Metodología I

Magíster en Ciencias Sociales

Pablo Pérez Ahumada
Universidad de Chile
Departamento de sociología

Módulo 3
Regresión lineal y logística binaria

REGRESIÓN LINEAL

Aspectos básicos

¿Qué es la regresión lineal?

- Modelo estadístico

- Se usa para
  - Analizar existencia de relaciones entre variables (sólo si esa relación es *lineal*)
  - Inferir si esas relaciones son estadísticamente significativas
  - Predecir cómo cambiaría va variable dependiente si cambian los valores de las variables independientes

- Dos tipos de regresión
  1. Simple (una variable independiente)
  2. Múltiple (más de dos variables independientes)

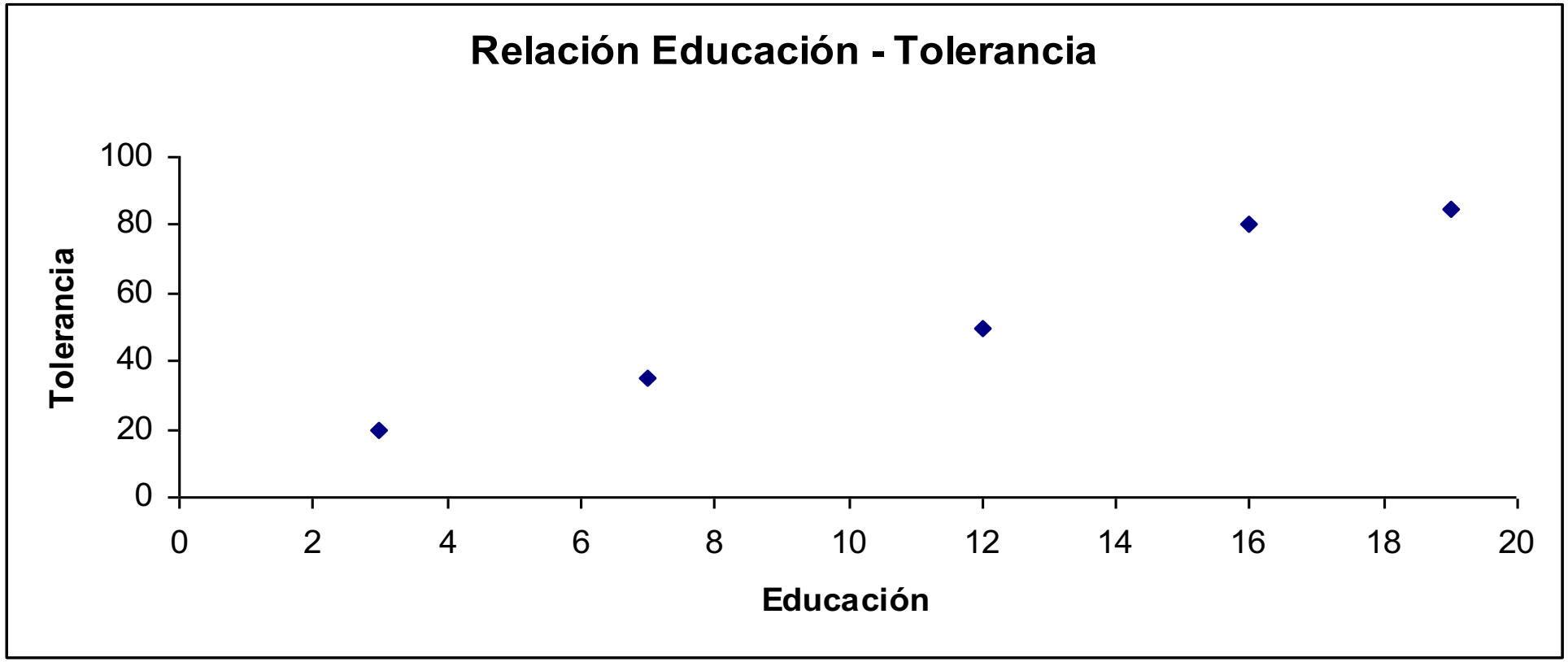
Estimación de la recta de regresión

- Ejemplo: se tiene 5 casos y se quiere ver si existe relación entre años de educación (años) y grado de tolerancia (puntaje de 1 a 100)

	Educación en años (X)	Nivel de tolerancia (Y)
Caso 1	3	20
Caso 2	7	35
Caso 3	12	50
Caso 4	16	80
Caso 5	19	85

Estimación de la recta de regresión

- Representación gráfica



Estimación de la recta de regresión

- Una recta puede ser graficada calculando los componentes de su ecuación:

$$\acute{Y} = a + \beta x$$

- Donde
  - $\acute{Y}$ = Valor de Y que se quiere predecir
  - α = intersección de la recta con eje y.
    - Es también el valor estimado cuando X = 0. También recibe el nombre de *constante.*
    - Debido a que es una estimación, es común que a sea un valor sin sentido analítico (incluso negativo).
  - β = **coeficiente de regresión.** Muestra la pendiente de la recta de regresión → "efecto" estimado sobre Y para un cambio de 1 unidad de X

Coeficiente de regresión (β)

- Coeficiente de regresión *b* (beta)

$$\beta = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2}$$

- Donde:
  - β = coeficiente de regresión (pendiente)
  - X = variable independiente
  - Y = variable dependiente
  - $\bar{X}$ = media de la variable independiente
  - $\bar{Y}$ = media de la variable dependiente

- ¿Qué nos indica esta fórmula? **Básicamente,** un análisis de variación conjunta de X e Y

Coeficiente de regresión (β)

Cálculo de coeficiente de regresión

	X (años educación)	Y (nivel de tolerancia)	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X}) * (Y - \bar{Y})$	$(X - \bar{X})^2$
Caso 1	3	20	-8,4	-34	285,6	70,56
Caso 2	7	35	-4,4	-19	83,6	19,36
Caso 3	12	50	0,6	-4	-2,4	0,36
Caso 4	16	80	4,6	26	119,6	21,16
Caso 5	19	85	7,6	31	235,6	57,76
	$\bar{X} = 11{,}4$	$\bar{Y} = 54$			Σ = 722	Σ = 169,2

$$\beta = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \frac{722}{169{,}2} = \mathbf{4,267}$$

Coeficiente α (intercepto)

- Luego de calcular b, se puede calcular α como:

$$a = \bar{Y} - \beta\bar{X}$$

- *Donde:*
  - α = intercepto
  - β = coeficiente de regresión
  - $\bar{X}$ = media de la variable independiente
  - $\bar{Y}$ = media de la variable dependiente
    - Nota: una propiedad de la recta de regresión es que siempre pasa por las coordenadas $(\bar{X}, \bar{Y})$. Esto es, pasa por los valores promedios de las variables X e Y.

Coeficiente α (intercepto)

- Cálculo de **α**:

$$a = \bar{Y} - \beta\bar{X}$$
$$a = 54 - 4{,}267 * 11{,}4$$
$$a = 5{,}355$$

Ecuación de la recta

- Habiendo calculado b y α, entonces la ecuación de la recta es

$$\acute{Y} = a + \beta x$$

$$\acute{Y} = 5{,}355 + 4{,}267x$$

Ecuación de la recta

- Habiendo calculado b y α, entonces la ecuación de la recta es

$$\acute{Y} = a + \beta x$$

$$\acute{Y} = 5{,}355 + 4{,}267x$$

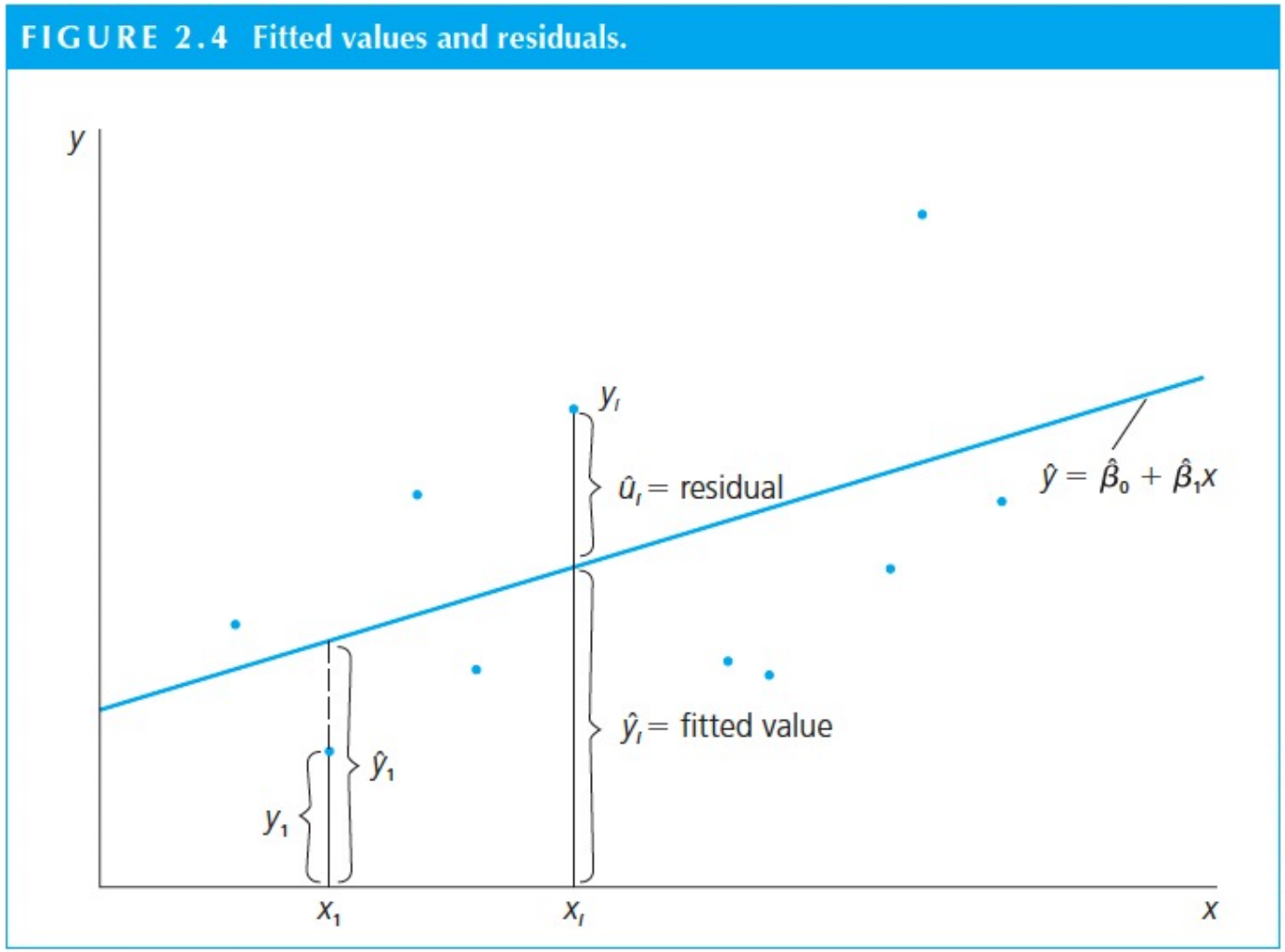
- Interpretación:

  Ante un cambio de una unidad de X (un año de nivel educacional), la variable Y (nivel de tolerancia) aumentará en 4,287 unidades

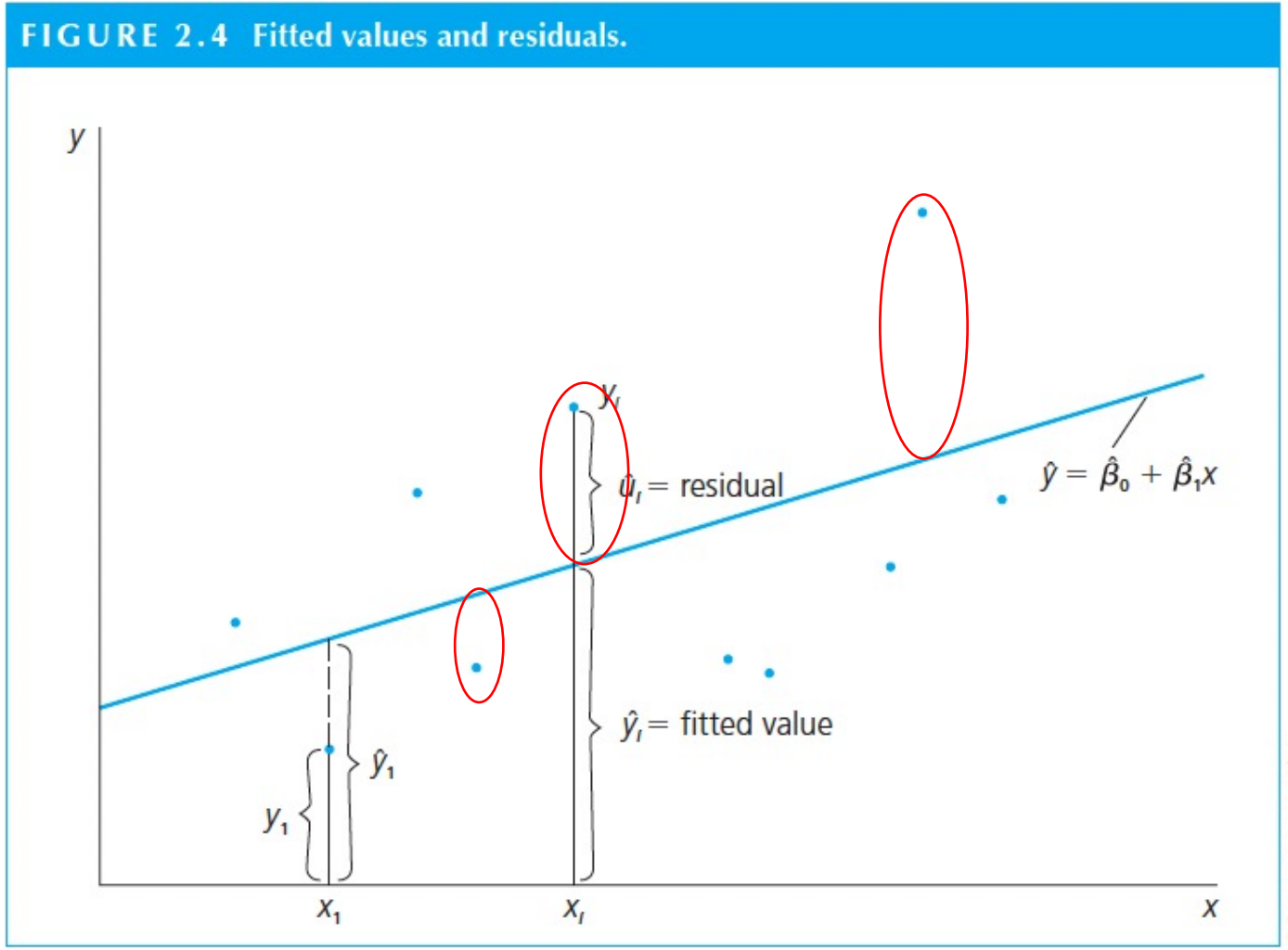
Recta de regresión y estimación de MCO (OLS en inglés)

- Calculada de esta manera, esta ecuación representa la **mejor estimación** (línea recta) de la relación entre X e Y.
  - Esta recta está basada en el método de **Mínimos Cuadrados Ordinarios**, y representa el estimador más eficiente e insesgado de α y β.

- ¿Por qué es la "mejor recta"? Porque **minimiza la suma de residuos** (distancia entre casos observados y la estimación que la recta hace de ella)

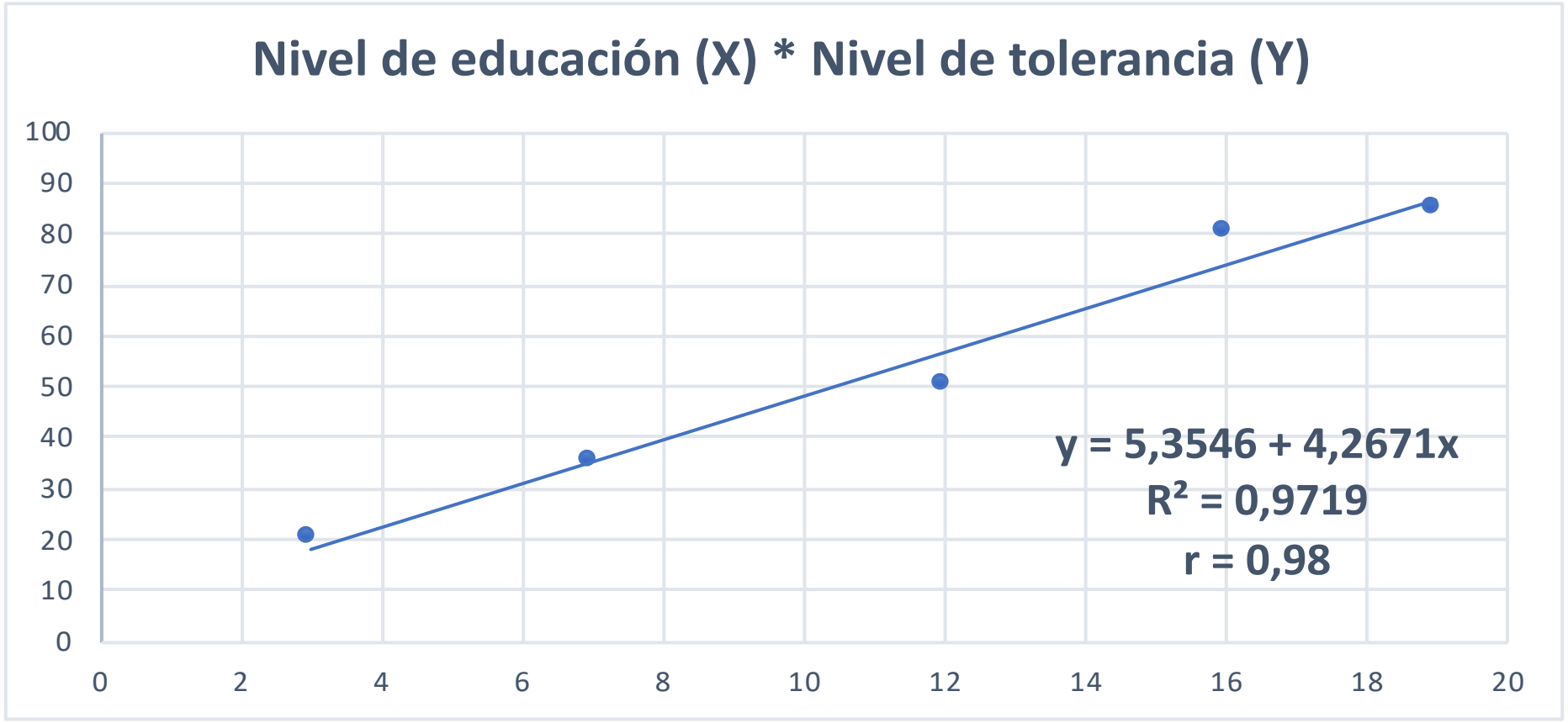


Wooldridge, J.. 2013. Introduction to Econometrics. A modern apporach, p. 33.



Wooldridge, J.. 2013. Introduction to Econometrics. A modern apporach, p. 33.

Ejemplo: recta de regresión (relación entre educación y tolerancia)



Midiendo el ajuste del modelo: El concepto de R^2

- Además de los coeficientes propios de la recta, es posible tener una medida conocida como R^2

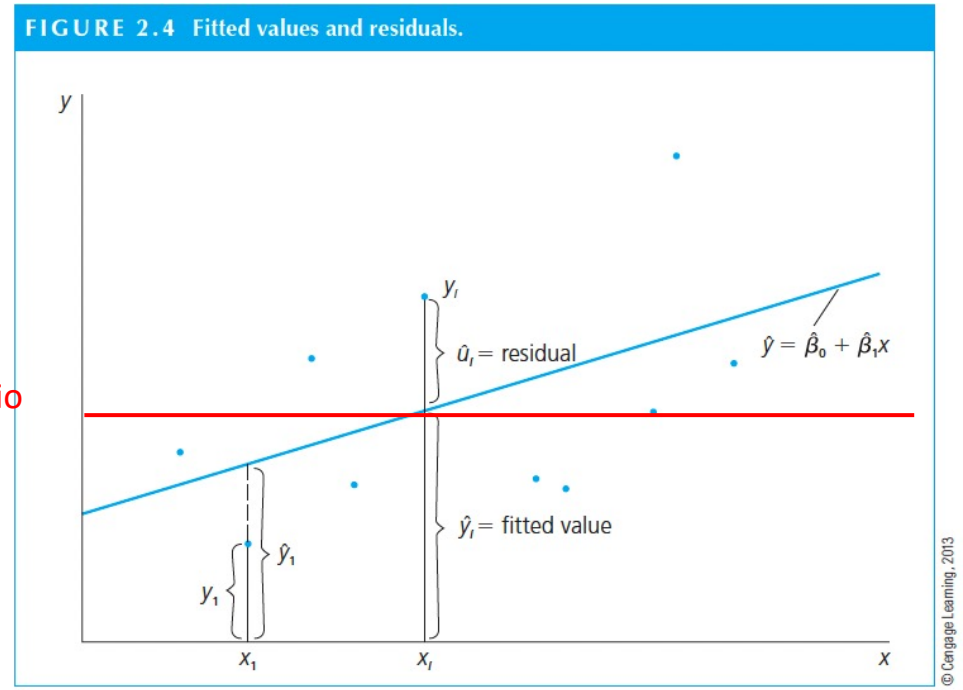
- Éste nos indica la proporción (o %) de variación de Y que es explicado por las variaciones de X (o, en el caso de las regresiones múltiples, de todas las Xs incluidas)
  - R^2 alto (más cercano a 1), indica que la recta se ajusta mejor a los puntajes reales, es decir, que la distancia entre lo predicho y lo observado es menor

El concepto de R^2

- Conceptualmente, el R^2 se calcula a partir de la suma de distancias al cuadrado

$$SC_{total} = SC_{reg} + SC_{error}$$

Promedio de Y

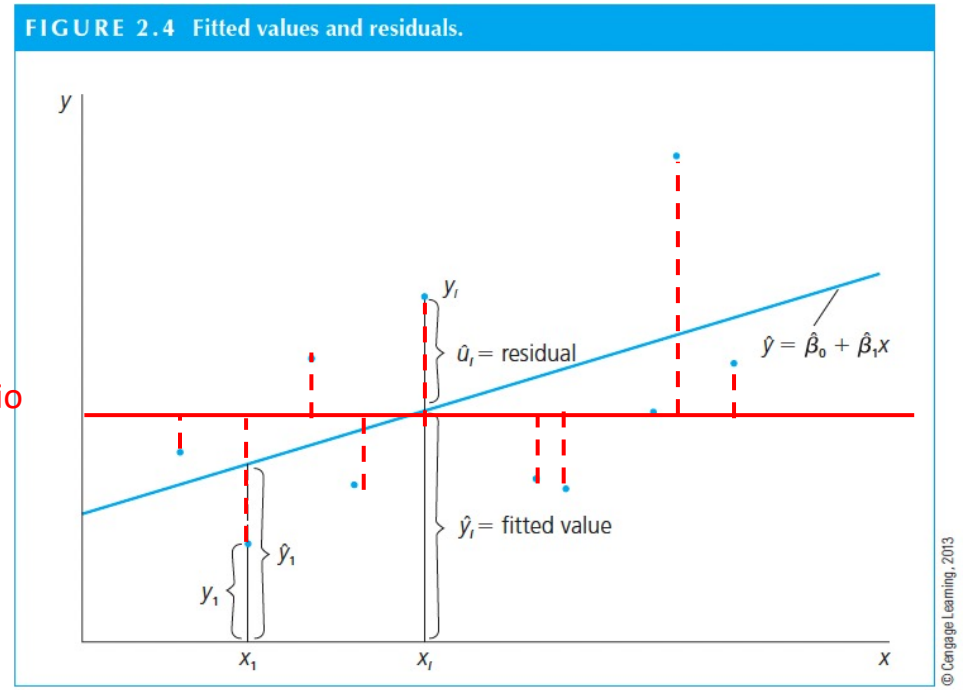


El concepto de R^2

- Conceptualmente, el R^2 se calcula a partir de la suma de distancias al cuadrado

$$SC_{total} = SC_{reg} + SC_{error}$$

SC_{total} = distancias entre valor observado de Y y su promedio



Promedio de Y

El concepto de R^2

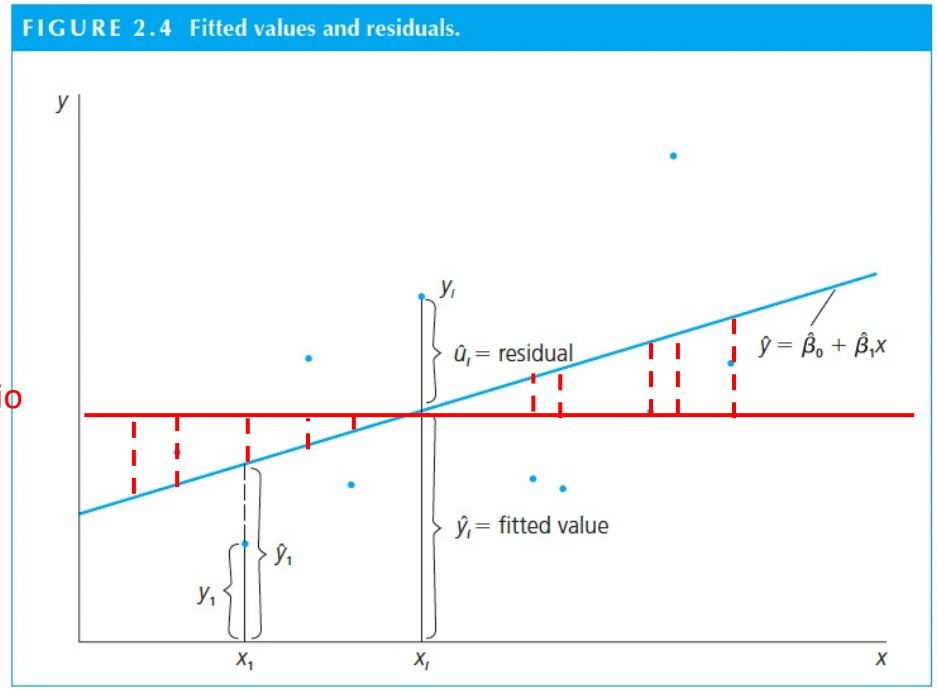
- Conceptualmente, el R^2 se calcula a partir de la suma de distancias al cuadrado

$$SC_{total} = SC_{reg} + SC_{error}$$

SC_{total} = distancias entre valor observado de Y y su promedio

SC_{reg} = distancias entre valor predicho de Y y su promedio

Promedio de Y



El concepto de R^2

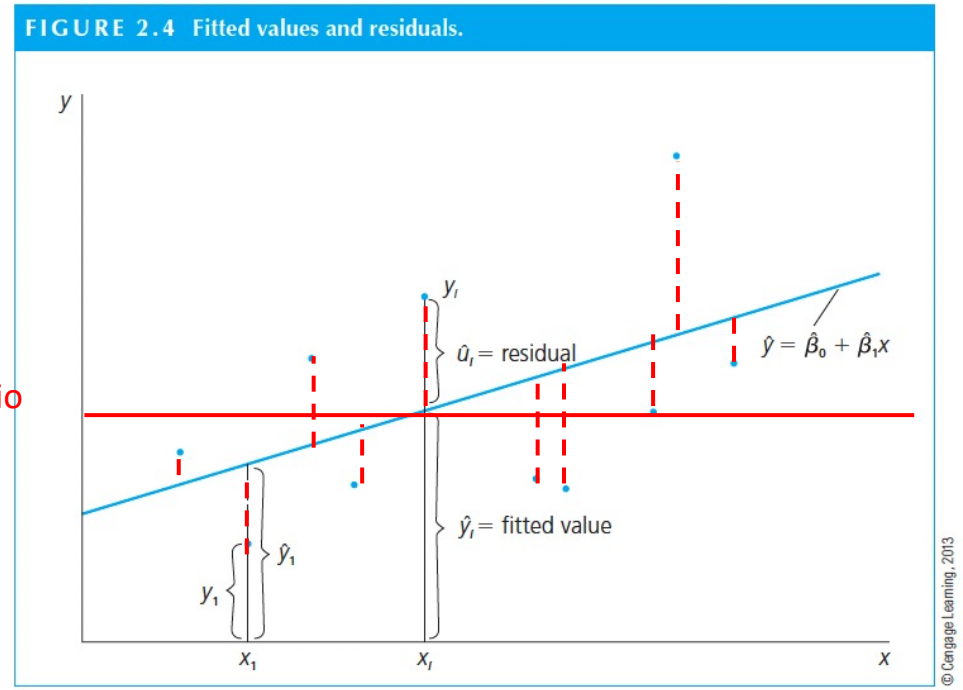
- Conceptualmente, el R^2 se calcula a partir de la suma de distancias al cuadrado

$$SC_{total} = SC_{reg} + SC_{error}$$

SC_{total} = distancias entre valor observado de Y y su promedio

SC_{reg} = distancias entre valor predicho de Y y su promedio

SC_{error} = distancias entre valor observado de Y y lo predicho por la recta



Promedio de Y

El concepto de R^2

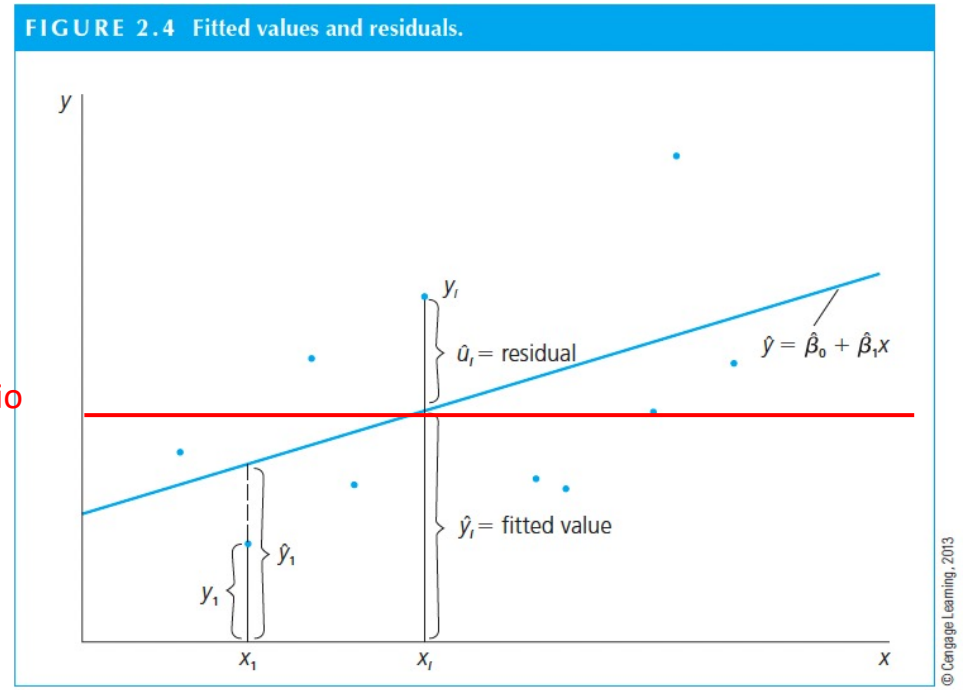
- Conceptualmente, el R^2 se calcula a partir de la suma de distancias al cuadrado

$$SC_{total} = SC_{reg} + SC_{error}$$

SC_{total} = distancias entre valor observado de Y y su promedio

SC_{reg} = distancias entre valor predicho de Y y su promedio

SC_{error} = distancias entre valor observado de Y y lo predicho por la recta



Promedio de Y

Así, el R^2

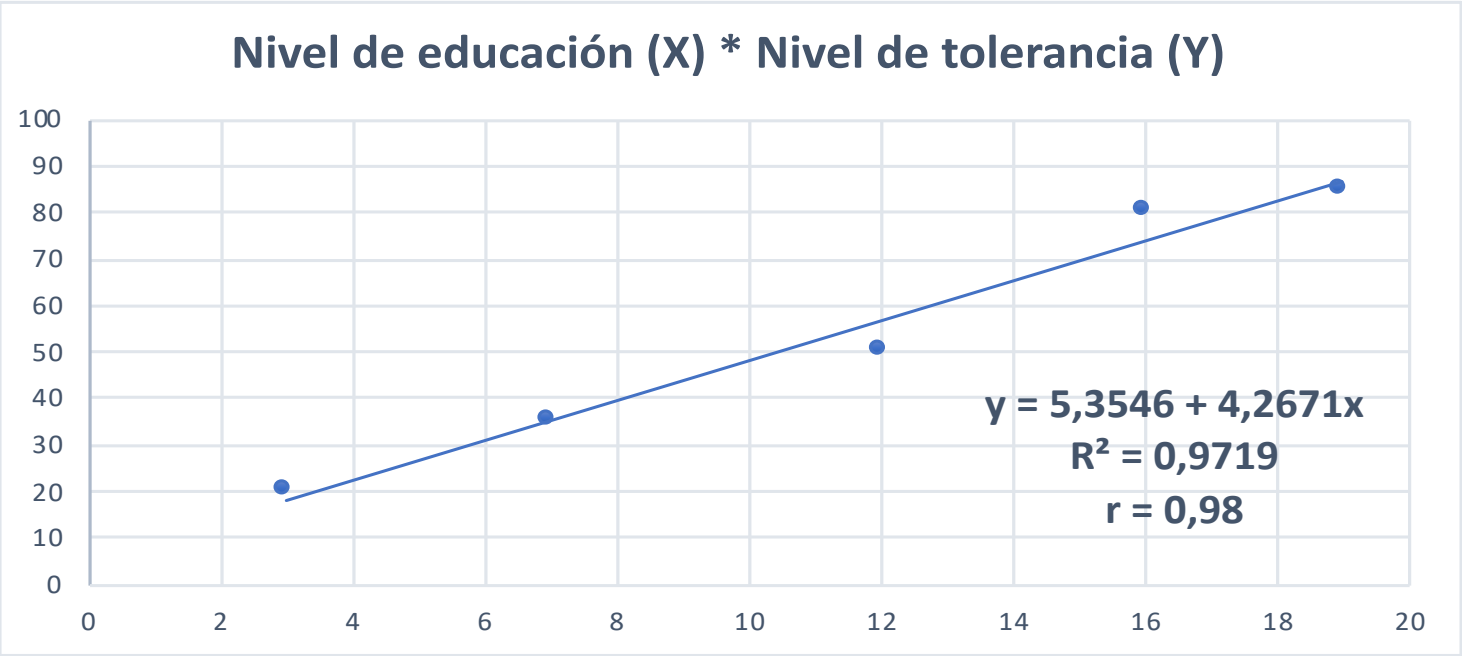
$$R^2 = \frac{SC_{reg}}{SC_{error}} = 1 - \frac{SC_{error}}{SC_{total}}$$

El concepto de R^2

- R^2 es la forma principal que tenemos de evaluar el ajuste (“calidad”) del modelo

El concepto de R^2

- R^2 es la forma principal que tenemos de evaluar el ajuste (“calidad”) del modelo



Acá, el R^2 es muy alto, porque la distancia entre la recta y los puntos es muy **baja**.

En otras palabras, la recta se ajusta muy bien a los datos (puntos)

El concepto de R^2

- Sin embargo, hay un problema
  - R^2 aumenta siempre que se incluyen más variables
  - Aspecto clave en regresión múltiple

El concepto de R^2

- Sin embargo, hay un problema
  - R^2 aumenta siempre que se incluyen más variables
  - Aspecto clave en regresión múltiple

- Pero también hay una solución

  - R^2 ajustado, que penaliza por la cantidad de predictores incluidos

EJEMPLO

Regresión lineal simple (un solo predictor)

Ejemplo

- ¿Existe una relación entre la desigual distribución de poder entre clases y el nivel de extensión de los derechos sindicales?

- **Variable dependiente:** extensión de los derechos sindicales
  - Puntaje 1 a 10; mayor puntaje mayor derecho a la sindicalización (Kucera & Sari, 2019)

  (*LR_Overall_Rev* en la base de datos usada acá)

- **Variable independiente:** disparidad de poder entre clases
  - Puntaje 0 a 4; mayor puntaje más desigualdad de poder (Varieties of Democracy Dataset).

  (*v2pepwrses_osp_Rev* en la base de datos)

- Análisis para 78 países (2017 o año más reciente disponible)

- **Variable dependiente:** extensión de los derechos sindicales
  - Puntaje 1 a 10; mayor puntaje mayor derecho a la sindicalización (Kucera & Sari, 2019)

  (*LR_Overall_Rev* en la base de datos usada acá)

- **Variable independiente:** disparidad de poder entre clases
  - Puntaje 0 a 4; mayor puntaje más desigualdad de poder (Varieties of Democracy Dataset).

  (*v2pepwrses_osp_Rev* en la base de datos)

- Análisis para 78 países (2017 o año más reciente disponible). Más detalles en artículo

Original Research Article

IJCS

International Journal of Comparative Sociology
1–18

DOI: 10.1177/00207152231163846
journals.sagepub.com/home/cos
SAGE

Trade union strength, business power, and labor policy reform: The cases of Argentina and Chile in comparative perspective

Pablo Pérez Ahumada
University of Chile, Chile

Abstract
In this article, I explain why pro-labor reforms succeed or fail. Focusing on the cases of Argentina and Chile, I show that labor reforms are more successful in extending trade union rights when unions successfully build associational power and employers are less able to do so. Consistent with this argument, a quantitative analysis of time-series cross-sectional data from 78 countries suggests that the level of class power disparity is negatively correlated with the extension of workers' collective rights. At the end of the article, I discuss how these results have implications for the study of labor reforms and power resources.

Keywords
Employer associations, labor reforms, labor rights, Latin America, power resources, trade unions

Salida de R (tabla paquete *texreg*)

```
================================
                          m1
--------------------------------
(Intercept)             9.513 ***

v2pepwrses_osp_Rev     -1.783 ***

--------------------------------
R^2                     0.310
Adj. R^2                0.301
Num. obs.              78
================================
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1
```

Ecuación de la recta

$$\acute{Y} = a + \beta x$$

$$\acute{Y} = 9{,}513 - 1{,}783x$$

Interpretación:

Salida de R (tabla paquete *texreg*)

```
==============================
                         m1
------------------------------
(Intercept)            9.513 ***

v2pepwrses_osp_Rev    -1.783 ***

------------------------------
R^2                    0.310
Adj. R^2               0.301
Num. obs.             78
==============================
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1
```

Ecuación de la recta

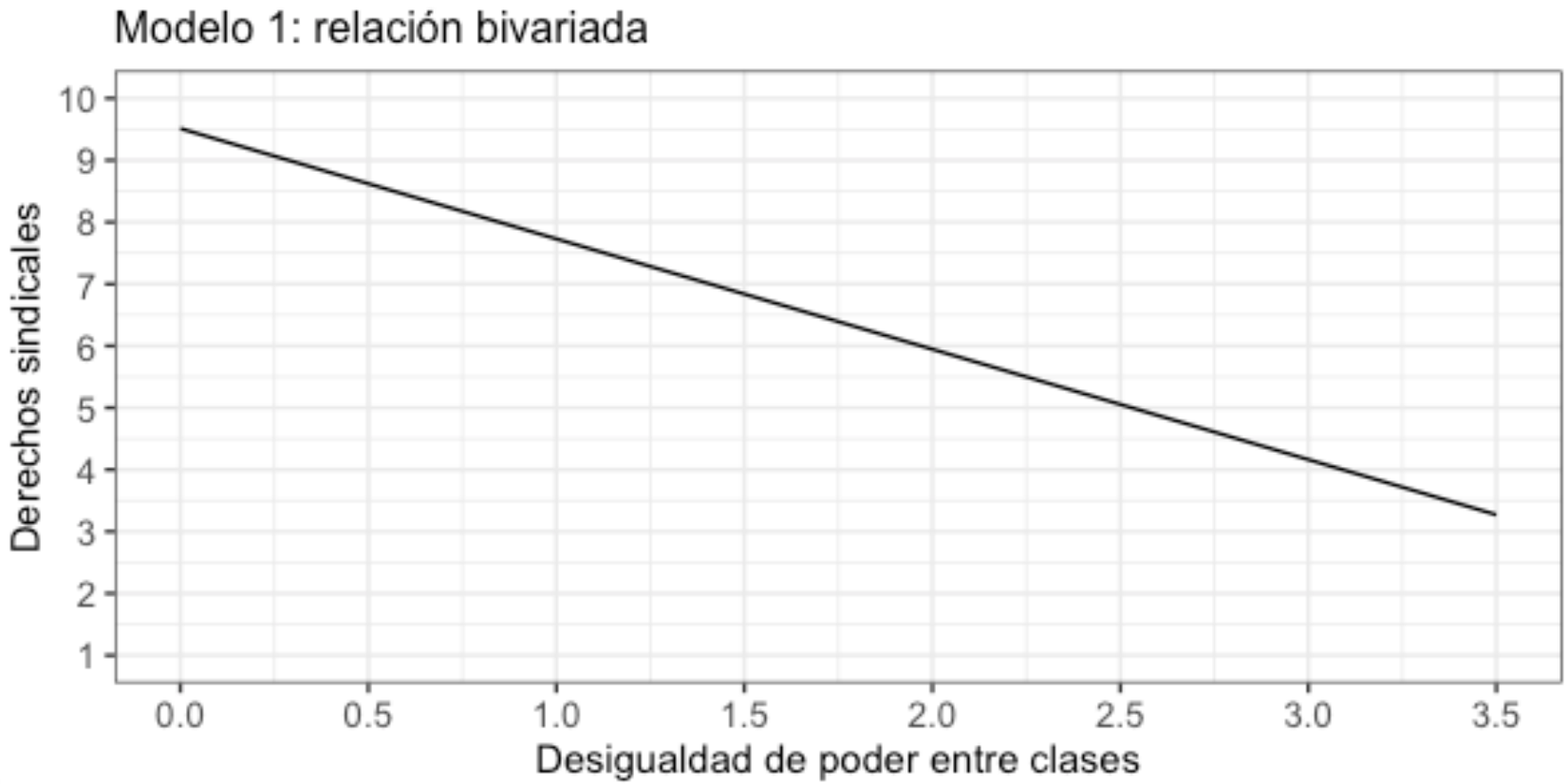
$$\acute{Y} = a + \beta x$$

$$\acute{Y} = 9{,}513 - 1{,}783x$$

Interpretación:

- Existe una relación negativa entre poder de clase y derechos laborales.
- Por cada unidad en que aumenta el índice de desigualdad de poder entre clases, el índice de derechos laborales disminuye en 1,8 puntos

Representación gráfica (paquete *ggplot2* en *R*)



INFERENCIA ESTADÍSTICA

Inferencia estadística

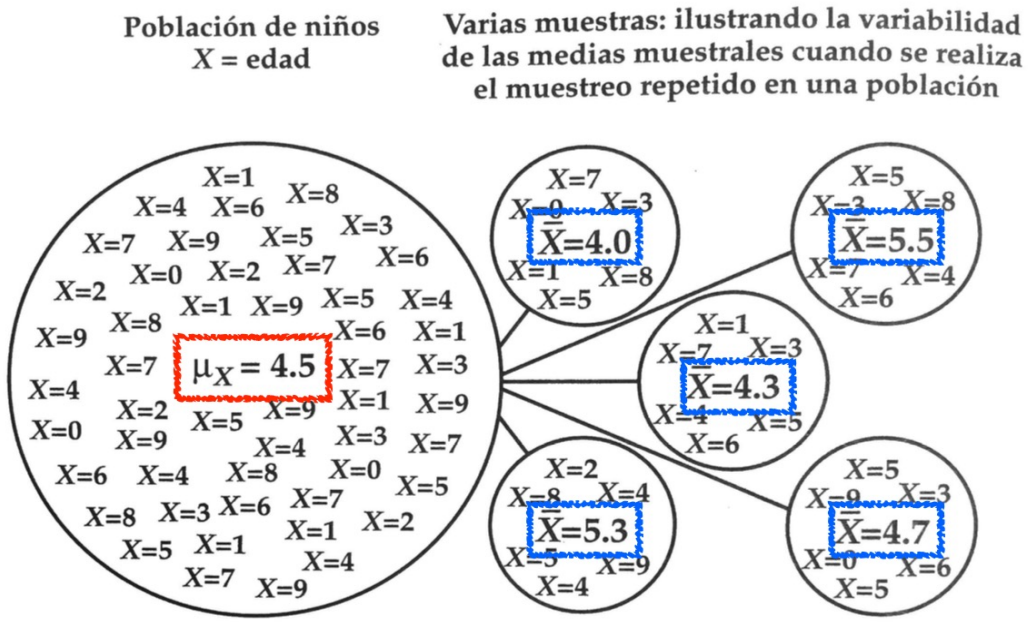
- ¿Cómo sabemos si el coeficiente de regresión β es estadísticamente significativo?

- ¿Cómo sabemos si nuestros resultados se pueden extrapolar a la población?

Aspectos esenciales

- Conceptos básicos:
  - Población → parámetros
  - Muestra → estadísticos

	Muestra (estadísticos)	Población (parámetros)
Promedio	$\bar{X}$	μ_x
Desviación estándar	S_X	σ_x



Fuente: Ritchey, 2008: 208

Distribución muestral

- Forma que toman los resultados de varias muestras, luego de **muestreos sucesivos**

- Al realizar muestreos repetidos, se ha comprobado que:
    1. los resultados varían de una muestra a otra

Distribución muestral

- Forma que toman los resultados de varias muestras, luego de **muestreos sucesivos**

- Al realizar muestreos repetidos, se ha comprobado que:
  1. los resultados varían de una muestra a otra
  2. dichos resultados están ligeramente errados de los valores reales de los parámetros de la población

Distribución muestral

- Forma que toman los resultados de varias muestras, luego de **muestreos sucesivos**

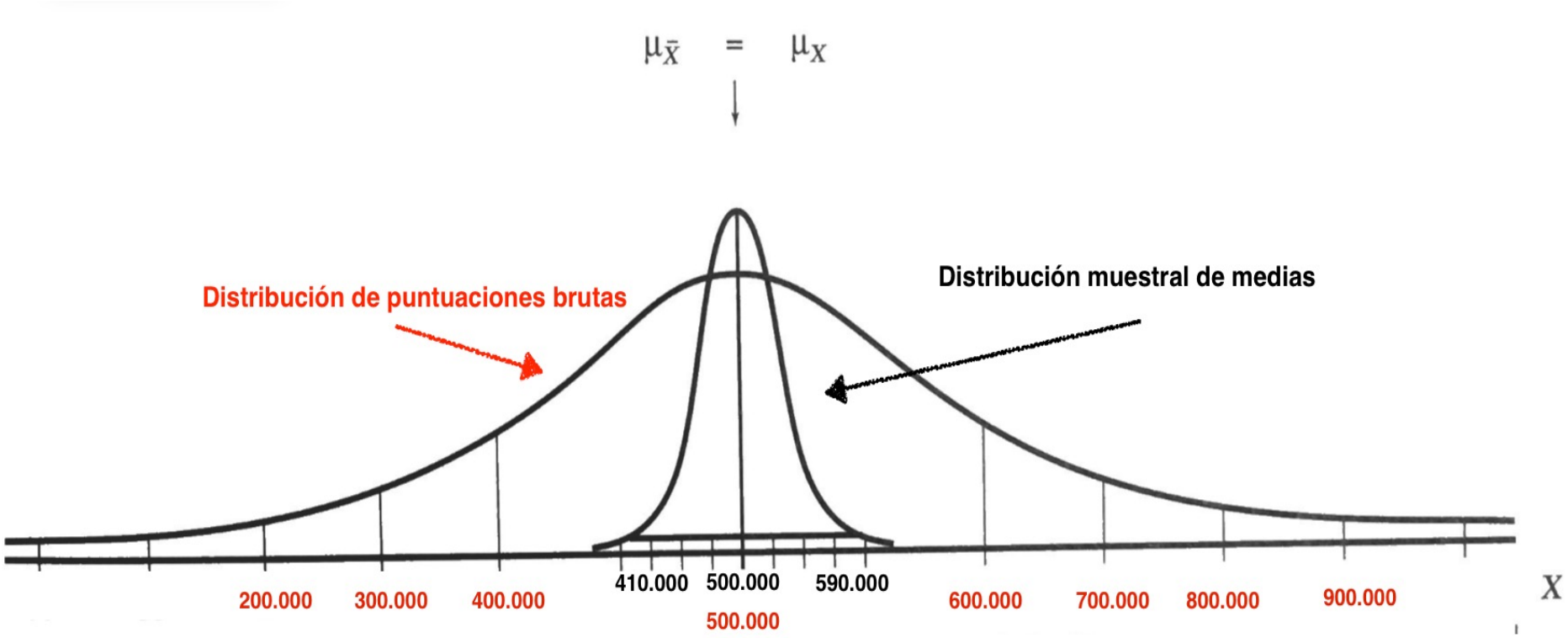
- Al realizar muestreos repetidos, se ha comprobado que:
  1. los resultados varían de una muestra a otra
  2. dichos resultados están ligeramente errados de los valores reales de los parámetros de la población
  3. *el **error es sistemático***, tiene patrones reconocibles y por lo tanto es predecible. Esto, porque:

Distribución muestral

- Forma que toman los resultados de varias muestras, luego de **muestreos sucesivos**

- Al realizar muestreos repetidos, se ha comprobado que:
  1. los resultados varían de una muestra a otra
  2. dichos resultados están ligeramente errados de los valores reales de los parámetros de la población
  3. *el **error es sistemático***, tiene patrones reconocibles y por lo tanto es predecible. Esto, porque:
     I. Las medias muestrales tienden a agruparse en torno a la media poblacional (Teorema Límite Central)
     II. La variabilidad de los muestreos se puede predecir de forma matemática a partir de *curvas de probabilidad (histogramas)*
     III. A mayor tamaño de la muestra, menor es el rango de los errores en muestras repetidas (Ley de los Grandes Números)

Distribución muestral



Notar que mientras las puntuaciones brutas (en rojo) varían mucho, las medias muestrales no varían tanto (su dispersión en torno a la media poblacional es mucho menor)

Error estándar

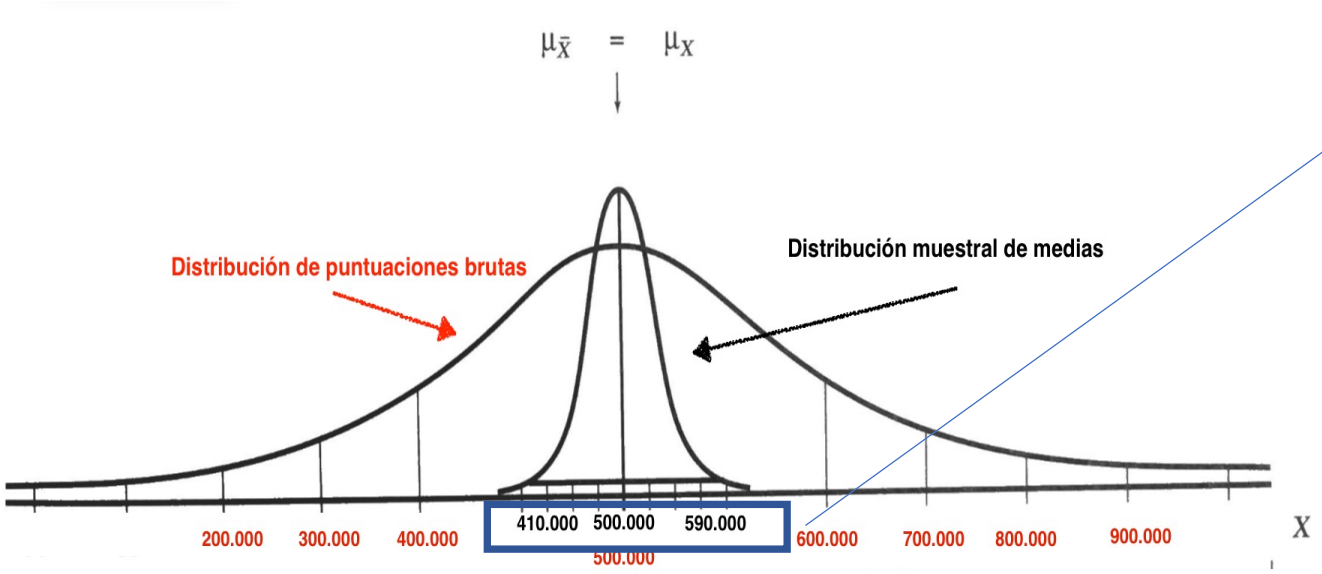
- Corresponde a la “desviación estándar” de una distribución muestral.

- Es importante porque mide la dispersión del error de muestreo que ocurre cuando se muestrea repetidamente una población (Ritchey 2008, p. 211).

Error estándar

- Corresponde a la “desviación estándar” de una distribución muestral.

- Es importante porque mide la dispersión del error de muestreo que ocurre cuando se muestrea repetidamente una población (Ritchey 2008, p. 211).



Notar que mientras las puntuaciones brutas (en rojo) varían mucho, las medias muestrales no varían tanto (su dispersión en torno a la media poblacional es mucho menor)

Error estándar

$$S_{\bar{X}} = \frac{S_X}{\sqrt{n}}$$

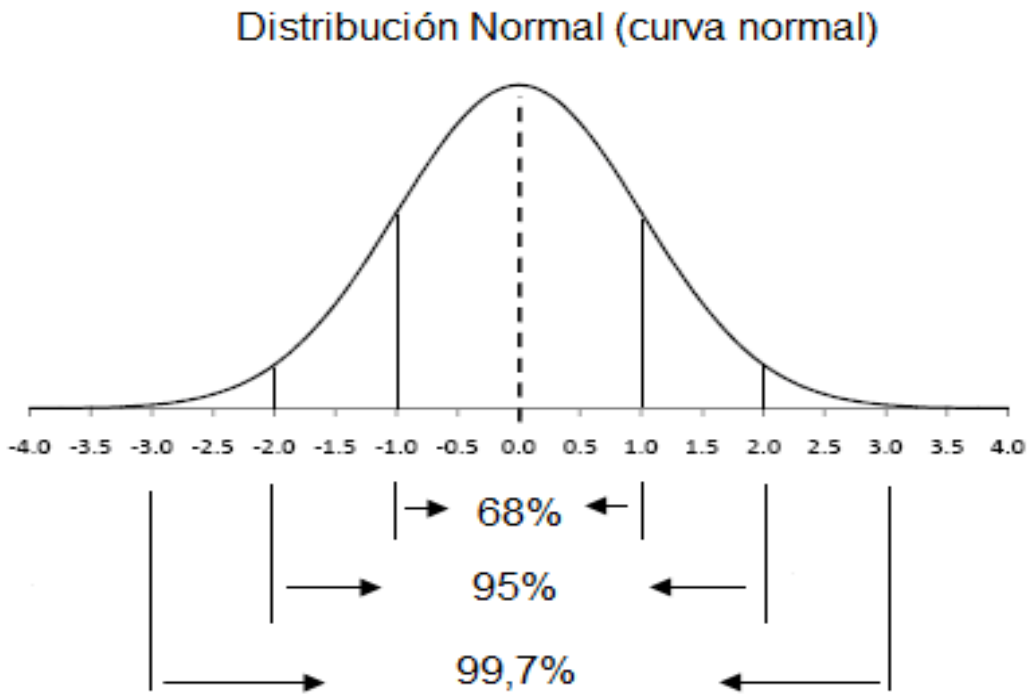
Donde:

S_X = desviación estándar de la muestra

n = tamaño de la muestra

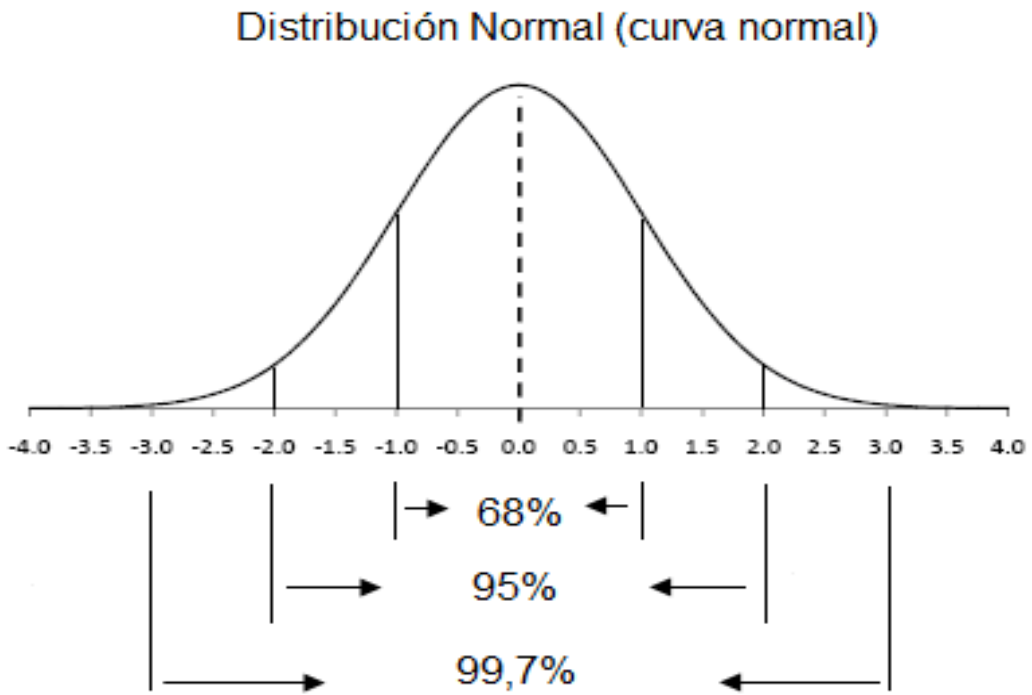
Error estándar y distribución de probabilidades

- La distribución muestral se comporta de forma normal (cuando tenemos muestras grandes)



Error estándar y distribución de probabilidades

- La distribución muestral se comporta de forma normal (cuando tenemos muestras grandes)



Gracias a ello, podemos saber que:

- Casi todas las medias muestrales (99,7%) caen dentro de 3 unidades de EE en ambas direcciones
- Cerca del 95% de las puntuaciones caen dentro de 2 unidades de EE
- Alrededor del 68% de las puntuaciones caen dentro de 1 unidad de EE

Valor - p

- A partir de lo anterior, podemos calcular la probabilidad de error de nuestras estimaciones

- Esta probabilidad está indicada en el **valor-p**
  - Indica la probabilidad de encontrar los resultados obtenidos en nuestra muestra (ej., coeficiente beta = - 1,789) *cuando la hipótesis nula (Ho) a nivel poblacional es verdadera*

- En palabras más imples, el valor-p muestra qué tan ***incompatibles*** son nuestros datos con la Ho

Inferencia y regresión

- En regresión nos interesa saber si existe una relación estadísticamente significativa entre variables (por ej., si la desigualdad de poder entre clases afecta el nivel de extensión de derechos sindicales)

- Esto se expresa en el contraste de 2 hipótesis:

$$Ho: \beta x = 0$$

$$Ha: \beta x \neq 0$$

Prueba T

- En regresión, la prueba estadística asociada a los coeficientes beta es la Prueba T . Ésta se calcula como una razón entre el coeficiente beta obtenido y el error estándar

$$t = \frac{b_j}{EE(b_j)}$$

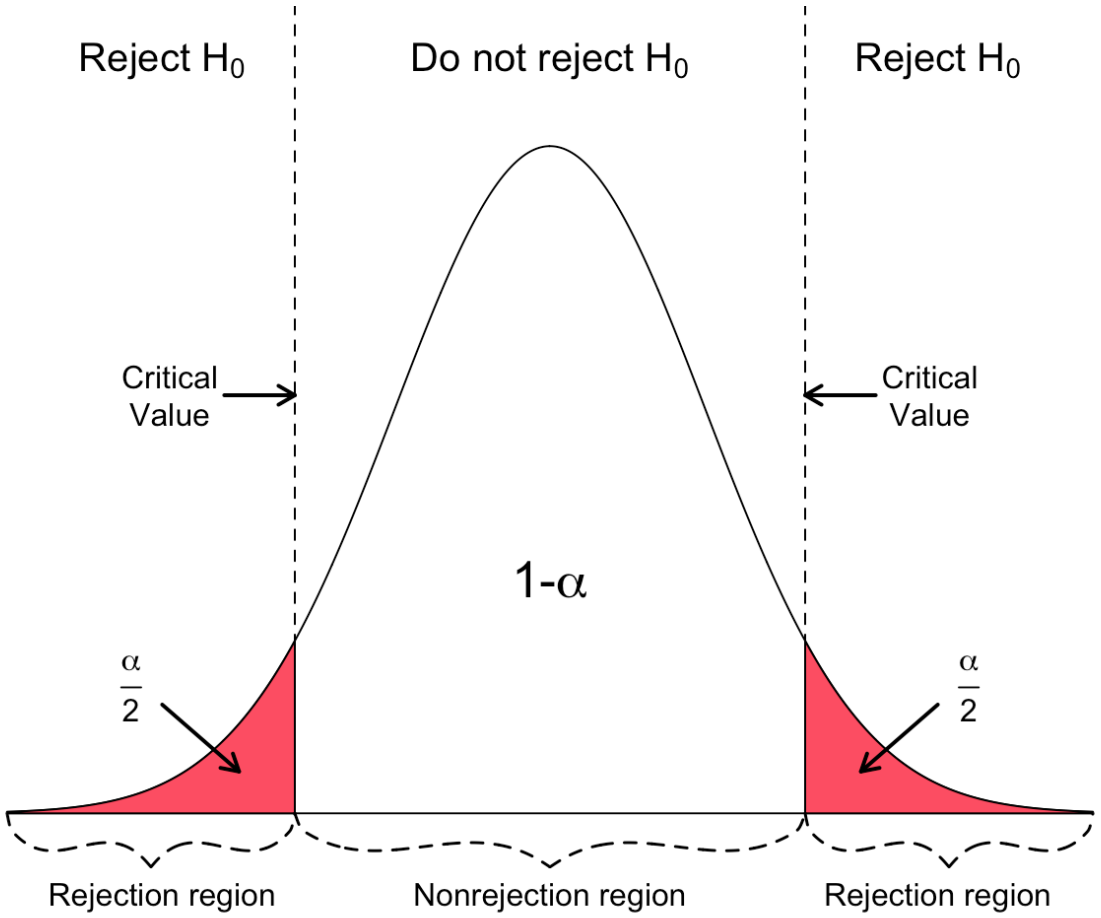
Prueba T

- En regresión, la prueba estadística asociada a los coeficientes beta es la Prueba T . Ésta se calcula como una razón entre el coeficiente beta obtenido y el error estándar

$$t = \frac{b_j}{EE(b_j)}$$

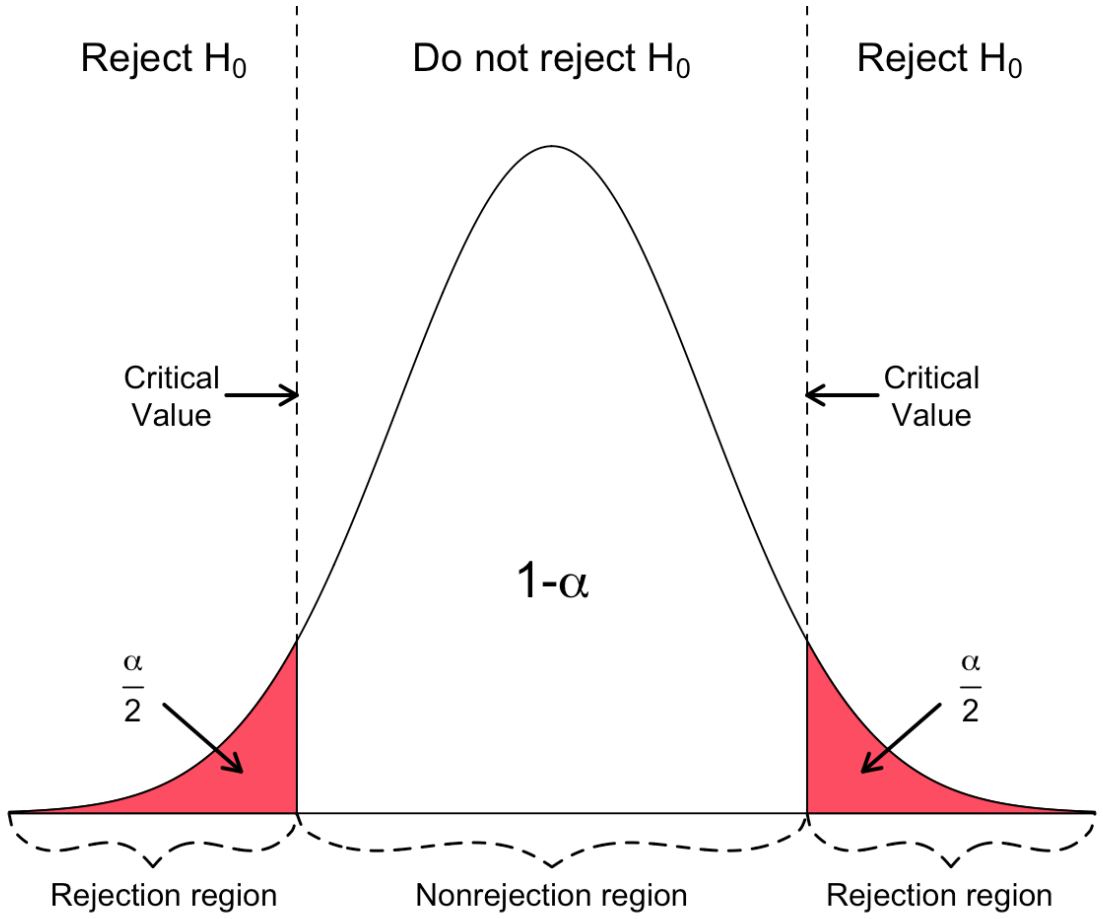
- El valor de t se compara con un “valor crítico”, que se obtiene de una tabla que muestra cómo sería dicho valor según
  - nivel de significancia con el que queremos hacer la estimación (α = 0,05; NC = 95%)
  - grados de libertad del modelo (n-k-1) (n = tamaño muestra; k = número de predictores)

Ejemplo zonas de aceptación/rechazo Ho, prueba de sign. bilateral (2 colas)



Fuente: https://www.geo.fu-berlin.de/en/v/soga/Basics-of-statistics/Hypothesis-Tests/Introduction-to-Hypothesis-Testing/Critical-Value-and-the-p-Value-Approach/index.html

Ejemplo zonas de aceptación/rechazo Ho, prueba de sign. bilateral (2 colas)



Idea central:

Si obtengo un resultado altamente inusual (*en condiciones en que Ho es cierta*), mi resultado muestral estará en la zona roja

Como el valor crítico de t depende de α y de GL, a medida en que el α disminuya será más difícil caer en la zona roja, por lo que será más difícil rechazar Ho

α = 0,05 –> NC = 95%
α = 0,01 –> NC = 99%
α = 0,001 –> NC = 99,9%

Fuente: https://www.geo.fu-berlin.de/en/v/soga/Basics-of-statistics/Hypothesis-Tests/Introduction-to-Hypothesis-Testing/Critical-Value-and-the-p-Value-Approach/index.html

Vuelta a nuestro ejemplo de regresión simple

- ¿Existe una relación entre la desigual distribución de poder entre clases y el nivel de extensión de los derechos sindicales?

Resultado de R (comando *summary*)

```
Call:
lm(formula = LR_Overall_Rev ~ v2pepwrses_osp_Rev, data = LaborRights_Data2017)

Residuals:
    Min      1Q  Median      3Q     Max
-4.9384 -0.8893  0.1265  1.1800  3.4490

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         9.5126     0.4601   20.67  < 2e-16 ***
v2pepwrses_osp_Rev -1.7832     0.3048   -5.85 1.17e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.676 on 76 degrees of freedom
Multiple R-squared:  0.3105,    Adjusted R-squared:  0.3014
F-statistic: 34.22 on 1 and 76 DF,  p-value: 1.167e-07
```

Modelo de regresión

Resultado de R (comando *summary*)

```
Call:
lm(formula = LR_Overall_Rev ~ v2pepwrses_osp_Rev, data = LaborRights_Data2017)

Residuals:
    Min      1Q  Median      3Q     Max
-4.9384 -0.8893  0.1265  1.1800  3.4490

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         9.5126     0.4601   20.67  < 2e-16 ***
v2pepwrses_osp_Rev -1.7832     0.3048   -5.85 1.17e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.676 on 76 degrees of freedom
Multiple R-squared:  0.3105,    Adjusted R-squared:  0.3014
F-statistic: 34.22 on 1 and 76 DF,  p-value: 1.167e-07
```

Modelo de regresión

Descripción de los residuos

Resultado de R (comando *summary*)

```
Call:
lm(formula = LR_Overall_Rev ~ v2pepwrses_osp_Rev, data = LaborRights_Data2017)

Residuals:
    Min      1Q  Median      3Q     Max
-4.9384 -0.8893  0.1265  1.1800  3.4490

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         9.5126     0.4601   20.67  < 2e-16 ***
v2pepwrses_osp_Rev -1.7832     0.3048   -5.85 1.17e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.676 on 76 degrees of freedom
Multiple R-squared:  0.3105,    Adjusted R-squared:  0.3014
F-statistic: 34.22 on 1 and 76 DF,  p-value: 1.167e-07
```

Modelo de regresión

Descripción de los residuos

Coeficientes, error estándar, valor de prueba T, valor-p y nivel de significancia

Resultado de R (comando *summary*)

```
Call:
lm(formula = LR_Overall_Rev ~ v2pepwrses_osp_Rev, data = LaborRights_Data2017)

Residuals:
    Min      1Q  Median      3Q     Max
-4.9384 -0.8893  0.1265  1.1800  3.4490

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         9.5126     0.4601   20.67  < 2e-16 ***
v2pepwrses_osp_Rev -1.7832     0.3048   -5.85 1.17e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.676 on 76 degrees of freedom
Multiple R-squared:  0.3105,    Adjusted R-squared:  0.3014
F-statistic: 34.22 on 1 and 76 DF,  p-value: 1.167e-07
```

Modelo de regresión

Descripción de los residuos

Coeficientes, error estándar, valor de prueba T, valor-p y nivel de significancia

Calidad del modelo

R^2 y R^2 ajustado

Prueba F: muestra en qué medida el modelo mejora la capacidad explicativa (de la varianza de Y) en relación a un modelo *sin* predictores

- Ho: ambos modelos son iguales
- Ha: modelo con predictores explica más varianza que modelo nulo

Presentación convencional de modelo de regresión, comando *textreg* en R (errores estándares entre paréntesis)

```
================================
                         m1
--------------------------------
(Intercept)               9.513 ***
                         (0.460)
v2pepwrses_osp_Rev       -1.783 ***
                         (0.305)
--------------------------------
R^2                       0.310
Adj. R^2                  0.301
Num. obs.                78
================================
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1
```

Sugerencias:

- reportar sólo el R^2 ajustado
- Reportar siempre el número de observaciones

REGRESIÓN MÚLTIPLE

Regresión múltiple / inferencia estadística

- **Idea clave:** Usar la recta recién descrita para estimar la recta de regresión “real”, definida como

$$Y = \alpha + \beta X + \varepsilon_i$$

Donde e_i: término de error *aleatorio*

Regresión múltiple

- Cuando se trabaja con una regresión múltiple, el modelo general que se pretende estimar es:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n + \varepsilon_i$$

Regresión múltiple

- Ahora el efecto de una variable X_1 (expresado en el coeficiente β_1) se interpreta de modo similar a una correlación parcial—es decir, *manteniendo controladas las otras variables del modelo* (X_2, X_3, X_n, etc.)

- Formas comunes de expresar esto:
  - El impacto de X sobre Y, *manteniendo constante el efecto de las otras variables*, es de…
  - *Ceteris paribus*, la relación entre X e Y es…
  - El efecto *neto* de X sobre Y es de…

REGRESIÓN MÚLTIPLE: ejemplo

Ejemplo regresión múltiple

- La relación existente entre desigual distribución de poder entre clases y nivel de extensión de los derechos sindicales, ¿se mantiene robusta al mantener otras variables relevantes?

- ¿Qué son “variables relevantes”?

Ejemplo regresión múltiple

- La relación existente entre desigual distribución de poder entre clases y nivel de extensión de los derechos sindicales, ¿se mantiene robusta al mantener otras variables relevantes?

- ¿Qué son “variables relevantes”?

- Variables de control incluidas en este análisis
  - **Controles económicos**: PIB per cápita (*GDPpp_log*); Inversión extranjera directa (*FDI_inflow*)
  - **Controles políticos**: Grado de democracia (*v2x_libdem_InPerc*); Gobierno de Izquierda (1 =sí, 0 = no)

¿Cómo cambia el coeficiente de desigualdad de poder (*v2pepwrses_osp_Rev)* a medida que se van agregando controles?

	m1	m2: econ	m3: pol
(Intercept)	9.513 ***	3.053 †	2.514
	(0.460)	(1.744)	(1.784)
v2pepwrses_osp_Rev	-1.783 ***	-1.111 **	-0.688 †
	(0.305)	(0.329)	(0.370)
GDPpp_log		0.584 ***	0.431 *
		(0.156)	(0.187)
FDI_inflow		0.011	0.012
		(0.010)	(0.010)
v2x_libdem_InPerc			0.022 †
			(0.011)
LeftGvt			0.481
			(0.399)
R^2	0.310	0.430	0.471
Adj. R^2	0.301	0.407	0.434
Num. obs.	78	78	78

*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1

Ojo:
Gobierno de Izquierda es una variable categórica

¿Cómo se interpreta esto?

	m1	m2: econ	m3: pol
(Intercept)	9.513 ***	3.053 †	2.514
	(0.460)	(1.744)	(1.784)
v2pepwrses_osp_Rev	-1.783 ***	-1.111 **	-0.688 †
	(0.305)	(0.329)	(0.370)
GDPpp_log		0.584 ***	0.431 *
		(0.156)	(0.187)
FDI_inflow		0.011	0.012
		(0.010)	(0.010)
v2x_libdem_InPerc			0.022 †
			(0.011)
LeftGvt			0.481
			(0.399)
R^2	0.310	0.430	0.471
Adj. R^2	0.301	0.407	0.434
Num. obs.	78	78	78

*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1

Ojo:
Gobierno de Izquierda es una variable categórica

¿Cómo se interpreta esto?

En estos casos, se debe tomar como referencia la categoría *omitida*

	m1	m2: econ	m3: pol
(Intercept)	9.513 ***	3.053 †	2.514
	(0.460)	(1.744)	(1.784)
v2pepwrses_osp_Rev	-1.783 ***	-1.111 **	-0.688 †
	(0.305)	(0.329)	(0.370)
GDPpp_log		0.584 ***	0.431 *
		(0.156)	(0.187)
FDI_inflow		0.011	0.012
		(0.010)	(0.010)
v2x_libdem_InPerc			0.022 †
			(0.011)
LeftGvt			0.481
			(0.399)
R^2	0.310	0.430	0.471
Adj. R^2	0.301	0.407	0.434
Num. obs.	78	78	78

*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1

Ojo:
Gobierno de Izquierda es una variable categórica

¿Cómo se interpreta esto?

En estos casos, se debe tomar como referencia la categoría *omitida*

	m1	m2: econ	m3: pol
(Intercept)	9.513 ***	3.053 †	2.514
	(0.460)	(1.744)	(1.784)
v2pepwrses_osp_Rev	-1.783 ***	-1.111 **	-0.688 †
	(0.305)	(0.329)	(0.370)
GDPpp_log		0.584 ***	0.431 *
		(0.156)	(0.187)
FDI_inflow		0.011	0.012
		(0.010)	(0.010)
v2x_libdem_InPerc			0.022 †
			(0.011)
LeftGvt			0.481
			(0.399)
R^2	0.310	0.430	0.471
Adj. R^2	0.301	0.407	0.434
Num. obs.	78	78	78

*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1

En R, la categoría omitida se puede definir construyendo variables dummy.
Si se incluye directamente una variable categórica (factor), la primera categoría será la omitida