Unidad 3. RLS: Inferencia

Erika R. Badillo
erika.badilloen@unaula.edu.co

Facultad de Economía
Universidad Autónoma Latinoamericana

En este tema

- Pruebas de hipótesis
- Intervalos de confianza

Lecturas

- Wooldridge, Jeffrey (2013). *Introducción a la econometría*. 5a edición, Cengage Learning. Cap. 4
- Gujarati, D. y Porter, D. (2010). *Econometría*. 5a edición, Mc Graw Hill. Cap. 5

El modelo de RLS presenta la siguiente estructura:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

En la estimación de los parámetros del modelo tenemos

Parámetro	**Estimador**	**Varianza estimada**
β_0	$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \overline{X}$	$Var(\widehat{\beta}_0) = \widehat{\sigma}^2_{\widehat{\beta}_0} = \frac{\widehat{\sigma}^2_u \sum X_i^2}{n \sum x_i^2}$
β_1	$\widehat{\beta}_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum x_i yi}{\sum x_i^2}$	$Var(\widehat{\beta}_1) = \widehat{\sigma}^2_{\widehat{\beta}_1} = \frac{\widehat{\sigma}^2_u}{\sum x_i^2}$
σ^2_u	$\widehat{\sigma}^2_u = \frac{\sum \widehat{u}^2}{n-2}$	-

σ^2_u o σ^2 es la varianza del error y $\sqrt{\sigma^2_u} = \sigma$ es llamada la desviación estándar del error

$E(Y/X) =$? y $Var(Y/X) =$? Demuestre.

Introducción

- El método estadístico intenta decir cosas sobre los parámetros poblacionales con base en los estimadores

- En el caso del modelo de RLS, consiste en decir algo acerca de β_0 y β_1 con base en $\widehat{\beta}_0$ y $\widehat{\beta}_1$

- Lo anterior implica construir intervalos de confianza y pruebas de hipótesis para β_0 y β_1

Introducción

- Para hacer inferencia estadística clásica se requiere un estadístico con distribución muestral conocida, con sólo un parámetro desconocido, justo aquel sobre el cual se hará la inferencia
- Recordando las propiedades de $\widehat{\beta}_1$:
  - es lineal en u_i: $\widehat{\beta}_1 = \beta_1 + \sum C_i u_i$ con $C_i = \frac{x_i}{\sum x_i^2}$
  - es un estimador insesgado $E(\widehat{\beta}_1) = \beta_1$
  - es eficiente: $\hat{\sigma}^2_{\widehat{\beta}_1} = \frac{\hat{\sigma}^2_u}{\sum x_i^2}$ es la mínima posible
  - Si $u_i \sim NID(0, \sigma^2_u)$, al ser $\widehat{\beta}_1$ una combinación lineal de normales, es inmediato decir que

$$\widehat{\beta}_1 \sim NID(\beta_1, \sigma^2_{\widehat{\beta}_1})$$

- Sin embargo, con la anterior distribución no es posible hacer inferencia estadística clásica acerca de β_1 pues se desconoce $\sigma^2_{\widehat{\beta}_1}$ ($\sigma^2_{\widehat{\beta}_1} = \frac{\sigma^2_u}{\sum x_i^2}$ y se desconoce σ^2_u)
- Para resolver esto y tener distribución muestral conocida se recurre al cálculo de la distribución t-student, la cual resulta de dividir una distribución normal estándar por la raíz cuadrada del cociente de chi-cuadrado y sus grados de libertad

Introducción

- Se tiene que

$$\frac{\widehat{\beta}_1-\beta_1}{\sigma_{\widehat{\beta}_1}} \sim N(0,1)$$

- Se parte del hecho que:

$$\frac{\sum \widehat{u}_i^2}{\sigma_u^2} \sim \chi^2_{N-2} \text{ gdl}$$

- Se tiene también que un estimador insesgado de σ_u^2 es:

$$\hat{\sigma}_u^2 = \frac{\sum \widehat{u}_i^2}{N-2}$$

- Ahora, despejando $\sum \widehat{u}_i^2$, se tiene

$$\sum \widehat{u}_i^2 = \widehat{\sigma}_u^2(N-2)$$

Por tanto,

$$\frac{\widehat{\sigma}_u^2(N-2)}{\sigma_u^2} \sim \chi^2_{N-2} \text{ gdl}$$

- Lo que se pretende es construir:

$$\frac{N(0,1)}{\sqrt{\frac{\chi^2_{N-2}}{N-2}}} \sim t_{N-2}$$

- Entonces reemplazando:

$$\frac{\frac{\widehat{\beta}_1-\beta_1}{\sigma_{\widehat{\beta}_1}}}{\sqrt{\frac{\widehat{\sigma}^2_u(N-2)}{\sigma^2_u(N-2)}}} = \frac{\frac{\widehat{\beta}_1-\beta_1}{\sigma_u/\sqrt{\sum x_i^2}}}{\sqrt{\frac{\widehat{\sigma}^2_u(N-2)}{\sigma^2_u(N-2)}}} = \frac{\widehat{\beta}_1-\beta_1}{\widehat{\sigma}_u/\sqrt{\sum x_i^2}} = \frac{\widehat{\beta}_1-\beta_1}{\widehat{\sigma}_{\widehat{\beta}_1}} \sim t_{N-2} \text{ gdl}$$

  Esta distribución de probabilidad si es una distribución muestral, ya que la única incógnita es β_1, el parámetro al cual se pretende hacer la inferencia

- En conclusión se tiene un estadístico $\frac{\widehat{\beta}_1-\beta_1}{\widehat{\sigma}_{\widehat{\beta}_1}}$ con una distribución muestral conocida (la t-student de $N-2$ gdl) y que cumple el requisito de sólo tener β_1 como desconocido, $\widehat{\beta}_1$ y $\widehat{\sigma}_{\widehat{\beta}_1}$ se obtienen de la muestra

Pruebas de hipótesis

- Ahora se quiere verificar estadísticamente una afirmación como la siguiente: $\beta_1 = \beta_{1o}$, esto es, verificar la hipótesis nula (H_o): $H_o : \beta_1 = \beta_{1o}$
- En estadística las hipótesis se rechazan o no se rechazan
- Lo importante en la inferencia estadística es:
  - suponer que H_o es cierta
  - encontrar la distribución muestral bajo H_o
  - observar la realidad bajo el supuesto de H_o cierta
  - si lo observado es poco probable $\Longrightarrow$ rechazar H_o
    si lo observado es probable $\Longrightarrow$ no rechazar H_o
- La decisión se toma con base en el valor del estadístico de prueba obtenido con los datos disponibles
- En consecuencia, las hipótesis nulas (H_o) que se verifican son del tipo "igualdad a", ya que es bajo este supuesto que se dan las distribuciones muestrales conocidas
- Cuando se esta bajo hipótesis nulas del tipo $>$, $<$ o $\neq$ se tienen otras distribuciones

Pruebas de hipótesis

Tenemos que el método estadístico de toma de decisiones implica:

- Formular una hipótesis nula (en términos de igualdad) y una hipótesis alternativa

$$\begin{aligned} H_o &: \beta_1 = \beta_{1o} \\ H_A &: \beta_1 < \beta_{1o} \text{ ó} \\ &\quad \beta_1 \neq \beta_{1o} \text{ ó} \\ &\quad \beta_1 > \beta_{1o} \end{aligned}$$

- Hay que encontrar la distribución muestral del estadístico apropiado, bajo H_o

$$\text{Bajo } H_o \ \frac{\widehat{\beta}_1 - \beta_{1o}}{\widehat{\sigma}_{\widehat{\beta}_1}} \sim t_{N-2} \text{ gdl}$$

- Dado esto se define el nivel de significancia aceptable en la prueba (α)

- No se debe olvidar que cualquier decisión que se tome se hace en condiciones de incertidumbre:

H_o **Realidad** **Decisión**	**Cierta**	**Falsa**
Rechaza	Error tipo I	Decisión correcta
No rechaza	Decisión correcta	Error tipo II

- $Prob$(Cometer error tipo I) $= \alpha \Longrightarrow$ Nivel de significancia
- $1 - Prob$(Cometer error tipo II) $\Longrightarrow$ Potencia de la prueba

Pruebas de hipótesis

La mecánica es

- Se formula el contraste o hipótesis

$$\begin{aligned} H_o &: \beta_1 = \beta_{1o} \\ H_A &: \beta_1 < \beta_{1o} \text{ ó} \\ &\quad \beta_1 \neq \beta_{1o} \text{ ó} \\ &\quad \beta_1 > \beta_{1o} \end{aligned}$$

- Bajo H_o cierto, el estadístico de prueba:

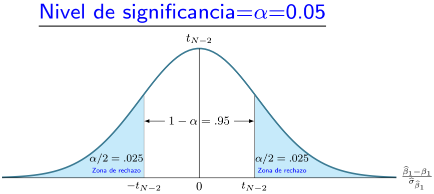
$$\frac{\widehat{\beta}_1 - \beta_{1o}}{\widehat{\sigma}_{\widehat{\beta}_1}} \sim t_{N-2} \text{ gdl}$$

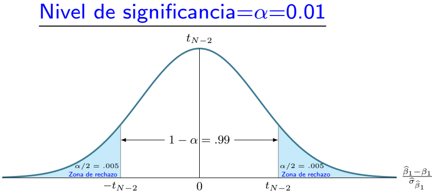
- Se establece una regla de decisión en función de H_o:
  - si $H_A : \beta_1 < \beta_{1o} \Longrightarrow t_o = \frac{\widehat{\beta}_1 - \beta_{1o}}{\widehat{\sigma}_{\widehat{\beta}_1}} < -t_{(N-2),\alpha} \Longrightarrow$ Rechazo H_o
  - si $H_A : \beta_1 \neq \beta_{1o} \Longrightarrow |t_o| = \frac{\widehat{\beta}_1 - \beta_{1o}}{\widehat{\sigma}_{\widehat{\beta}_1}} > t_{(N-2),\alpha/2} \Longrightarrow$ Rechazo H_o
  - si $H_A : \beta_1 > \beta_{1o} \Longrightarrow t_o = \frac{\widehat{\beta}_1 - \beta_{1o}}{\widehat{\sigma}_{\widehat{\beta}_1}} > t_{(N-2),\alpha} \Longrightarrow$ Rechazo H_o

- *p-value*: la probabilidad del límite derecho de H_o bajo el supuesto de que es cierta. En otras palabras, nivel de significancia más bajo al cual puede rechazarse una hipótesis nula
- La regla es rechazar H_o si

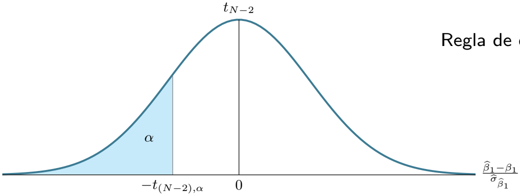
$$p\text{-value} < \alpha$$

- Gráficamente sería:





- Antiguamente se popularizaron dos números mágicos: el 5 % y el 1 %. Hoy día se precisa tener en cuenta de nuevo la naturaleza de los datos
- Se postula una regla de decisión. Lo usual ha sido hacerlo de acuerdo a la H_A
- Si es del tipo $\beta_1 < \beta_{1o}$, se trata de una cola situada a la izquierda



Regla de decisión: rechazar H_o al nivel de significancia α si $t_o < -t_{(N-2),\alpha}$ con $t_o = \frac{\widehat{\beta}_1 - \beta_{1o}}{\widehat{\sigma}_{\widehat{\beta}_1}}$

- Hoy día cada vez se generaliza el uso del *p-value*:

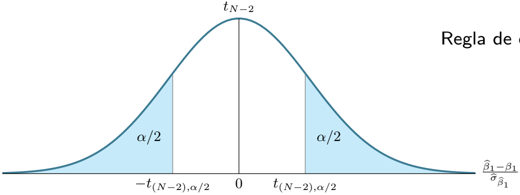
$$\textit{p-value} = \int_{-\infty}^{t_o} t_{N-2} dt$$

Esto es exactamente el nivel marginal de significancia en el cual se puede rechazar H_o. La regla de decisión sería:

Rechazar H_o si *p-value*$< \alpha$

Pruebas de hipótesis

- Si la hipótesis alternativa es del tipo $\beta_1 \neq \beta_{1o}$, se tiene un ensayo de dos colas



Regla de decisión: rechazar H_o al nivel de significancia α si $|t_o| > t_{(N-2),\alpha/2}$ con $t_o = \frac{\widehat{\beta}_1 - \beta_{1o}}{\widehat{\sigma}_{\widehat{\beta}_1}}$

- En términos del *p-value*:

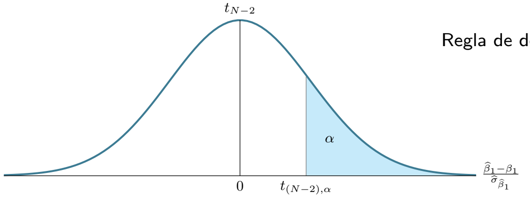
$$\textit{p-value} = 2\int_{t_o}^{\infty} t_{N-2}\,dt$$

La regla de decisión sería:

Rechazar H_o si *p-value* $< \alpha$

Pruebas de hipótesis

- Si el ensayo es del tipo $\beta_1 > \beta_{1o}$, se tiene una prueba con una cola a la derecha



Regla de decisión: rechazar H_o al nivel de significancia α si $t_o > t_{(N-2),\alpha}$ con $t_o = \frac{\widehat{\beta}_1 - \beta_{1o}}{\widehat{\sigma}_{\widehat{\beta}_1}}$

- En términos del *p-value*:

$$\textit{p-value} = \int_{t_o}^{\infty} t_{N-2} dt$$

La regla de decisión sería:

Rechazar H_o si *p-value* $< \alpha$

Ejemplo - Stata

Se tiene datos de corte transversal de 526 trabajadores correspondientes a 1976: *wage (salario en dólares por hora), educ (años de escolaridad), exper (años de experiencia laboral), female (mujer, indicador del género), y married (casado, indicador del estado marital).*

```
reg wage educ
```

Source	SS	df	MS			
				Number of obs	=	526
				F(1, 524)	=	103.36
Model	1179.73204	1	1179.73204	Prob > F	=	0.0000
Residual	5980.68225	524	11.4135158	R-squared	=	0.1648
				Adj R-squared	=	0.1632
Total	7160.41429	525	13.6388844	Root MSE	=	3.3784

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.5413593	.053248	10.17	0.000	.4367534	.6459651
_cons	-.9048516	.6849678	-1.32	0.187	-2.250472	.4407687

Se va a contrastar las siguientes pruebas de hipótesis:

$$H_o : \beta_0 = 0 \quad H_o : \beta_1 = 0$$

$$H_A : \beta_0 \neq 0 \quad H_A : \beta_1 \neq 0$$

La regla de decisión es rechazar H_o si

$$|t_o| => t_{(N-2),\alpha/2} \text{ o}$$

$$p\text{-value} < \alpha$$

```
* Calculando to
scalar to_B0 = (coef[1,2] - 0)/sqrt(var_B0)
dis to_B0
-1.3210133

scalar to_B1 = (coef[1,1] - 0)/sqrt(var_B1)
dis to_B1
10.166746

* Calculando t de la tabla, con 526-2=524 gdl y nivel
* de significancia alpha=0.05
scalar ttabla = invttail(524, .025)
dis ttabla
1.9645015

* A partir de la regla p-value<alpha
scalar pvalue_B0 = 2*ttail(524,abs(to_B0))
dis pvalue_B0
.18707349

scalar pvalue_B1 = 2*ttail(524,t0_B1)
dis pvalue_B1
2.783e-22

* Conclusión: |to_B0|<ttabla (1.3210133<1.9645015) o pvalue>alpha
* (.18707349>0.05) lo que indica que a un nivel de significancia
* de 0.05 no se rechaza Ho: B0=0. Para B1 |to_B1|>ttabla
* (10.166746>1.9645015) o pvalue<alpha ((2.783e-22)<0.05) lo que
* indica que se rechaza Ho: B1=0 a un nivel de significancia
* de 0.05
```

Pruebas de hipótesis

Ejemplo - Stata

```
reg wage educ
```

Source	SS	df	MS			
Model	1179.73204	1	1179.73204	Number of obs	=	526
Residual	5980.68225	524	11.4135158	F(1, 524)	=	103.36
				Prob > F	=	0.0000
				R-squared	=	0.1648
				Adj R-squared	=	0.1632
Total	7160.41429	525	13.6388844	Root MSE	=	3.3784

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.5413593	.053248	10.17	0.000	.4367534	.6459651
_cons	-.9048516	.6849678	-1.32	0.187	-2.250472	.4407687

Se va a contrastar la siguiente prueba de hipótesis:

$$H_o : \beta_1 = 0.6$$
$$H_A : \beta_1 \neq 0.6$$

La regla de decisión es rechazar H_o si

$$|t_0| > t_{(N-2),\alpha/2} \text{ o}$$
$$\textit{p-value} < \alpha$$

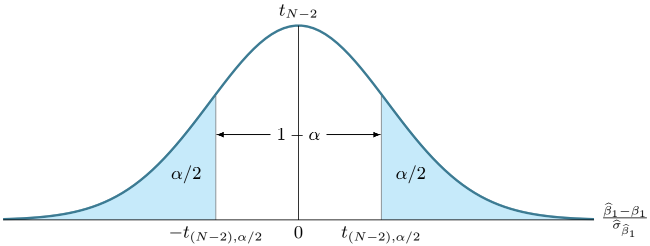
```
* Calculando to
scalar to = (coef[1,1] - 0.6)/sqrt(var_B1)
dis to
-1.101275

* Calculando t de la tabla, con 526-2=524 gdl y nivel
* de significancia alpha=0.05
scalar ttabla = invttail(524, .025)
dis ttabla
1.964502

* A partir de la regla p-value<alpha
scalar pvalue = 2*ttail(524,abs(to))
dis pvalue
0.2712825

* Conclusión: A un nivel de significancia del 5% no
* podemos rechazar la hipótesis que nuestros datos son
* compatibles con B1=0.6
```

Intervalos de confianza

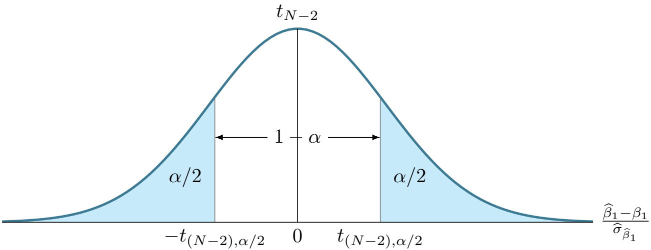
- Definición
  - Es la probabilidad de que dos valores extremos contengan el parámetro desconocido
  - Son unos límites probabilísticos que contienen al verdadero parámetro (en este caso β_1) con una probabilidad de $1-\alpha$ (nivel de confianza)
- Para construir un intervalo de confianza para β_1 se parte de la distribución muestral

$$\frac{\widehat{\beta}_1-\beta_1}{\widehat{\sigma}_{\widehat{\beta}_1}} \sim t_{N-2} \text{ gdl}$$

- Gráficamente lo que la distribución dice es:



Intervalos de confianza



Lo que esta gráfica dice es:

$$Prob\left[-t_{(N-2),\alpha/2} \leq \frac{\widehat{\beta}_1-\beta_1}{\widehat{\sigma}_{\widehat{\beta}_1}} \leq t_{(N-2),\alpha/2}\right] = 1-\alpha$$

Intervalos de confianza

Se despeja lo desconocido:

$$Prob\left[-t_{(N-2),\alpha/2} \leq \frac{\widehat{\beta}_1 - \beta_1}{\widehat{\sigma}_{\widehat{\beta}_1}} \leq t_{(N-2),\alpha/2}\right] = 1 - \alpha$$

$$Prob\left[\widehat{\beta}_1 - \widehat{\sigma}_{\widehat{\beta}_1} t_{(N-2),\alpha/2} \leq \beta_1 \leq \widehat{\beta}_1 + \widehat{\sigma}_{\widehat{\beta}_1} t_{(N-2),\alpha/2}\right] = 1 - \alpha$$

Interpretación:

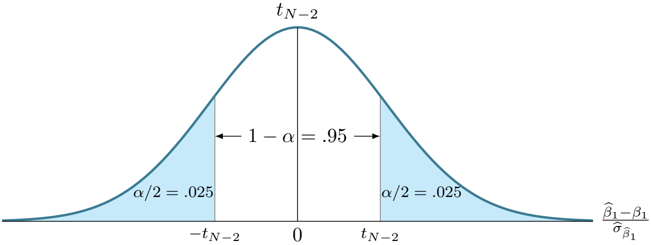
- La probabilidad de que el intervalo que va desde $\widehat{\beta}_1 - \widehat{\sigma}_{\widehat{\beta}_1} t_{(N-2),\alpha/2}$ hasta $\widehat{\beta}_1 + \widehat{\sigma}_{\widehat{\beta}_1} t_{(N-2),\alpha/2}$ contenga el verdadero valor de β_1 es $1 - \alpha$
- El intervalo de confianza $\widehat{\beta}_1 \pm \widehat{\sigma}_{\widehat{\beta}_1} t_{(N-2),\alpha/2}$ contiene a β con una probabilidad de $1 - \alpha$
- En el $(1 - \alpha)\,\%$ de los casos el intervalo contendrá el parámetro β_1

$$\begin{aligned} IC_{(1-\alpha)}(\beta_1) &= \widehat{\beta}_1 \pm \widehat{\sigma}_{\widehat{\beta}_1} t_{(N-2),\alpha/2} \\ &= \text{Estimador} \pm \text{Error de estimación x (Valor t-student)} \end{aligned}$$

Intervalos de confianza

El valor crítico

Por lo general se ha fijado un valor crítico de $\alpha = 0.05$, esto es la masa de probabilidad en las colas de la distribución, con $\alpha/2 = 0.025$ en cada cola



Intervalos de confianza

Ejemplo - Stata

```
reg wage educ
```

Source	SS	df	MS			
Model	1179.73204	1	1179.73204	Number of obs	=	526
Residual	5980.68225	524	11.4135158	F(1, 524)	=	103.36
				Prob > F	=	0.0000
				R-squared	=	0.1648
				Adj R-squared	=	0.1632
Total	7160.41429	525	13.6388844	Root MSE	=	3.3784

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.5413593	.053248	10.17	0.000	.4367534	.6459651
_cons	-.9048516	.6849678	-1.32	0.187	-2.250472	.4407687

```
* Generando el intervalo de confianza para B0, inferior
scalar IC_inf_B0 = coef[1,2] - sqrt(var_B0)*invttail(524, .025)
dis IC_inf_B0
* -2.250472
scalar IC_sup_B0 = coef[1,2] + sqrt(var_B0)*invttail(524, .025)
dis IC_sup_B0
* 0.4407687
* Generando el intervalo de confianza para B1, inferior
scalar IC_inf_B1 = coef[1,1] - sqrt(var_B1)*invttail(524, .025)
dis IC_inf_B1
* 0.4367534
* Generando el intervalo de confianza para B1, superior
scalar IC_sup_B1 = coef[1,1] + sqrt(var_B1)*invttail(524, .025)
dis IC_sup_B1
* 0.6459651
```

La estrecha relación entre los intervalos de confianza y las pruebas de hipótesis

Note que se rechaza H_o cuando

$$-t_{(N-2),\alpha/2} < t_0 < t_{(N-2),\alpha/2}$$

o cuando

$$-t_{(N-2),\alpha/2} < \frac{\widehat{\beta}_1 - \beta_1}{\widehat{\sigma}_{\widehat{\beta}_1}} < t_{(N-2),\alpha/2}$$

o cuando

$$\widehat{\beta}_1 - \widehat{\sigma}_{\widehat{\beta}_1} t_{(N-2),\alpha/2} < \beta_1 < \widehat{\beta}_1 + \widehat{\sigma}_{\widehat{\beta}_1\,(N-2),\alpha/2}$$

- Los valores extremos de este intervalo son idénticos a los valores extremos del intervalo de confianza que se dedujo en la diapositiva 20
- Se rechaza la hipótesis nula $H_o : \beta_1 = \beta_{1o}$ a un nivel de significancia α, cuando β_{1o} cae por fuera del correspondiente $100(1-\alpha)\,\%$ intervalo de confianza
- En nuestro ejemplo wage-educ, $\beta_{1o} = 0$ no cae dentro del intervalo de confianza del 95 % (0.436, 0.645), y por tanto, usando este enfoque, nosotros de nuevo rechazamos $H_o : \beta_1 = 0$ a un nivel de significancia del 5 %
- De hecho, se rechaza cualquier hipótesis nula donde β_{1o} no esta contenido en el intervalo de confianza (0.436, 0.645)
- Por otro lado, si el IC(95 %) contiene el cero, β_1 no es significativo al 5 %