

Clase 13: Regresión lineal

STATA



 python

Econometría

- La econometría cuenta con la ayuda de las matemáticas, la estadística y la economía para identificar relaciones entre variables
- Intenta simplificar la realidad para estudiarla en un modelo econométrico
- La correlación lineal es un primer acercamiento al análisis
- Las regresiones permiten ir un paso adelante y ayudan a establecer relaciones de causa-efecto entre variables

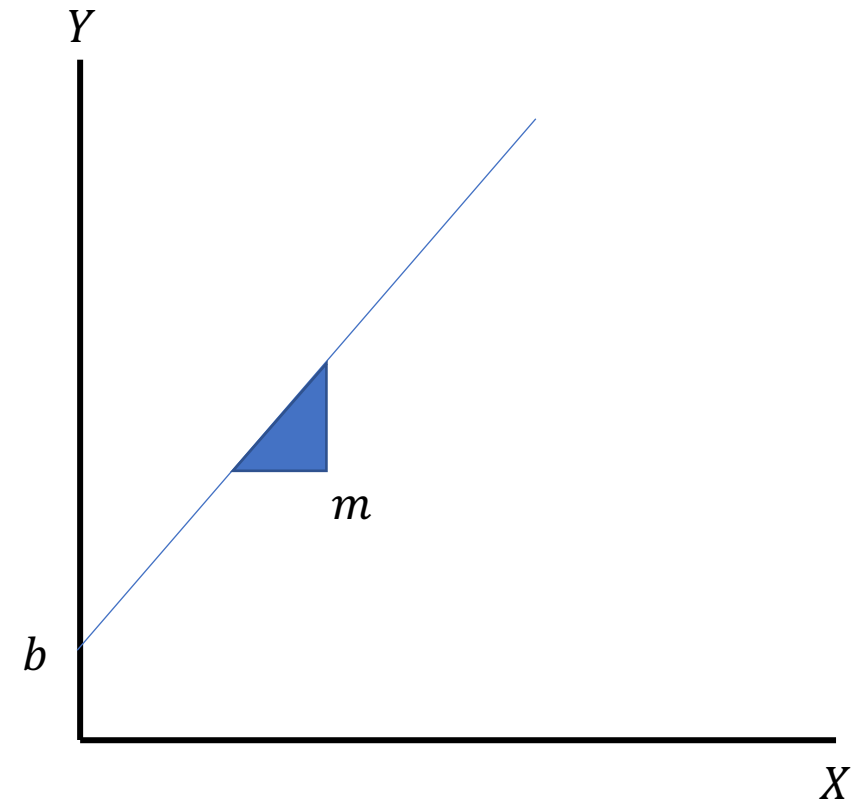
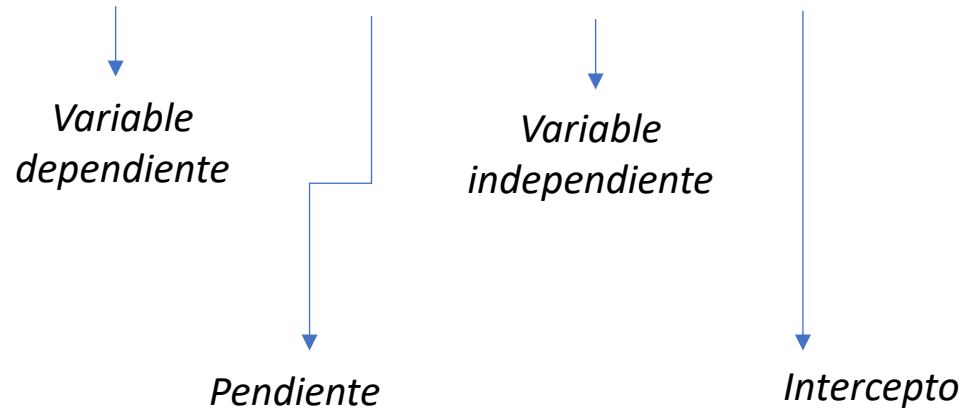
Coeficiente de correlación

- El coeficiente de correlación identifica el grado y sentido de la relación lineal entre un par de variables
- Está acotado en el intervalo $-1 \leq r \leq 1$
- Valores cercanos a $|1|$ indican una fuerte relación (negativa o positiva). Valores cercanos a cero indican la presencia de poca relación.
- Debe emplearse para variables continuas
- ¡Pero! Correlación no es causalidad



Regresión: Función lineal

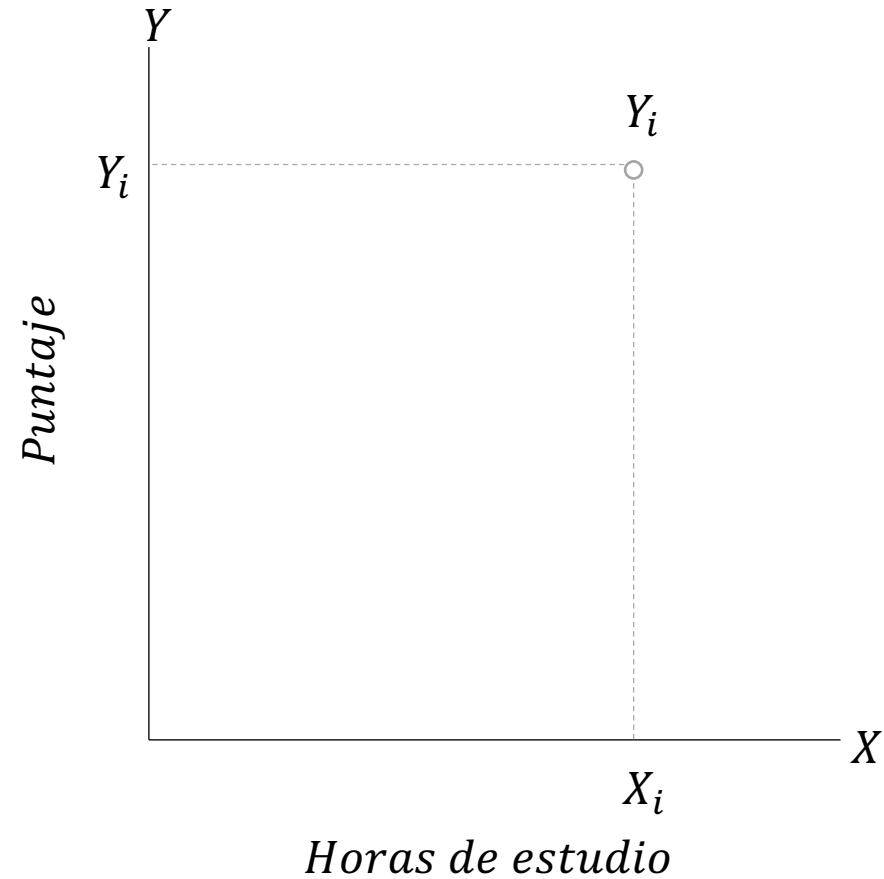
$$Y = m \times X + b$$



Función de regresión poblacional y función de regresión muestral

Suponga el ejemplo de resultados en una prueba estandarizada Y en función de las horas de estudio X .

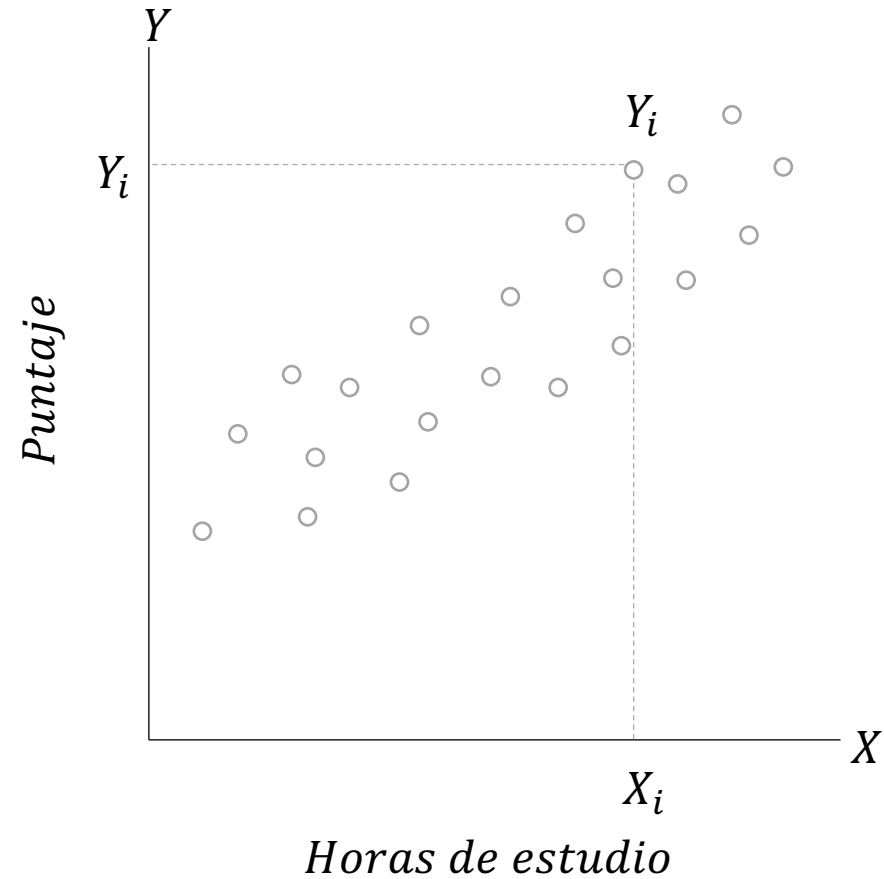
Se selecciona una observación i , para las horas de estudio, denominado X_i y su correspondiente puntaje en la prueba estandarizada, denominado Y_i .



Función de regresión poblacional y función de regresión muestral

Suponga el ejemplo de resultados en una prueba estandarizada Y en función de las horas de estudio X .

Se selecciona una observación i , para las horas de estudio, denominado X_i y su correspondiente puntaje en la prueba estandarizada, denominado Y_i .

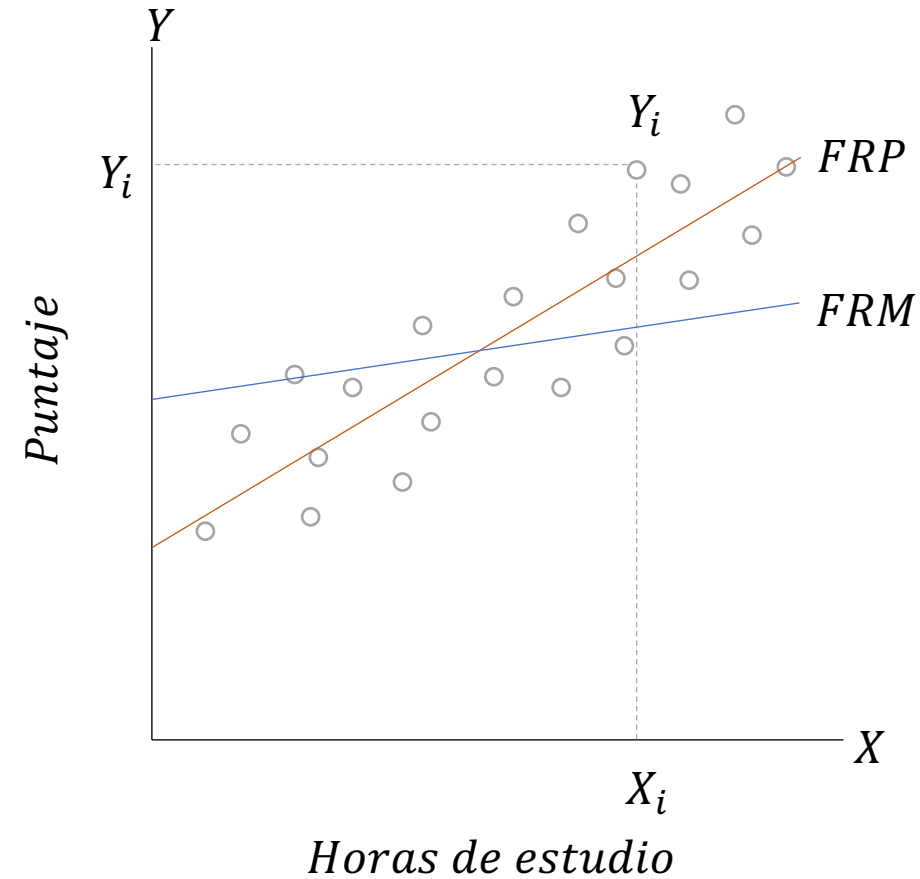


Función de regresión poblacional y función de regresión muestral

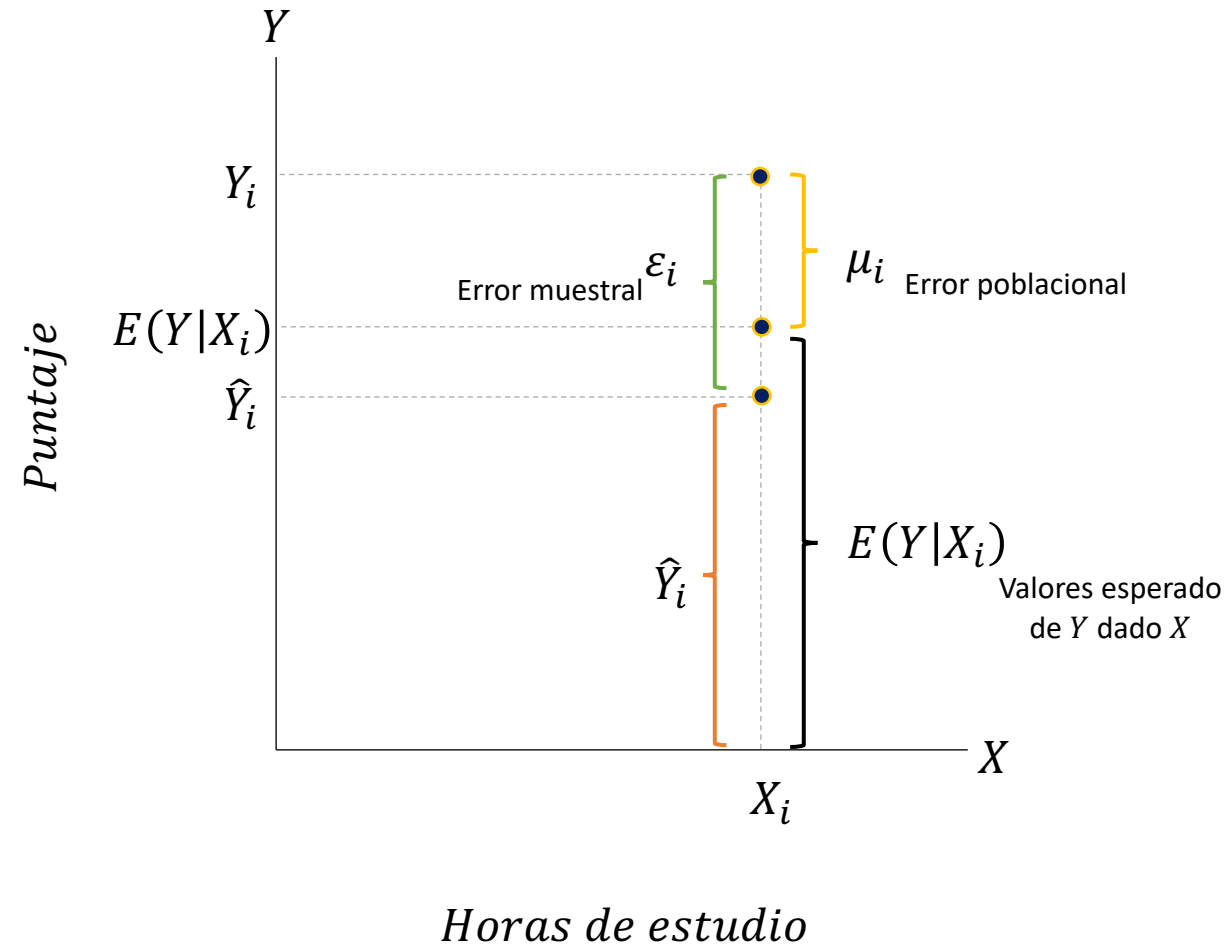
La FRP es la línea que ajusta la nube de puntos correspondiente al diagrama de dispersión de la población.

Cuando se hace la predicción con la FRP se obtiene el valor esperado de Y_i , dado X_i , que se simboliza con la expresión $E(Y_i | X_i)$.

La FRM corresponde a la predicción de los valores estimados \hat{Y}_i .



Función de regresión muestral



¿Cómo se estima la FRM?

- Existen diferentes métodos. Dependen de las características de la variable dependiente incluida.
- El modelo más empleado por su facilidad y su integralidad es el modelo de *Mínimos Cuadrados Ordinarios* (MCO).
- Este método asegura que se obtengan medidas de la relación entre las variables independientes y la variable dependiente conocidas como β .
- Si el investigador cumple con los supuestos del modelo, puede obtener estimadores (β) con excelentes propiedades.

MCO: regresión simple

$$FRP: Y_i = \beta_0 + \beta_1 X_i + \mu_i$$



$$FRM: Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \varepsilon_i$$



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

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$



$$\hat{\beta}_{1_{MCO}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{cov(X, Y)}{var(X)}$$



$$\hat{\beta}_{0_{MCO}} = \bar{Y} - \hat{\beta}_1 \bar{X}$$

MCO: verificación

- Una vez estimada la regresión se debe verificar su validez estadística y también la de los estimadores obtenidos.
- La validez de la regresión se estima con una prueba de hipótesis, indicando que ninguna variable independiente (X) es relevante para explicar a la variable dependiente (Y). Para tal fin, se usa la distribución *F de Fisher*.
- Los estimadores se verifican de manera individual y se comparan con el tamaño de sus errores estándar (como la desviación estándar pero de los estimadores). Para tal fin, se usa la distribución *t de Student*.

Ejemplo

STATA



Ejemplo

OLS Regression Results

Dep. Variable:	wage	R-squared:	0.165
Model:	OLS	Adj. R-squared:	0.161
Method:	Least Squares	F-statistic:	45.79
Date:	Tue, 18 May 2021	Prob (F-statistic):	3.87e-35
Time:	14:51:09	Log-Likelihood:	-6854.3
No. Observations:	935	AIC:	1.372e+04
Df Residuals:	930	BIC:	1.374e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-426.1507	110.596	-3.853	0.000	-643.197	-209.104
educ	74.8600	6.222	12.031	0.000	62.649	87.071
exper	13.7998	3.228	4.275	0.000	7.465	20.135
married	179.8177	39.470	4.556	0.000	102.358	257.278
tenure	7.6976	2.475	3.111	0.002	2.841	12.554

Omnibus:	235.973	Durbin-Watson:	1.797
Prob(Omnibus):	0.000	Jarque-Bera (JB):	734.428
Skew:	1.228	Prob(JB):	3.32e-160
Kurtosis:	6.580	Cond. No.	180.