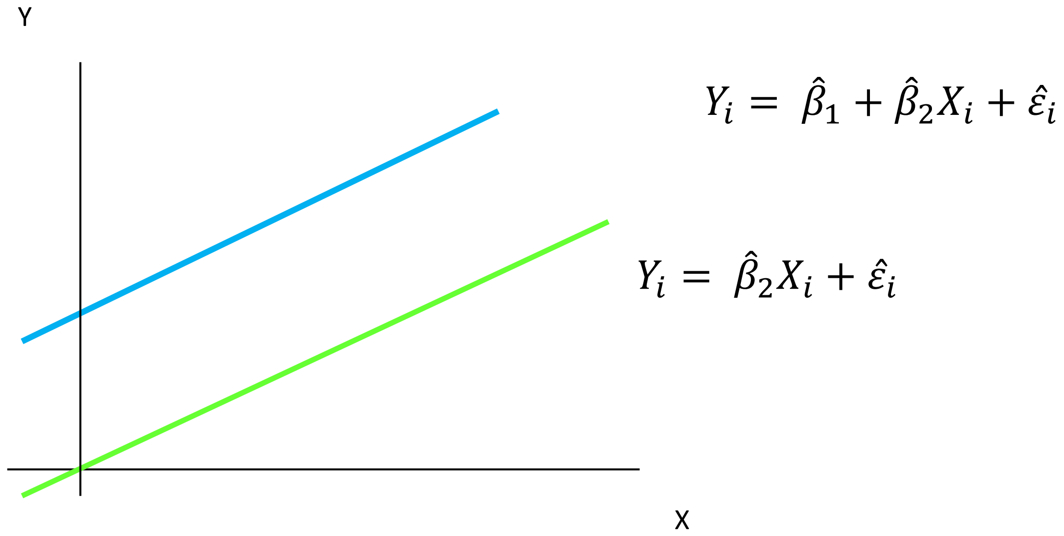
EXTENSIÓN DEL MODELO DE REGRESIÓN LINEAL DE DOS VARIABLES

REGRESIÓN A TRAVÉS DEL ORIGEN



“A MENOS QUE EXISTA UNA EXPECTATIVA A PRIORI MUY FUERTE ES ACONSEJABLE APEGARSE AL MODELO CONVENCIONAL CON PRESENCIA DE INTERSECCIÓN”

¿QUÉ OCURRE CON LOS MODELOS SIN INTERCEPTO?

- Si el modelo que se va estimar teóricamente debe llevar intersección y ésta no se incluye, se puede cometer un error de especificación y con ello violar un supuesto del MCRL, dando origen a estimadores **NO MELI.**

- La $\sum \hat{\varepsilon}_i$ que siempre es cero en modelos con intercepto puede no serlo cuando está ausente $\hat{\beta}_1$. Es decir que los residuos ya no tienen media muestral igual a cero $E(\varepsilon_i|X_i) \neq 0$.

- El coeficiente de determinación (R^2) que siempre es no negativo podría llegar a ser negativo, por tanto el coeficiente de determinación convencional no resulta adecuado para evaluar la bondad de ajuste del modelo. Para ello se utiliza el R^2 simple, que se interpreta de la manera ya conocida pero no se puede comparar con el R^2 tradicional.

$$R^2 simple = \frac{(\sum X_i Y_i)^2}{\sum {X_i}^2 \sum {Y_i}^2}$$

- Si en el modelo poblacional el intercepto es diferente de cero, y se usa un modelo muestral con $\hat{\beta}_1 = 0$, los estimadores MCO de la pendiente estarán sesgados.

- El costo de estimar $\hat{\beta}_1$ cuando realmente $\hat{\beta}_1 = 0$, es que las varianzas de los estimadores MCO son mayores (no mínimas).

En resumen:

MODELO CON INTERCEPTO	MODELO SIN INTERCEPTO
$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$	$Y_i = \hat{\beta}_2 X_i + \hat{u}_i$
$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum {x_i}^2}$	$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum {X_i}^2}$
$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum {x_i}^2}$	$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum {X_i}^2}$
$\hat{\sigma}^2 = \frac{\sum {\hat{u}_i}^2}{n-2}$	$\hat{\sigma}^2 = \frac{\sum {\hat{u}_i}^2}{n-1}$

“Si la intercepción efectivamente está ausente, el coeficiente de la pendiente puede ser estimado con mucha más precisión que cuando el término de la intersección está incluido” H. Theil (1978)

Ejemplo 1[1]: Considere el modelo de asignación de precios de activos de capital (CAMP) de la teoría moderna de portafolios, expresada

$$(ER_i - r_f) = \beta_i (ER_m - r_f) \qquad (^2)$$

El modelo puede expresarse como:

$R_i - r_f = \beta_i (R_m - r_f) \quad + u_i$

$R_i - r_f = \alpha_i + \beta_i (R_m - r_f) \quad + u_i$ MODELO DEL MERCADO

Si se cumple el CAMP, entonces $\alpha_i = 0$

La siguiente regresión muestra los datos referentes a los rendimientos excedentes (Yt) en %, de índice de 104 acciones del sector de bienes de consumo cíclico y los rendimientos excedentes (Xt) en % del índice de todo el mercado de valores del Reino Unido para 1980:01-1999:12

[1] Revisar Haim Levy y Marsahall Sarnta, (1984). ***Portfolio and Investment Selection: Theory and Practice.*** Cap. 14

[2] ER_i: tasa esperada de rendimiento del título i. ER_m: tasa esperada del rendimientos del portafolios del mercado como se representa. r_f: tasa de rendimiento libre de riesgo. B_i: coeficiente BETA (medida de riesgo sistemático, es decir, riesgo que no se ha eliminado con la diversificación)

```
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## Xt   1.1555     0.0744   15.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.549 on 239 degrees of freedom
## Multiple R-squared:  0.5023, Adjusted R-squared:  0.5003
## F-statistic: 241.2 on 1 and 239 DF,  p-value: < 2.2e-16
```

¿Qué diferencias encuentra en los dos modelos?

```
## Coefficients:
##                Estimate   Std. Error   t value   Pr(>|t|)
## (Intercept)  -0.44748    0.36294      -1.233     0.219
## Xt            1.17113    0.07539      15.535    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.543 on 238 degrees of freedom
## Multiple R-squared:  0.5035, Adjusted R-squared:  0.5014
## F-statistic: 241.3 on 1 and 238 DF,  p-value: < 2.2e-16
```

Ejemplo 2: PIB en millones e inversión doméstica privada bruta millones (GPDIM), EEUU 1988-1997.

Dependent Variable: PIB
Sample: 1988 1997

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GPDIM	0.007259	0.001385	5.240237	0.0005

R-squared	-54.203821	Mean dependent var	6.440900
Adjusted R-squared	-54.203821	S.D. dependent var	0.455016
S.E. of regression	3.380733	Akaike info criterion	5.368702
Sum squared resid	102.8642	Schwarz criterion	5.398960
Log likelihood	-25.84351	Hannan-Quinn criter.	5.335508
Durbin-Watson stat	0.550763		

¿Cuál de los dos modelos es mejor? ¿Por qué?

Dependent Variable: PIB
Sample: 1988 1997

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.018338	0.245104	28.63408	0.0000
GPDIM	-0.000840	0.000318	-2.645248	0.0295

R-squared	0.466572	Mean dependent var	6.440900
Adjusted R-squared	0.399894	S.D. dependent var	0.455016
S.E. of regression	0.352485	Akaike info criterion	0.929238
Sum squared resid	0.993965	Schwarz criterion	0.989755
Log likelihood	-2.646191	Hannan-Quinn criter.	0.862851
F-statistic	6.997338	Durbin-Watson stat	0.724787
Prob(F-statistic)	0.029472		

ESCALAS Y UNIDADES DE MEDICIÓN

Los cambios en las escalas y en las unidades de medida se realizan con la finalidad expresar números en menores cantidades que permitan ser manejados fácilmente.

Sea $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\varepsilon}_i$ (1)

Defina: $Y_i{}^* = w_1 Y_i$

$X_i{}^* = w_2 X_i$

$\hat{\varepsilon}_i{}^* = w_1 \hat{\varepsilon}_i$

w_1 y w_2 Se denominan **FACTORES DE ESCALA** ($w_1 =$ ó $\neq w_2$)

$$Y_i{}^* = \hat{\beta}_1{}^* + \hat{\beta}_2{}^* X_i{}^* + \hat{\varepsilon}_i{}^* \qquad (2)$$

Aplicando en método de MCO a (2) se obtiene:

$$\hat{\beta}_1{}^* = \bar{Y}^* - \hat{\beta}_2{}^* \bar{X}^*$$

$$\hat{\beta}_2{}^* = \frac{\sum x_i{}^* y_i{}^*}{\sum x_i{}^{*2}}$$

$$Var(\hat{\beta}_1{}^*) = \frac{\sum X_i{}^{*2}}{n \sum x_i{}^{*2}} \sigma^{*2}$$

$$Var\left(\hat{\beta}_2{}^{*}\right) = \frac{\sigma^{*2}}{\sum x_i{}^{*2}}$$

$$\sigma^{*2} = \frac{\sum \hat{\varepsilon}_i{}^{*2}}{n-k}$$

Se pueden establecer las relaciones entre los dos conjuntos de parámetros estimados:

$$\hat{\beta}_2{}^{*} = \left(\frac{w_1}{w_2}\right)\hat{\beta}_2$$

$$\hat{\beta}_1{}^{*} = w_1\hat{\beta}_1$$

$$\sigma^{*2} = w_1{}^2\hat{\sigma}^2$$

$$Var\left(\hat{\beta}_1{}^{*}\right) = w_1{}^2 Var(\hat{\beta}_1)$$

$$Var\left(\hat{\beta}_2{}^{*}\right) = \left(\frac{w_1}{w_2}\right)^2 Var(\hat{\beta}_2)$$

$$R^2{}_{xy} = R^2{}_{x^*y^*}$$

Cambios en el origen:

$$Y_i^{\ *} = \ a + Y_i$$
$$X_i^{\ *} = \ b + X_i$$

Si $(a = \text{ó} \neq b)$

$$\hat{\beta}_2^{\ *} = \hat{\beta}_2$$
$$\rho_{xy} = \ \rho_{x^*y^*}$$

"LA TRANSFORMACION DE LA ESCALA (Y, X) A LA ESCALA (Y^*, X^*) NO AFECTA LAS PROPIEDADES DE LOS ESTIMADORES MCO"

Ejemplo 2: Inversión doméstica privada bruta (GPDIBL) en función del PIB (GDPB), ambos en miles de millones de dólares del 2000, EEUU 1990-2005.

```
## Call:
## lm(formula = GPDIBL ~ GDPB)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.346 -3.746  0.500  3.441  7.346
##
## Coefficients:
##             Estimate   Std. Error   t value    Pr(>|t|)
## (Intercept)  11.1250     2.4580      4.526     0.000475 ***
## GDPB         -0.3088     0.2542     -1.215     0.244499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.687 on 14 degrees of freedom
## Multiple R-squared:  0.09537,    Adjusted R-squared:  0.03076
## F-statistic: 1.476 on 1 and 14 DF,  p-value: 0.2445
```

Inversión doméstica privada bruta (GPDIM) en función del PIB (GDPPM) ambos en millones de dólares, EEUU 1990-2005.

```
## Call:
## lm(formula = GPDIM ~ GDPM)
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -7.346 -3.746  0.500  3.441  7.346
##
## Coefficients:
##              Estimate    Std. Error  t value    Pr(>|t|)
##(Intercept)   11.1250     2.4580      4.526    0.000475 ***
## GDPM         -0.3088     0.2542     -1.215    0.244499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.687 on 14 degrees of freedom
## Multiple R-squared:  0.09537,    Adjusted R-squared:  0.03076
## F-statistic: 1.476 on 1 and 14 DF,  p-value: 0.2445
```

REGRESIÓN SOBRE LAS VARIABLES ESTANDARIZADAS

La influencia que tiene sobre la interpretación de los efectos de las variables las unidades de medida en que éstas sean expresadas, puede evitarse si se expresan ambas variables como ***variables estandarizadas***.

Una variable estandarizada se expresa como:

$$Y_i^* = \frac{Y_i - \bar{Y}}{ee_y} \qquad\qquad X_i^* = \frac{X_i - \bar{X}}{ee_x}$$

La propiedad más resaltante de las variables estandarizadas es que el valor de su media siempre es cero y su desviación estándar es siempre 1.

$$Y_i^* \sim (0,1) \qquad\qquad X_i^* \sim (0,1)$$

La regresión con variables estandarizadas queda:

$$Y_i^* = \beta_2^* X_i^* + \varepsilon_i^*$$

¿POR QUÉ?, ¿QUÉ PASA CON EL COEFICIENTE DEL

PARA TENER EN CUENTA:

- Los coeficientes de regresión de las variables estandarizadas se conocen como coeficientes beta[3].

- Después de estandarizadas las variables ya no importa las unidades de medida.

- Los coeficientes beta se interpretan en términos de desviaciones estándar, es decir, se mide el efecto en las unidades de la desviación estándar.

- Los coeficientes dentro de la misma ecuación son comparables entre sí. Se pueden utilizar como una medida de fuerza relativa de las distintas regresoras.

[3] No se deben confundir estos coeficientes beta con los coeficientes beta de la teoría del portafolio.

Ejemplo 4: Regresión de inversión sobre PIB

VARIABLES ORIGINALES

Matriz de correlación

	PIB	INV
PIB	1.000000	0.164923
INV	0.164923	1.000000

Dependent Variable: INV
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	900.2423	58.96998	15.26611	0.0000
PIB	0.001281	0.002708	0.472948	0.6489

R-squared	**0.027200**	Mean dependent var	916.1100
Adjusted R-squared	-0.094401	S.D. dependent var	146.5929
S.E. of regression	153.3561	Akaike info criterion	13.08026
Sum squared resid	188144.8	Schwarz criterion	13.14078
Log likelihood	-63.40130	F-statistic	0.223680
Durbin-Watson stat	0.269266	Prob(F-statistic)	0.648885

VARIABLES ESTANDARIZADAS

Matriz de correlación

	INV	PIB
INV	1.000000	0.164923
PIB	0.164923	1.000000

Dependent Variable:INV
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.33E-10	0.330817	-4.02E-10	1.0000
PIB	0.164923	0.348712	0.472948	0.6489

R-squared	**0.027200**	Mean dependent var	-1.00E-10
Adjusted R-squared	-0.094401	S.D. dependent var	1.000000
S.E. of regression	1.046136	Akaike info criterion	3.104940
Sum squared resid	8.755204	Schwarz criterion	3.165457
Log likelihood	-13.52470	F-statistic	0.223680
Durbin-Watson stat	0.269266	Prob(F-statistic)	0.648885

¿Cómo interpretaría los parámetros estimados de cada regresión?

Lecturas obligatorias:

- ✓ Gujarati, D. y Porter, D. (2010). Econometría. 5ta. Edición McGraw Hill. Capítulo 6.
- ✓ Wooldridge, J. (2010). Introducción a la econometría un enfoque moderno. Capítulos 6.