Modelos de Regresión y Series de Tiempo (MRST)
2025 - 02

Clase 6 – MRLS
Inferencia sobre el modelo de regresión
Intervalos de confianza

Docente: Natalia Jaramillo Quiceno

Escuela de Ingenierías

natalia.jaramilloq@upb.edu.co

Universidad Pontificia Bolivariana

Regresión lineal simple

Recordemos las propiedades de β_1

Propiedades de $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$

$\hat{\beta}_1$ es una función lineal de variables aleatorias independientes $Y_1, Y_2, \dots, Y_n$, cada una de las cuales está normalmente distribuida

Así, se tiene que:

- $\boldsymbol{E}(\widehat{\boldsymbol{\beta}}_{\mathbf{1}}) = \beta_1$: estimador insesgado. La distribución de $\hat{\beta}_1$ siempre está centralizada en el valor β_1.

- $\mathbf{V}(\widehat{\boldsymbol{\beta}}_{\mathbf{1}}) = \frac{\sigma^2}{S_{xx}}$ y $\mathbf{se}(\widehat{\boldsymbol{\beta}}_{\mathbf{1}}) = \frac{\sigma}{\sqrt{S_{xx}}}$ (**se:** error estándar estimado)

- El estimador $\hat{\beta}_1$ tiene una distribución normal.

Regresión lineal simple

Recordemos las propiedades de β_0

Propiedades de $\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$

Para el intercepto, se puede demostrar de la misma manera que:

- $\boldsymbol{E}(\widehat{\boldsymbol{\beta}}_{\mathbf{0}}) = \beta_0$: estimador insesgado. La distribución de $\hat{\beta}_0$ siempre está centralizada en el valor β_0.
- $\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathbf{0}}) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]$ y
- $\mathbf{se}(\widehat{\boldsymbol{\beta}}_{\mathbf{0}}) = \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}$ (**se:** error estándar estimado)
- El estimador $\hat{\beta}_0$ tiene una distribución normal.

Regresión lineal simple

Intervalos de confianza para β_0 y β_1

Un intervalo de confianza de $100(1-\alpha)\%$ para la pendiente β_1 se determina con:

$$\hat{\beta}_1 - t_{\alpha/2,n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2,n-2}\underbrace{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}_{se(\hat{\boldsymbol{\beta}}_1)}$$

De manera similar, para β_0

$$\hat{\beta}_0 - t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]} \le \beta_0 \le \hat{\beta}_0 + t_{\alpha/2,n-2}\underbrace{\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]}}_{se(\hat{\boldsymbol{\beta}}_0)}$$

- El ancho de los intervalos de confianza para β_0 y β_1 es una medida de la calidad general de la recta de regresión.
- Cuando el I.C. de β_1 **no contiene el cero**, se dice que:

La variable respuesta “Y” está relacionada linealmente con la variable predictora o independiente “X”

Regresión lineal simple
Intervalos de confianza para β_1

- Un IC al $100(1-\alpha)\%$ de confianza para la pendiente β_1 es:

$$\hat{\beta}_1 - t_{\alpha/2,n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2,n-2}\underbrace{\boxed{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}}_{se(\hat{\beta}_1)}$$

- Construya in IC al 95% para la pendiente β_1 del ejemplo básico (inversiones en I+D y ganancias de la empresa).

Primero, busquemos los datos que necesitamos en el resumen (summary) del modelo:

```
Coefficients:
            Estimate  Std. Error t value Pr(>|t|)
(Intercept)  20.0000      2.6458   7.559  0.00164 **
inv           2.0000      0.4583   4.364  0.01202 *
---
```

$\hat{\beta}_1$ $\quad$ $se(\hat{\beta}_1)$

¿Y el valor de t?

Tablas, Excel o R ☺

Definir valor de α

Calcular g.l del error$(n-2)$

```
qt(0.025,4,lower.tail = FALSE)
```

$$t_{\alpha/2,n-2} = t_{0.025,4} = 2.776$$

Regresión lineal simple
Intervalos de confianza para β_1

- Un IC al $100(1-\alpha)\%$ de confianza para la pendiente β_1 es:

$$\hat{\beta}_1 - t_{\alpha/2,n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2,n-2}\underbrace{\boxed{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}}_{se(\hat{\beta}_1)}$$

- Construya in IC al 95% para la pendiente β_1 del ejemplo básico (inversiones en I+D y ganancias de la empresa).

$$\hat{\beta}_1 = 2$$

$$se(\hat{\beta}_1) = 0{,}4582$$

$$t_{\alpha/2,n-2} = t_{0.025,4} = 2.776$$

¿Cómo lo interpretamos?

Sin Límites

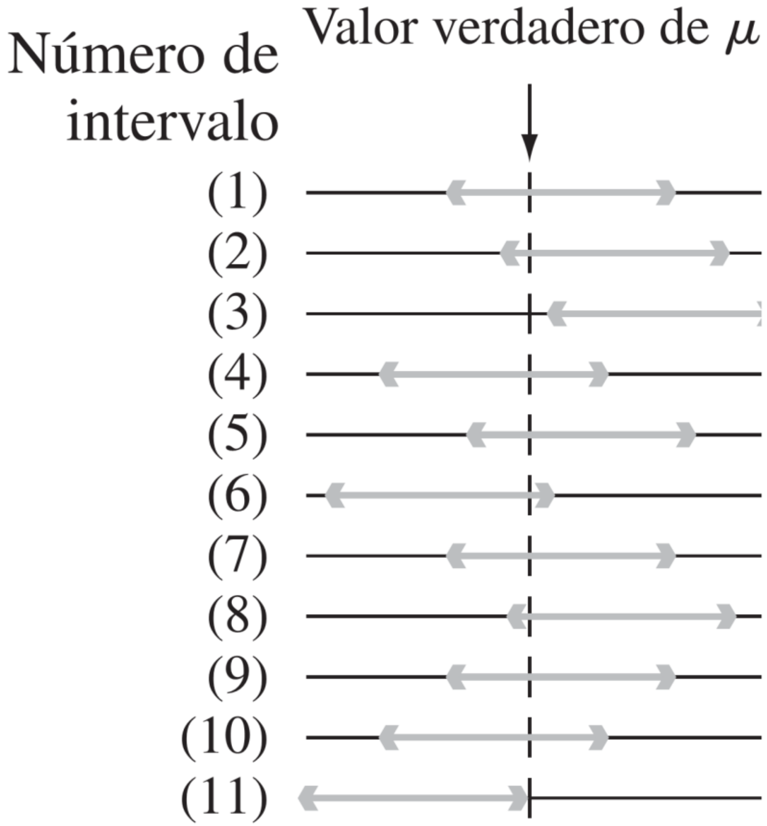
Intervalos de confianza

Interpretación de un intervalo de confianza y otros niveles de confianza

Para el caso de un 95% de confianza

Una **interpretación correcta** de 95% de confianza se basa en la interpretación de probabilidad de frecuencia relativa a largo plazo.

- Suponga que calcula varios intervalos, cada uno con una muestra diferente.
- El 95% de los intervalos calculados contendrán a μ.



Regresión lineal simple

Intervalos de confianza para β_1 en RStudio

Comandos

```
##Luego de ajustar el modelo!!!

#Activar paquete stats
library(stats)

#Generar intervalo de confianza para b0 y b1, por defecto este código lo crea al 95%
confint(modelo)
```

Resultado

```
> confint(modelo)
                2.5 %    97.5 %
(Intercept) 12.654217 27.345783
inv          0.727673  3.272327
```

Regresión lineal simple

IC para la respuesta media (μ)

Una aplicación importante de un modelo de regresión es estimar la **respuesta media** ($\boldsymbol{\mu}$), para determinado valor de la variable independiente x.

Un estimador puntual de la media de Y para un valor x_0 es

$$\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0 = \bar{y} + \hat{\beta}_1 (x_0 - \bar{x})$$

La varianza de $\hat{\mu}_{Y|x_0}$ es

$$V\left(\hat{\mu}_{Y|x_0}\right) = \sigma^2 \left[\frac{1}{n} + \frac{\boxed{(x_0 - \bar{x})^2}}{S_{xx}}\right]$$

El error de la estimación depende de la distancia entre x_0 y $\bar{x}$!!!

Así, un intervalo de confianza al $100(1-\alpha)\%$ para la respuesta media ($\boldsymbol{\mu}$), dado un valor de x_0, está dado por

$$\hat{\boldsymbol{\mu}}_{Y|x_0} \pm t_{\alpha/2,n-2} * \sqrt{\hat{\boldsymbol{\sigma}}^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]}$$

Universidad Pontificia Bolivariana

Regresión lineal simple

IC para la respuesta media

Volviendo al ejemplo de las ganancias…

Con:

$$\boldsymbol{\hat{\mu}}_{\boldsymbol{Y|x_0}} = \mathbf{20 + 2}\boldsymbol{x_0}$$

Suponga que hay interés en estimar las ganancias cuando la inversión en I+D es:

$$\boldsymbol{x_0} = \mathbf{9}$$

Construya un IC del 95% para $\boldsymbol{\mu}_{\boldsymbol{Y|9}}$

$\bar{x} =$ $\qquad n =$ $\qquad \leq \boldsymbol{\mu}_{\boldsymbol{Y|9}} \leq$

$\hat{\sigma}^2 =$ $\qquad S_{xx} =$

$\boldsymbol{t}_{0.025,4} =$

Regresión lineal simple
IC para la respuesta media

Siguiendo con el ejemplo de las ganancias…

Con:

$$\hat{\mu}_{Y|x_0} = 20 + 2x_0$$

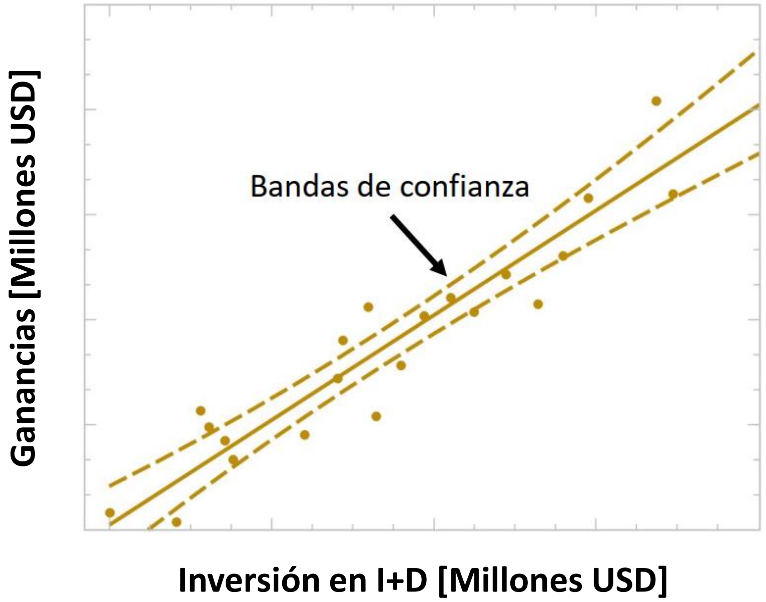
Construya intervalos para la respuesta media cuando:

$$x_0 = 2$$
$$x_0 = 11$$

$$\leq \mu_{Y|2} \leq$$

$$\leq \mu_{Y|11} \leq$$

¿el ancho de los intervalos obtenidos es el mismo?



Regresión lineal simple

IC para la respuesta media en RStudio

Comandos

```
#Activar paquete stats
library(stats)

#Crear vector, con el valor de x con el que se desea estimar Y
new.dat <- data.frame(inv=9)

#Generar intervalo de confianza para la respuesta media, por defecto este código lo crea al 95%
predict(modelo, newdata = new.dat, interval = 'confidence')
```

Resultado

```
> predict(modelo, newdata = new.dat, interval = 'confidence')
  fit      lwr      upr
1  38 31.72376 44.27624
```

MUCHAS GRACIAS

Natalia Jaramillo Quiceno
e-mail: natalia.jaramilloq@upb.edu.co