(3.52)**	(3.49)**	(3.67)**	(3.29)**	(3.17)**	(3.32)**	(1.76)†	(1.60)	(1.74)†
Size of Client Portfolio	.147	.133	.100	.123	.125	.097	.096	.087	.062	.036	.022	.004
	(2.17)*	(1.95)*	(1.47)*	(1.78)†	(1.78)†	(1.37)	(1.32)	(1.16)	(.83)	(.47)	(.28)	(-.05)
Distance from headquarters (miles)	-.061	-.022	-.028	-.140	-.137	-.142	.015	.030	.026	-.040	.009	.005
	(-.85)	(-.30)	(-.38)	(-1.93)†	(-1.82)†	(-1.90)†	(.19)	(.37)	(.32)	(-.50)	(.12)	(.06)
Team Effect (US$)	.391	.372	.345	.309	.315	.292	.310	.392	.272	.311	.296	.275
	(5.57)**	(5.05)**	(4.73)**	(4.38)**	(4.18)**	(3.88)**	(4.08)**	(3.64)**	(3.38)**	(3.92)**	(3.57)**	(3.31)**
R^2	0.386	0.399	0.423	0.368	0.368	0.386	0.282	0.284	0.298	0.213	0.238	0.253
Adjusted R^2	0.370	0.377	0.399	0.352	0.346	0.360	0.264	0.258	0.269	0.193	0.211	0.222
R^2 Change		0.013	0.024**		0.095	5.517**		0.002	3,759*		0.025*	0.015*
F-statistic	24.85**	18.58**	17.84**	23.01**	16.32**	15.29**	15.53**	11,11**	10.32**	8.23**	8.23**	8.23**

Note. The table reports standardized coefficients with t-values in parentheses.
†p < .10; *p < .05; **p < .01.

2. Test de Significación Global (Test F)

- Es un test que se pregunta si todos los regresores incluidos tienen al mismo tiempo valores que son cero

$$H_0 : \beta_1 = 0, \beta_2 = 0, \ldots, \beta_k = 0$$

- Hipótesis que se testea con el siguiente estadístico

$$F = \frac{SSR / K}{SSE / T - (K+1)} \, sigue \; F_{K, T-(K+1)}$$

- Está emparentado con el R2. Si F es muy grande, entonces la hipótesis nula se rechaza. Comparo el Significance F (p-value) con el alpha que me guste (para la significatividad)

Caso Compass Maritme Services

Tom Roberts, founding partner of Compass Maritime Services, a New Jersey-based firm specializing in the sale and purchase of ships ("S&P brokers" in the industry parlance), valuation, shipping research, and consulting, was approached in May 2008, by a client who wanted to purchase a caponize bulk carrier. Roberts asked Basil Karatzas, the firm's Director of Projects and Finance, to find a suitable ship and value it. Karatzas identified the Bet Performer, an 11-year old bulk carrier, and his challenge is to use the available data to determine an appropriate offer price for the client.