Econometría I

Test de Parámetros de MCO

Carlos A. Yanes | Departamento de Economía | 2024-03-17

😕

Preguntas de la sesión anterior?

Preliminar

La última vez:

1. Trabajamos la primera estimación de modelos (MCO).
2. Miramos las demostraciones del **M.C.O**.
3. Analizamos los parámetros del **Modelo**.
4. Hablé rápidamente lo de **Normalidad**.
5. Hoy, inferencia y prueba de hipótesis de forma teórica.

Normalidad

Estadístico JB 🆒

> Es un test creado por Jarque-Bera para mirar los picos de una distribución y concluir si esta lo hace de forma normal. En este caso, la inferencia se hace hacia los errores del modelo (ϵ_i)

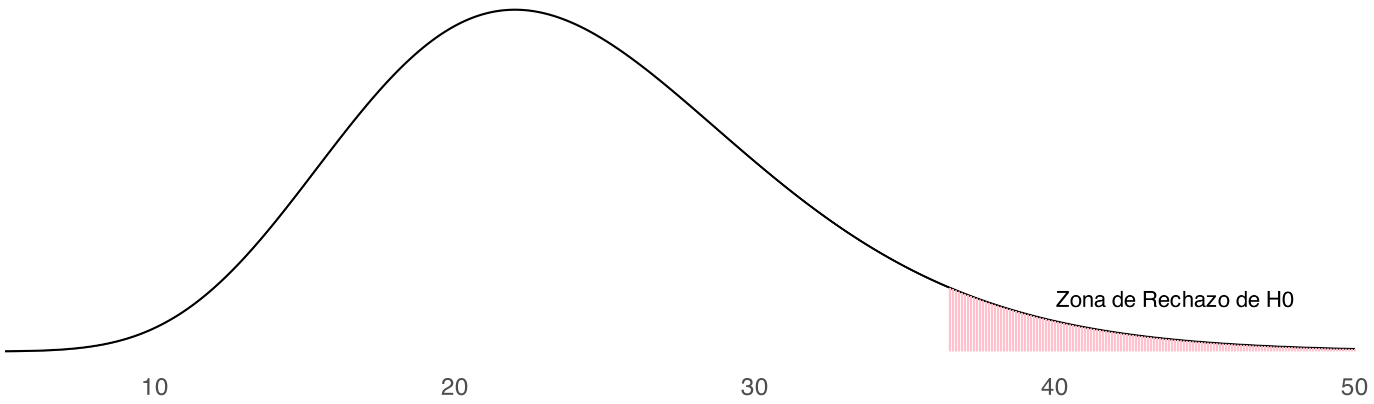
Debe ser menor al χ^2 *critico* con el objeto de constatar que:

$$H_0 = \epsilon_i \sim N(0, \sigma^2)$$
$$H_a = \epsilon_i \nsim N(0, \sigma^2)$$

- En caso de que ϵ_i no se distribuya `normal`, los $\beta's$ dejan de ser **eficientes** o *mínima varianza*, y no puede hacerse en principio inferencia estadística.
- Los intervalos de confianza no son **validos**.

Estadístico JB 🆒

Mire lo siguiente 🛑



Si por algún motivo nuestro estadístico **Jarque-Bera** cae en la zona de **rechazo** entonces podemos argumentar que nuestros residuos **NO** se distribuyen de forma normal.

Estadístico JB 🆒

Simulemos un modelo y extraigamos su residuo

```
regresion <- lm(y ~ x, datos)
u.hat <- resid(regresion)
```

Algo como:

```
#>              y     y.hat      u.hat
#> 1  -13.0974752 -3.5310510  -9.566424
#> 2  -11.6682696  0.9705791 -12.638849
#> 3  -17.0207068 -4.6934148 -12.327292
#> 4   -5.4526798  8.8149097 -14.267590
#> 5   -4.1929814  1.7811335  -5.974115
#> 6  -10.8124406 -4.6091709  -6.203270
#> 7   -0.9483357  2.6586865  -3.607022
#> 8    9.7244035  4.0528889   5.671515
#> 9    0.1861900  3.1496515  -2.963461
#> 10  13.4321802 -1.7469221  15.179102
```

Estadístico JB 🆒

$$JB = \left[\frac{s^2}{6} \times \frac{(k-3)^2}{24}\right] \sim \chi^2$$

Donde (s) es el tercer momento (Asimetría) y (k) viene siendo la curtosis. La prueba por naturaleza se hace con la tabla `ji-cuadrado`

```
library(moments) # Paquete estadístico
jarque.test(u.hat)
```

```
#> 
#> 	Jarque-Bera Normality Test
#> 
#> data:  u.hat
#> JB = 7.2082, p-value = 0.02721
#> alternative hypothesis: greater
```

Observe que la probabilidad de caer en la **zona de no rechazo** es tan solo de un 3.5%, -demasiado pequeña-. Necesitamos que $JB < \chi_{Critico}$. Para el caso, el estadístico nos da $JB = 7.20 > 3.84$

Estadístico JB 🆒

Observe que el valor crítico cambia de acuerdo al **nivel de significancia** de la prueba. Empecemos con el mas noble (10%)

```
qchisq(0.10,1, lower.tail = F)
```

```
#> [1] 2.705543
```

Si somos estrictos, entonces al 95% vamos a tener lo siguiente

```
qchisq(0.05,1, lower.tail = F)
```

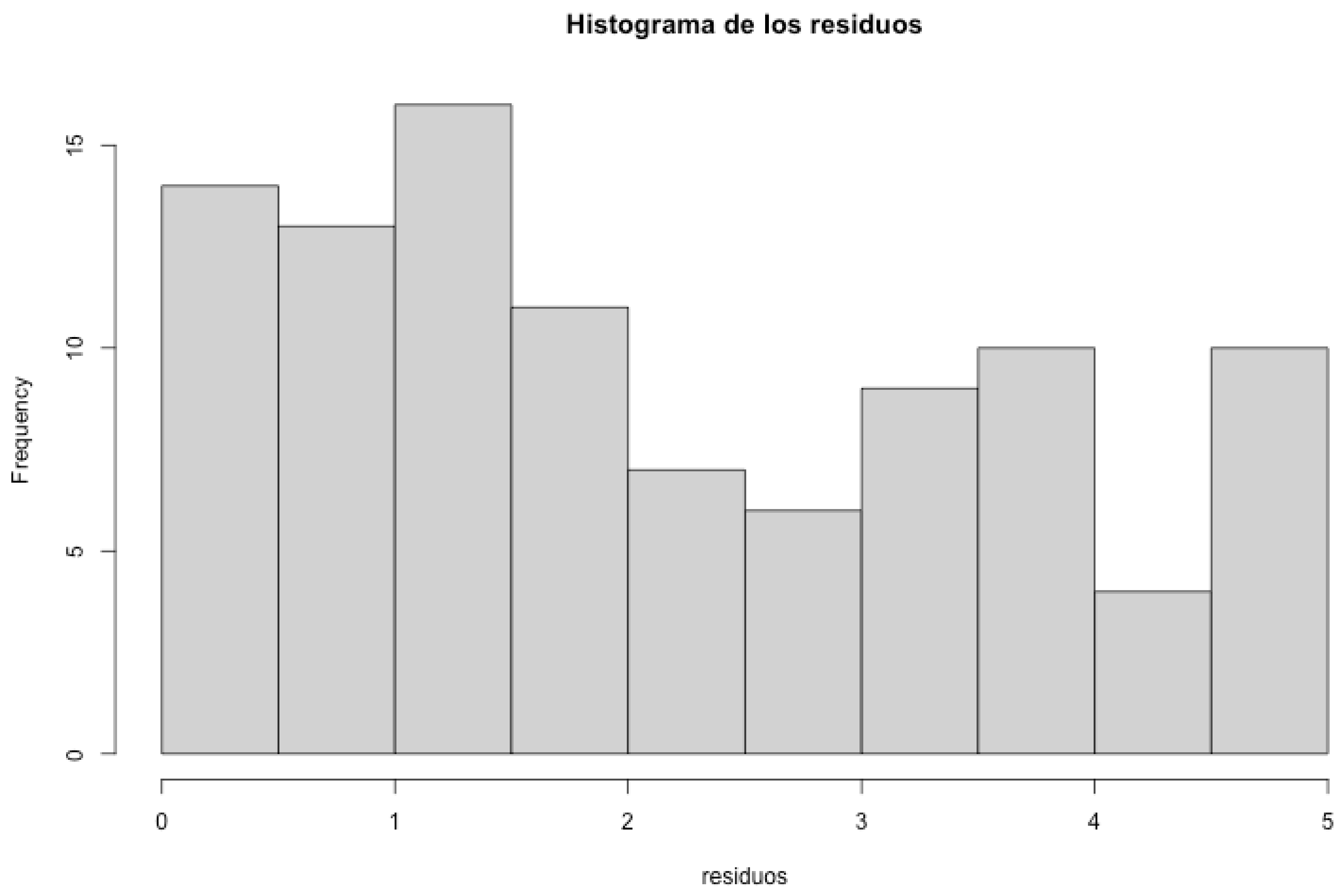
```
#> [1] 3.841459
```

Ahora miremos al nivel de testeo mas alto, esto es 99%

```
qchisq(0.01,1, lower.tail = F)
```

```
#> [1] 6.634897
```

Estadístico JB 🆒



Que sigue a continuación...

Errores o residuos del Modelo

De los errores y/o residuos ✅

Como vimos la última vez, nuestro problema que va con la **incertidumbre** es que no sabemos si nuestra estimación muestral está *cerca* o *lejos* del parámetro poblacional desconocido.

Sin embargo, no "todo está perdido". Podemos utilizar los errores $(e_i = y_i - \hat{y}_i)$ para tener una idea de lo **bien** que nuestro **modelo** explica la variación observada en y.

Cuando nuestro **modelo** parece estar haciendo un "buen" trabajo, podemos tener un poco más de confianza en su uso para aprender acerca de la relación entre y y x.

Ahora sólo tenemos que **formalizar** lo que significa realmente un "buen trabajo".

De los errores y/o residuos ✅

En primer lugar, estimaremos la varianza de u_i (recordemos: $\text{Var}(u_i) = \sigma^2$) utilizando nuestros errores al cuadrado (SEC):

$$SEC = \sigma^2 = \frac{\sum_i e_i^2}{n-k}$$

Donde k da el número de términos de pendiente e interceptos que estimamos ($p.e$, β_0 y β_1 darían $k = 2$).

σ^2 es un estimador **insesgado**.

A continuación, se muestra que la varianza de $\hat{\beta}_1$ (para la regresión lineal simple) es:

$$\text{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_i \left(x_i - \overline{x}\right)^2}$$

Que muestra que la **varianza** de nuestro estimador de la pendiente:

- Aumenta a medida que nuestras perturbaciones se vuelven más ruidosas
- Disminuye a medida que la varianza de x aumenta

Lo que sigue a continuación...

es intentarlo hacer mas interesante

De los errores y/o residuos

```
tidy(lm(Salario ~ Experiencia))
```

```
#> # A tibble: 2 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)  956.         37.4    25.5   1.53e-109
#> 2 Experiencia    0.202       3.03    0.0669 9.47e-   1
```

☐ Utilizamos el error estándar de $\hat{\beta}_1$, junto con $\hat{\beta}_1$ mismo, para aprender sobre el parámetro β_1.

Después de derivar la distribución de $\hat{\beta}_1$, tenemos dos opciones (relacionadas) para hacer inferencia estadística formal y entonces (aprender) sobre nuestro parámetro desconocido β_1:

- **Intervalos de confianza:** Utiliza la estimación y su error estándar para crear un intervalo que, al repetirse, generalmente contiene el verdadero parámetro.

- **Pruebas de hipótesis:** Determinan si hay evidencia estadística lo suficiente significativa para rechazar un valor o rango de valores de la hipótesis nula.

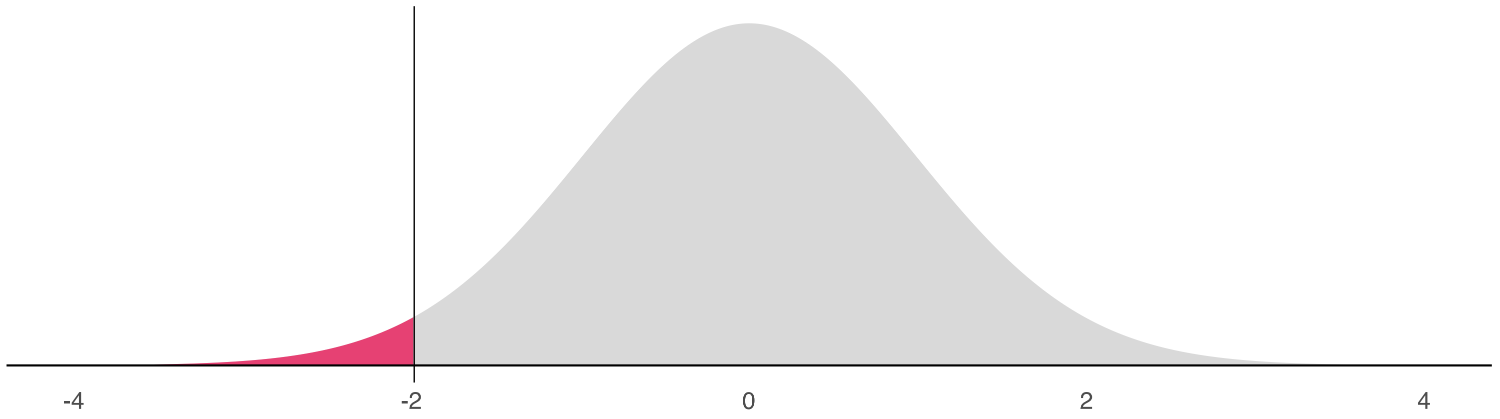
Intervalos de confianza

Intervalos de confianza

- Construimos intervalos de confianza a un nivel de $(1-\alpha)$ para β_1:

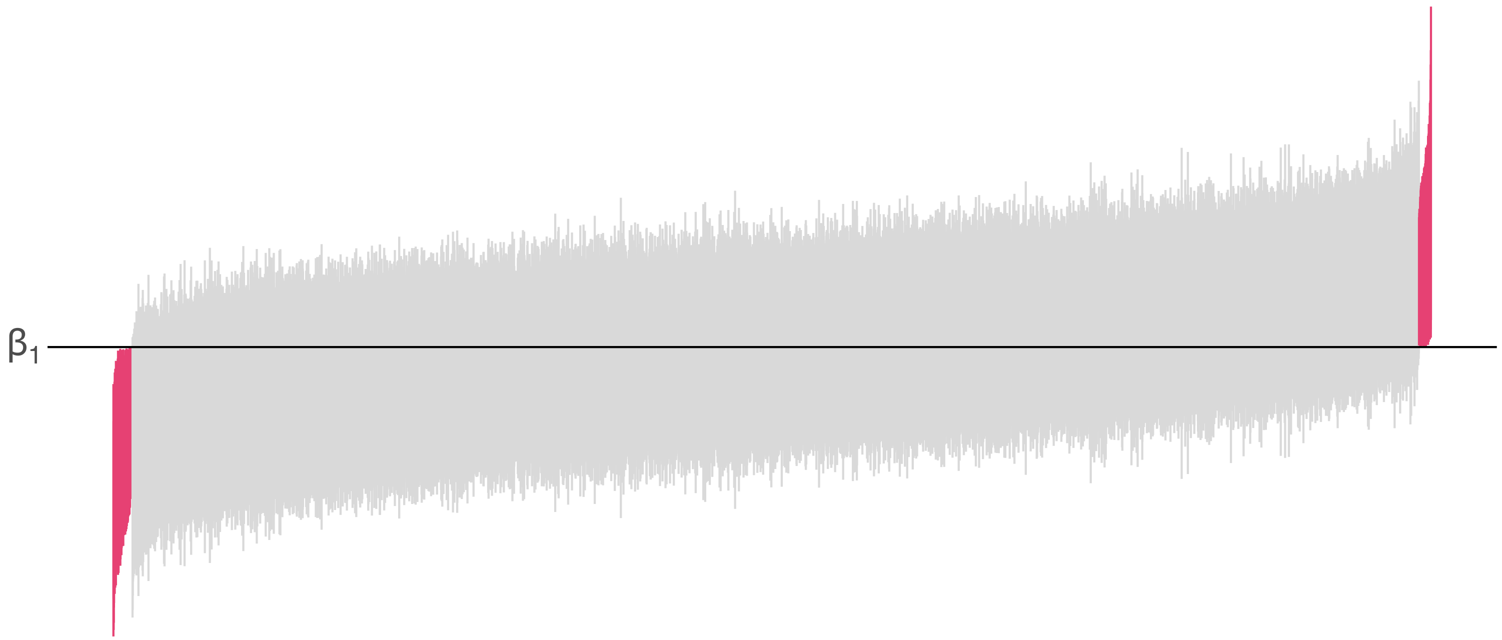
$$\hat{\beta}_1 \pm t_{\alpha/2,\ \text{df}}\ \hat{\text{SE}}\left(\hat{\beta}_1\right)$$

- Por ejemplo, con 100 obs., tenemos dos coeficientes, $(\hat{\beta}_0$ y $\hat{\beta}_1 \implies k=2),$ y tenemos un $\alpha = 0.025$ (Para un intervalo de confianza del 98%) nos brinda un **estadístico** de $t_{0.025,\ 98}$ = -1.98



Intervalos de confianza

Del lo anterior Tenemos certeza que con un 97.6% de confiabilidad nuestros intervalos de confianza contienen el verdadero valor de nuestro β_1.



Intervalos de confianza 🎙

- Construimos intervalos de confianza a un nivel de $(1-\alpha)$ para β_1:

$$\hat{\beta}_1 \pm t_{\alpha/2,\ \mathrm{df}}\ \hat{\mathrm{SE}}\left(\hat{\beta}_1\right)$$

Ejemplo:

```
lm(Salario ~ Experiencia) %>% tidy()
```

```
#> # A tibble: 2 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)  956.         37.4    25.5   1.53e-109
#> 2 Experiencia    0.202       3.03    0.0669 9.47e-  1
```

- Nuestro intervalo de confianza del 98% es en nuestro caso $0.202 \pm 1.98 \times 3.03 = [-5.7364,\ 6.1412]$

Recuerde que el valor crítico puede obtenerlo de:

```
qt(0.975,100)
```

```
#> [1] 1.983972
```

Intervalos de confianza

- Construimos intervalos de confianza a un nivel de $(1-\alpha)$ para β_1:

$$\hat{\beta}_1 \pm t_{\alpha/2,\ \mathrm{df}}\ \hat{\mathrm{SE}}\left(\hat{\beta}_1\right)$$

Directamente en **R**:

```
modelo.1<- lm(Salario~Experiencia)
confint(modelo.1)
```

```
#>                   2.5 %      97.5 %
#> (Intercept) 882.185279 1029.024598
#> Experiencia  -5.736443    6.141249
```

⚘ *si esta interesado(a)* en mirar los otros niveles de confianza es usar el código con la opción **level** *p.e*:

```
confint(modelo.1, level=0.99)
```

Intervalos de confianza

Qué significa el intervalo:

$$\left[-5.7364 \leq \hat{\beta}_i \leq 6.1412\right]$$

Informalmente: El intervalo de confianza nos da una región (intervalo) en la que podemos depositar cierta confianza para contener el parámetro estimado.

Más formalmente: Si con nuestras muestras de la población repetimos el proceso n veces y construimos intervalos de confianza para cada una de estas, $(1-\alpha)$ por ciento de nuestros intervalos (*p.e*, 97.5%) contendrán el parámetro poblacional *en algún lugar del intervalo*.

Hipótesis 👍

> **Pruebas de hipótesis**: En muchas aplicaciones, queremos saber algo más que una estimación puntual o un rango de valores. Queremos saber qué dicen nuestras pruebas estadísticas sobre las **teorías** existentes.

- Queremos comprobar las hipótesis planteadas por funcionarios, políticos, economistas, científicos, amigos, vecinos raros, etc.

Ejemplos:

☆ ¿El aumento de la presencia policial **reduce la delincuencia**?

☆ ¿Construir un muro gigante **reduce la delincuencia**?

☆ ¿Influye el cierre de un gobierno **en la economía**?

☆ ¿Se reduce con la **legalización del uso** de cannabis el numero de casos de conducir (manejar un vehículo) bajo los efectos del alcohol o con consumo de las opioides?

☆ ¿Las normas de calidad del aire **aumentan la salud** y/o **reducen el empleo**?

Hipótesis 👍

Las pruebas de hipótesis se basan en resultados e intuiciones muy similares.

Aunque no cabe duda de que existe **incertidumbre**, podemos elaborar pruebas estadísticas fiables (rechazar o no rechazar una hipótesis planteada).

En **MCO** las pruebas de hipotesis se hacen a los parámetros:

β_1 es igual al valor (c), *p.e.*, planteamos que $H_o: \ \beta_1 = c$

Luego esta el *test* para hacerlo:

$$t_{\text{estadístico}} = \frac{\hat{\beta}_1 - c}{\hat{\text{SE}}\left(\hat{\beta}_1\right)}$$

> Note que c regularmente se iguala a cero (0) y la hipotesis nula pasa a ser $H_o: \ \beta_1 = 0$.

Hipótesis

Para un nivel α y un test de **dos colas**, vamos a rechazar la *hipotesis nula* cuando ocurra lo siguiente:

$$\left|t_{\text{estadístico}}\right| > \left|t_{1-\alpha/2,\ df}\right|$$

Lo que significa que nuestro **estadístico de prueba es más extremo que el valor crítico**.

De otra forma, podemos calcular el **valor p** que acompaña a nuestro estadístico de prueba, que nos da efectivamente la probabilidad de ver nuestro estadístico de prueba *o un estadístico de prueba más extremo* si la hipótesis nula fuera cierta.

Los **valores p** muy pequeños (generalmente < 0,05) *significan* que sería poco probable ver nuestros resultados si la hipótesis nula fuera realmente cierta; tendemos a rechazar la hipótesis nula para valores p inferiores a 0,05.

En **R**:

```
library(broom) # Para tener el p-value
modelo.1<- lm(Salario~Experiencia)
glance(modelo.1)$p.value
```

```
#>     value
#> 0.9466878
```

Hipótesis 👍

```
lm(Salario~Experiencia) %>% tidy()
```

```
#> # A tibble: 2 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)  956.         37.4    25.5    1.53e-109
#> 2 Experiencia    0.202       3.03    0.0669 9.47e-  1
```

Ho: $\beta_1 = 0$ *vs.* Ha: $\beta_1 \neq 0$

$t_{\text{estadístico}} = 0.0669$ y el $t_{0.95,\,28} = 1.65$

El cual implica que p-value > 0.05

Entonces, no podemos **rechazar Ho**.

Hipótesis

Cómo se lee el siguiente modelo desde la probabilidad?

Modelo de Prueba					
	Estimate	Standard Error	t value	Pr(>ltl)	
(Intercept)	861.482	78.171	11.020	0.0000	***
Experiencia	18.837	13.924	1.353	0.1764	
I(Experiencia^2)	-0.794	0.579	-1.371	0.1707	

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05*

Residual standard error: 404.4 on 932 degrees of freedom

Multiple R-squared: 0.002018, Adjusted R-squared: -0.0001238

F-statistic: 0.9422 on 932 and 2 DF, p-value: 0.3901

Hipótesis 👍

» El p-value o **p-valor** nos dice que probabilildad tenemos de caer en la zona de **no rechazo** (la zona mas grande de toda la distribución).

> Cientificamente, esto implica la probabilidad que tenemos de cometer el error tipo I en las pruebas de hipótesis. *Esto es, usted* rechaza *Ho cuando ella es verdadera*

La formula de cálculo es:

$$\text{p-value} = 2 \times P(T_{n-1} > |t|) \equiv 2 \times (1 - Ft_{n-1}(|t|))$$

*Donde $|t|$ es el **valor crítico** y Ft es la función de densidad*

en **R**:

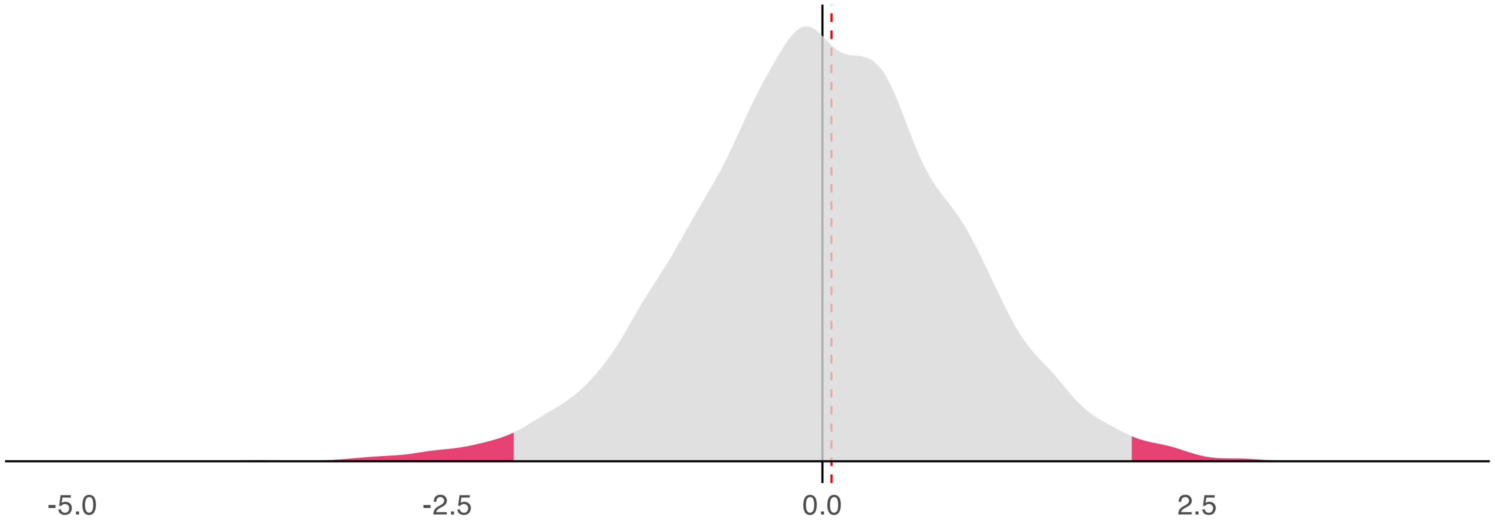
```
n<- 935 # Por el tamaño muestral del ejemplo de salarios/educación
t<- 0.0669 # Valor de T-calculado
(p<-2*(1-pt(abs(t), n-1)))
```

```
#> [1] 0.9466756
```

Hipótesis 👍

En nuestro ejemplo con los salarios, hay un 94% (por ciento) que nuestro t estadístico esté en la zona de **no rechazo** y por ende la **experiencia** *no explique las variaciones del salario*

La distribución de nuestro t estadístico es: (teniendo presente las zonas de rechazo).



Otras Métricas: R-Cuadrado

Otras Métricas

- **Suma Total de Cuadrados** (SST): Mide variación muestral total de y_i.

$$SST \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- **Suma Explicada de Cuadrados** (SSE): Mide variación de $\hat{y}_i$.

$$SSE \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

- **Suma de los Residuos al Cuadrado** (SSR): Mide variación en μ_i.

$$SSR \equiv \sum_{i=1}^{n} \hat{\mu}_i^2$$

La **variación total** en y puede ser expresada como la suma de la variación explicada y la no explicada:

$$SST = SSE + SSR$$

Otras Métricas

- **Coeficiente de determinación** R^2: Mide el grado de precisión del modelo, la proporción de la variación de la variable *dependiente* que es explicado por x.

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \qquad R^2 \in [0, 1]$$

Cuando se interpreta se multiplica por 100 para interpretarlo como porcentaje.

Un R^2 **cercano a cero** indica un ajuste bajo a la linea de M.C.O.

Un R^2 **cercano a uno**, x explica la mayoría de y.

Otras Métricas

- En **R** se puede implementar así:

```
modelo.1 <- lm(Salario ~ Experiencia)
sal.pred <- fitted(modelo.1) #Predichos
u.hat <- resid(modelo.1)

# R cuadrado puede obtenerse:
Sal <- datos$wage
var(sal.pred)/var(Sal) #Primera forma
```

```
#> [1] 4.794796e-06
```

```
1 - var(u.hat)/ var(Sal) #Segunda forma
```

```
#> [1] 4.794796e-06
```

```
cor(Sal, sal.pred)^2 # Tercera forma
```

```
#> [1] 4.794796e-06
```

Bibliografía

Álvarez, R. A. R., Calvo, J. A. P., Torrado, C. A. M., & Mondragón, J. A. U. (2013). *Fundamentos de econometría intermedia: teoría y aplicaciones*. Universidad de los Andes.

Stock, J. H., Watson, M. W., & Larrión, R. S. (2012). *Introducción a la Econometría*.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Gracias por su atención!

Alguna pregunta adicional?

Carlos Andres Yanes Guerra

cayanes@uninorte.edu.co

keynes37