Análisis de regresión

26 de agosto
(semana 3)

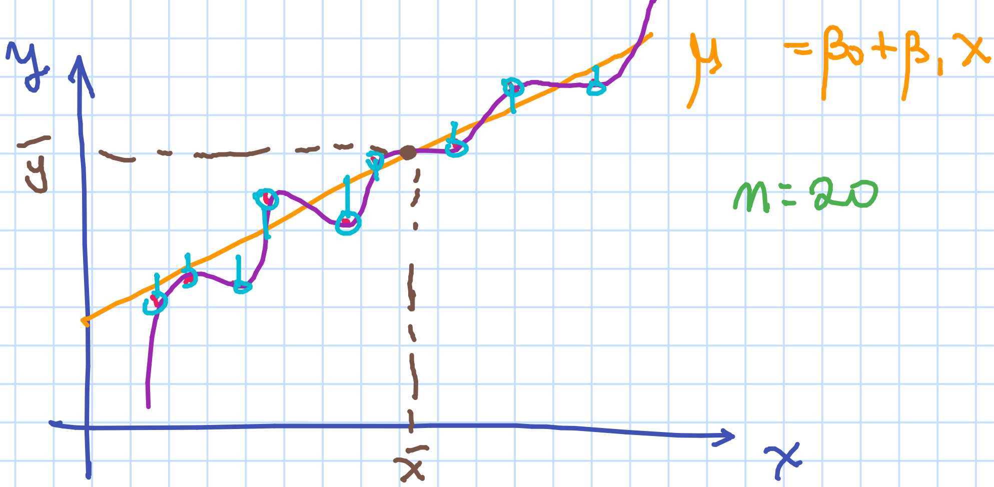
Plan de trabajo

1. Inferencia sobre los coeficientes de regresión
2. Predicción en la regresión lineal simple

Nota:

- No olviden aceptar la invitación al CLASSROOM y revisar acceso al DRIVE y MOODLE (https://micampus.unal.edu.co/)
- La primera entrega del trabajo final será para el domingo 4 de septiembre por CLASSROOM.
    - Llenen la lista que les compartí con los integrantes.
    - Si van a recoger sus datos, NO empiecen hasta que yo les dé el aval.
    - En el DRIVE, videos de repaso de estadística descriptiva.
- En la semana del 29 de agosto al 2 de septiembre no habrá clase. Usen ese tiempo para la 1era entrega y el taller 1.
- Impriman por favor material módulo 2 **para la próxima clase**.

Regresión lineal simple



$$\begin{cases} Y_k = \mu_k + \varepsilon_k \\ \mu_k = \beta_0 + \beta_1 x_k \\ \varepsilon_k \ \text{iid} \end{cases}$$

$Var(\varepsilon_k) = \sigma^2 \quad \forall k$

$E(\varepsilon_k) = 0 \quad \forall k$

Estimación de los coeficientes de regresión

- **Parámetros:** β_0 : intercepto, β_1 : pendiente

Método de mínimos cuadrados ordinarios: Suponga que se tiene el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i; \quad 1 \le i \le n$$

Donde:

$$1.\ E(\varepsilon_i) = 0\ \forall i$$

$$2.\ Var(\varepsilon_i) = \sigma^2\ \forall i$$

$$3.\ Cov(\varepsilon_i, \varepsilon_j) = 0\ \forall i \ne j$$

La estimación de mínimos cuadrados (dados los datos) es:

$$\left(\hat{\beta}_0, \hat{\beta}_1\right) = \arg\min_{\beta_0, \beta_1} Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right)^2$$

Solución del problema de minimización (I)

1. Puntos críticos.

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_0}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \ (1) \\ \dfrac{\partial Q}{\partial \beta_1}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -2\sum_{i=1}^{n} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \ (2) \end{cases}$$

$$\begin{cases} \overline{y} - \hat{\beta}_0 - \hat{\beta}_1 \overline{x} = 0 \therefore \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \ (1') \\ \sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \ (2') \end{cases}$$

Reemplazando $\hat{\beta}_0$ de (1') en (2') y despejando, se tiene que

$$\sum_{i=1}^{n} x_i y_i - \left(\overline{y} - \hat{\beta}_1 \overline{x}\right)\sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y} - n\hat{\beta}_1 \overline{x}^2 - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \therefore \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right) y_i}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}$$

Solución del problema de minimización (II)

2. Mínimo local (y global).

$$\begin{cases} \dfrac{\partial^2 Q}{\partial \beta_0^2}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -2\sum_{i=1}^{n}(-1) = 2n \\ \dfrac{\partial^2 Q}{\partial \beta_1^2}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -2\sum_{i=1}^{n} x_i\left(-x_i\right) = 2\sum_{i=1}^{n} x_i^2 \\ \dfrac{\partial^2 Q}{\partial \beta_0 \partial \beta_1}\left(\hat{\beta}_0, \hat{\beta}_1\right) = 2\sum_{i=1}^{n} x_i \end{cases}$$

$$\mathbf{J} = \begin{pmatrix} 2n & 2\sum_{i=1}^{n} x_i \\ 2\sum_{i=1}^{n} x_i & 2\sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

Luego, por el criterio de Sylvester, como

$$2n > 0 \text{ y } \det\left(\mathbf{J}\right) = 4n\sum_{i=1}^{n} x_i^2 - 4n^2\overline{x}^2 = 4n\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 > 0;$$

por ende, $\mathbf{J}$ es positiva definida, lo que implica que la solución es un mínimo local.

Como $\mathbf{J}$ no depende de $\hat{\beta}_0, \hat{\beta}_1$; quiere decir que Q es convexa y eso hace que la solución sea global.

No es difícil ver que $\lim_{\|\boldsymbol{\beta}\| \to \infty} Q\left(\beta_0, \beta_1\right) = \infty$; lo que completa la prueba.

Una definición importante y un lema

Mejor estimador linealmente insesgado (BLUE)

Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias involucradas en un proceso de estimación. Se dice que un estimador de la forma $\sum_{i=1}^{n} \phi_i Y_i$ con constantes ϕ_i no aleatorias y conocidas es un **BLUE** si

1) es un estimador insesgado,

2) es el estimador con menor varianza dentro de todos los estimadores lineales insesgados.

- Como siempre la definición no es constructiva. Usaremos el lema del máximo (Lema 11.2.7 de Casella & Berger) más adelante para verificar que un estimador es BLUE.

Best (mejor) $Var(\hat{\theta})$ es la más pequeña

Linear (lineal) $\hat{\theta} = \sum_{i=1}^{n} \phi_i Y_i$

Unbiased (insesgado) $E[\hat{\theta}] = \theta$

Estimator

$\hat{\theta}$ es un BUE de θ

Coeficientes de regresión (I)

Teorema de Gauss-Markov

La solución al problema de minimización propuesto es:

$$
\begin{cases}
\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x}_n) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}, \\
\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{x}_n
\end{cases}
$$

$\hat{\beta}_1 \xrightarrow[n\to\infty]{P} \beta_1$

$\hat{\beta}_0 \xrightarrow[n\to\infty]{P} \beta_0$

BLUE

bajo las condiciones del método descrito anteriormente, corresponde a los mejores estimadores linealmente insesgados de los parámetros del modelo de regresión lineal simple.

Además, son consistentes y tienen distribuciones asintóticas normales (por el TCL de Hájek - Sidak).

Idea de la prueba (I)

Es claro que los estimadores encontrados son lineales.

$$\begin{cases} \hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x}_n)Y_i}{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2} = \sum_{i=1}^{n} c_i Y_i, \text{ con } c_i = \dfrac{(x_i - \overline{x}_n)}{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2} \quad . \\ \hat{\beta}_0 = \overline{Y}_n - \hat{\beta}_1 \overline{x}_n = \dfrac{1}{n}\sum_{i=1}^{n} Y_i - \overline{x}_n \sum_{i=1}^{n} c_i Y_i = \sum_{i=1}^{n}\left(\dfrac{1}{n} - \overline{x}_n c_i\right)Y_i \end{cases}$$

Linealidad ✓

Ahora, vamos a probar que son insesgados:

$$E\left[\hat{\beta}_1\right] = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_n)E[Y_i]}{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_n)(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2} = \frac{\beta_0 \sum_{i=1}^{n}(x_i - \overline{x}_n) + \beta_1 \sum_{i=1}^{n} x_i (x_i - \overline{x}_n)}{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2} = \beta_1.$$

$$E\left[\hat{\beta}_0\right] = E\left[\overline{Y}_n\right] - E\left[\hat{\beta}_1\right]\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i) - \beta_1 \overline{x}_n = \frac{1}{n} n\beta_0 + \beta_1 \overline{x}_n - \beta_1 \overline{x}_n = \beta_0.$$

Insesgamiento ✓

Idea de la prueba (II)

Centremos la atención en el proceso de estimación de β_1.

Todo estimador lineal $\tilde{\beta}_1$ debe ser de la forma $\tilde{\beta}_1 = \sum_{i=1}^{n} \phi_i Y_i$.

Para ser insesgado, debe satisfacer que $E\left[\tilde{\beta}_1\right] = \sum_{i=1}^{n} \phi_i E\left[Y_i\right] = \beta_1$.

De este modo, debe satisfacer que $\sum_{i=1}^{n} \phi_i \left(\beta_0 + \beta_1 x_i\right) = \beta_0 \sum_{i=1}^{n} \phi_i + \beta_1 \sum_{i=1}^{n} \phi_i x_i = \beta_1$.

De allí se obtienen dos restricciones para los coeficientes:

$(1) \sum_{i=1}^{n} \phi_i = 0, \ (2) \sum_{i=1}^{n} \phi_i x_i = 1.$

independencia

Ahora, $Var\left(\tilde{\beta}_1\right) = Var\left(\sum_{i=1}^{n} \phi_i Y_i\right) = \sum_{i=1}^{n} \phi_i^2 Var\left(Y_i\right) = \sigma^2 \sum_{i=1}^{n} \phi_i^2$.

Por ende, un BLUE se puede encontrar solucionando el problema de minimizar

$\sum_{i=1}^{n} \phi_i^2$ sujeto a las restricciones (1) y (2). El lema del máximo garantiza que dicha solución es la que se obtiene con el estimador de mínimos cuadrados.

Todo estimador linealmente insesgado de β_1

$$\tilde{\beta}_1 = \sum_{i=1}^{n} \phi_i Y_i$$

① $Var(\tilde{\beta}_1) = \sigma^2 \sum_{i=1}^{n} \phi_i^2$

② Para que sea insesgado $\Longrightarrow$

(1) $\sum \phi_i = 0$

$\sum \phi_i x_i = 1$

$$\boxed{\begin{array}{l} \min \ \sigma^2 \sum_{i=1}^{n} \phi_i^2 \\ \text{s.a} \\ \quad \sum \phi_i = 0 \\ \quad \sum \phi_i x_i = 1 \end{array}}$$

TAREA: Resolver el problema de minimización y ver que $\phi_i = c_i$ (de $\hat{\beta}_1$)

Coeficientes de regresión (II)

Teorema de Gauss-Markov (II)

Si además 1 se tiene que $\{\varepsilon_i\}$ son una m.a. $N\left(0,\sigma^2\right)$, $\hat{\beta}_0$ y $\hat{\beta}_1$ son UMVUEs de sus respectivos parámetros y son los mismos estimadores ML. 2 Además, tienen distribución normal bivariada con

$$\left(\hat{\beta}_0,\hat{\beta}_1\right)^T \sim MVN_2\left(\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix},\begin{bmatrix}\dfrac{\sigma^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}_n\right)^2} & \dfrac{-\bar{x}_n\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}_n\right)^2}\\ \dfrac{-\bar{x}_n\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}_n\right)^2} & \dfrac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}_n\right)^2}\end{bmatrix}\right)$$

Idea de la prueba (I)

Si $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ son una m.a. $N\left(0, \sigma^2\right)$; $Y_1, Y_2, \ldots, Y_n$ heredan esa normalidad y son independientes entre sí, $Y_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$. Entonces,

$$L\left(\beta_0, \beta_1, \sigma^2 \middle| \mathbf{y}, \mathbf{x}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-\left(y_i - \beta_0 - \beta_1 x_i\right)^2}{2\sigma^2}\right] \text{ y}$$

$$l\left(\beta_0, \beta_1, \sigma^2 \middle| \mathbf{y}, \mathbf{x}\right) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2.$$

Para estimar a β_0, β_1:

$$\max l\left(\beta_0, \beta_1, \sigma^2 \middle| \mathbf{y}, \mathbf{x}\right) = \max\left\{-\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2\right\}$$

$$= \max\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2\right\} = \min\left\{\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2\right\}.$$

Luego, los estimadores ML coinciden con los estimadores MCO (OLS en inglés).

Idea de la prueba (II)

Retomando la densidad de Y, se puede ver que este modelo pertenece a la fam. exponencial de densidades triparamétrica con $\boldsymbol{\theta} = \left(\beta_0, \beta_1, \sigma^2\right)^T$ porque

$$f_{Y_i}\left(y_i, \boldsymbol{\theta}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-\left(y_i - \beta_0 - \beta_1 x_i\right)^2}{2\sigma^2}\right]$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-y_i^2}{2\sigma^2} + \frac{y_i\left(\beta_0 + \beta_1 x_i\right)}{\sigma^2} - \frac{\left(\beta_0 + \beta_1 x_i\right)^2}{2\sigma^2}\right].$$ De donde se tiene:

$$a\left(\boldsymbol{\theta}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left(\beta_0 + \beta_1 x_i\right)^2}{2\sigma^2}\right], \qquad b\left(y\right) = 1,$$

$$\mathbf{c}\left(\boldsymbol{\theta}\right) = \left(\frac{-1}{2\sigma^2}, \frac{\beta_0}{\sigma^2}, \frac{\beta_1}{\sigma^2}\right)^T, \qquad \mathbf{d}\left(y\right) = \left(y_i^2, y_i, x_i y_i\right)^T.$$

Luego, $\mathbf{d}\left(\mathbf{y}\right) = \left(\sum_{i=1}^{n} y_i^2, \sum_{i=1}^{n} y_i, \sum_{i=1}^{n} x_i y_i\right)^T$ es un vector de estadísticas suficientes y completas.

Como $\hat{\beta}_0, \hat{\beta}_1$ son insesgados y función de $\mathbf{d}\left(\mathbf{y}\right)$, estos son UMVUEs.

La normalidad multivariada se probará más adelante.

Finalmente,

- $$Var\left(\hat{\beta}_1\right)=Var\left(\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)Y_i}{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2}\right)=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2 Var\left[Y_i\right]}{\left\{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2\right\}^2}=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2\sigma^2}{\left\{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2\right\}^2}=\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2}.$$

- $$Var\left(\hat{\beta}_0\right)=Var\left(\overline{Y}_n-\hat{\beta}_1\overline{x}_n\right)=Var\left(\overline{Y}_n\right)+\overline{x}_n^2 Var\left(\hat{\beta}_1\right)-2Cov\left(\overline{Y}_n,\hat{\beta}_1\overline{x}_n\right)$$

$$=\frac{\sigma^2}{n}+\frac{\overline{x}_n^2\sigma^2}{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2}-2\overline{x}_n Cov\left(\overline{Y}_n,\hat{\beta}_1\right)$$

$$=\sigma^2\left[\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2+n\overline{x}_n^2}{n\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2}\right]-2\overline{x}_n Cov\left(\frac{1}{n}\sum_{i=1}^{n}Y_i,\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)Y_i}{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2}\right)$$

$$=\sigma^2\left[\frac{\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2}\right]-\frac{2\overline{x}_n}{n\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2}Cov\left(\sum_{i=1}^{n}Y_i,\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)Y_i\right)$$

$$= \sigma^2 \left[\frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} \left(x_i - \overline{x}_n \right)^2} \right] - \frac{2\overline{x}_n}{n \sum_{i=1}^{n} \left(x_i - \overline{x}_n \right)^2} \sum_{i=1}^{n} \sum_{j=1}^{n} Cov\left(Y_i, \left(x_j - \overline{x}_n \right) Y_j \right)$$

$$= \sigma^2 \left[\frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} \left(x_i - \overline{x}_n \right)^2} \right] - \frac{2\overline{x}_n}{n \sum_{i=1}^{n} \left(x_i - \overline{x}_n \right)^2} \sum_{i=1}^{n} Cov\left(Y_i, \left(x_i - \overline{x}_n \right) Y_i \right)$$

$$= \sigma^2 \left[\frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} \left(x_i - \overline{x}_n \right)^2} \right] - \frac{2\overline{x}_n}{n \sum_{i=1}^{n} \left(x_i - \overline{x}_n \right)^2} \sum_{i=1}^{n} \left(x_i - \overline{x}_n \right) \underbrace{Cov\left(Y_i, Y_i \right)}_{=Var\left(Y_i \right) = \sigma^2} = \sigma^2 \left[\frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} \left(x_i - \overline{x}_n \right)^2} \right].$$

- $Cov\left(\hat{\beta}_0, \hat{\beta}_1 \right) = Cov\left(\overline{Y}_n - \hat{\beta}_1 \overline{x}_n, \hat{\beta}_1 \right) = Cov\left(\overline{Y}_n, \hat{\beta}_1 \right) - Cov\left(\hat{\beta}_1 \overline{x}_n, \hat{\beta}_1 \right)$

$$= 0 - \overline{x}_n Cov\left(\hat{\beta}_1, \hat{\beta}_1 \right) = -\overline{x}_n Var\left(\hat{\beta}_1 \right) = \left[\frac{-\overline{x}_n \sigma^2}{\sum_{i=1}^{n} \left(x_i - \overline{x}_n \right)^2} \right].$$

Coeficientes de regresión (III)

Bajo normalidad, el estimador ML de σ^2 es

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2,$$

sin embargo, este estimador es sesgado. Un estimador insesgado de la varianza es

$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$. Sobre este estimador, es posible probar que:

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2),$$ y es independiente de $\hat{\beta}_0$ y $\hat{\beta}_1$.

- Se probará más adelante su insesgamiento, su distribución y su independencia.

Intervalos de confianza para los coeficientes de regresión bajo normalidad

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{Var\left(\hat{\beta}_1\right)}}=\sqrt{\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2}\left(\frac{\hat{\beta}_1-\beta_1}{\sigma}\right)\sim N\left(0,1\right)$$

indep.

$$\perp \quad \frac{\left(n-2\right)\hat{\sigma}^2}{\sigma^2}\sim\chi^2\left(n-2\right)$$

$$\frac{\hat{\beta}_0-\beta_0}{\sqrt{Var\left(\hat{\beta}_0\right)}}=\sqrt{\frac{n\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2}{\sum_{i=1}^{n}x_i^2}}\left(\frac{\hat{\beta}_0-\beta_0}{\sigma}\right)\sim N\left(0,1\right)$$

- **Cantidad pivote:**

$$\frac{\hat{\beta}_i-\beta_i}{\sqrt{V\hat{a}r\left(\hat{\beta}_i\right)}}\sim t\left(n-2\right)$$

$$IC_{100(1-\alpha)\%}\left(\beta_i\right)=\hat{\beta}_i\mp t_{1-\frac{\alpha}{2}}\left(n-2\right)\sqrt{V\hat{a}r\left(\hat{\beta}_i\right)}$$

Pruebas de hipótesis para los coeficientes de regresión bajo normalidad

(Sist. a dos colas) **(Sist. con cola a derecha)** **(Sist. con cola a izquierda)**

$$\begin{cases} H_0 : \beta_i = \tilde{\beta} \\ versus \\ H_1 : \beta_i \neq \tilde{\beta} \end{cases} \qquad \begin{cases} H_0 : \beta_i \leq \tilde{\beta} \\ versus \\ H_1 : \beta_i > \tilde{\beta} \end{cases} \qquad \begin{cases} H_0 : \beta_i \geq \tilde{\beta} \\ versus \\ H_1 : \beta_i < \tilde{\beta} \end{cases}$$

- **Valor calculado:** $t_C = \dfrac{\hat{\beta}_i - \tilde{\beta}}{\sqrt{V\hat{a}r\left(\hat{\beta}_i\right)}}$
- **Regla de decisión:**

(Sist. a dos colas) **(Sist. con cola a derecha)** **(Sist. con cola a izquierda)**

$$\tau\text{:"Rechazar } H_0 \text{ si}$$

$$\left|t_C\right| > t_{1-\frac{\alpha}{2}}\left(n-2\right)" \qquad t_C > t_{1-\alpha}\left(n-2\right)" \qquad t_C < t_{\alpha}\left(n-2\right)"$$

$$p - values =$$

$$2p\left(T_C > \left|t_C\right| \middle| H_0\right) \qquad p\left(T_C > t_C \middle| H_0\right) \qquad p\left(T_C < t_C \middle| H_0\right)$$

Intervalos de confianza para la media

$\mu_i = \beta_0 + \beta_1 x_i$: Valor esperado para una unidad con valor x_i de la covariable.

$$\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \text{ : Estimación puntual.}$$

$$E\left[\hat{\mu}_i\right] = E\left[\hat{\beta}_0 + \hat{\beta}_1 x_i\right] = \beta_0 + \beta_1 x_i = \mu_i$$

$$Var\left[\hat{\mu}_i\right] = Var\left[\hat{\beta}_0 + \hat{\beta}_1 x_i\right] = Var\left(\hat{\beta}_0\right) + x_i^2 Var\left(\hat{\beta}_1\right) + 2x_i \operatorname{cov}\left(\hat{\beta}_0, \hat{\beta}_1\right)$$

$$\hat{\mu}_i \sim N\left(\mu_i, Var\left[\hat{\mu}_i\right]\right)$$

- **Cantidad pivote:**

$$\frac{\hat{\mu}_i - \mu_i}{\sqrt{V\hat{a}r\left(\hat{\mu}_i\right)}} \sim t\left(n-2\right)$$

$$IC_{100(1-\alpha)\%}\left(\mu_i\right) = \hat{\mu}_i \mp t_{1-\frac{\alpha}{2}}\left(n-2\right)\sqrt{V\hat{a}r\left(\hat{\mu}_i\right)}$$

Intervalos de predicción

Sea x^* el valor de la covariable para una **nueva unidad** de la cual no se conoce su respuesta Y^*.¿Cómo construir un int. de **predicción** para Y^*?

$\mu^* = \beta_0 + \beta_1 x^*$: Valor esperado para una unidad con valor x^*.

$$\hat{\mu}^* \sim N\left(\mu^*, Var\left[\hat{\mu}^*\right]\right)$$

Sea $e^* = Y^* - \hat{\mu}^*$

$$E\left[e^*\right] = E\left[Y^* - \hat{\mu}^*\right] = 0$$

$$Var\left[e^*\right] = Var\left[Y^* - \hat{\mu}^*\right] = Var\left(Y^*\right) + Var\left[\hat{\mu}^*\right] = \sigma^2 + Var\left[\hat{\mu}^*\right]$$

$$e^* \sim N\left(0, \sigma^2 + Var\left[\hat{\mu}^*\right]\right)$$

- **Cantidad pivote:**

$$\frac{e^*}{\sqrt{\hat{\sigma}^2 + V\hat{a}r\left(\hat{\mu}^*\right)}} \sim t\left(n-2\right) \quad IP_{100(1-\alpha)\%}\left(Y^*\right) = \hat{\mu}^* \mp t_{1-\frac{\alpha}{2}}\left(n-2\right)\sqrt{\hat{\sigma}^2 + V\hat{a}r\left(\hat{\mu}^*\right)}$$