eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Tema 6

El Modelo de Regresión Lineal General

Inferencia y predicción

Pilar González y Susan Orbe

Dpto. Economía Aplicada III (Econometría y Estadística)

Objetivos de aprendizaje

- Comprender qué es la inferencia.
- Derivar la distribución del estimador MCO bajo los supuestos clásicos del modelo de regresión.
- Estimar los coeficientes del modelo de regresión por intervalo.
- Contrastar hipótesis sobre los valores poblacionales de los coeficientes.
- Predecir el valor de la variable endógena.

Contenido

Contenido

¿Qué es inferencia?

En los temas anteriores se ha especificado el modelo de regresión lineal que establece una relación entre variables económicas y se ha estimado por Mínimos Cuadrados Ordinarios.

Basados en los supuestos del modelo de regresión lineal y, dadas las estimaciones de los parámetros del modelo obtenidas con la muestra, el objetivo es obtener conclusiones sobre los valores de los parámetros poblacionales, midiendo a la vez su significación, es decir, la confianza que nos merecen.

Ejemplo. Consumo de pizza.

$$pizza_i = \beta_1 + \beta_2 renta_i + \beta_3 edad_i + \beta_4 M_i + \beta_5 B_i + \beta_6 U_i + \beta_7 P_i + u_i$$

Algunas preguntas de interés:

¿Afecta la renta al consumo de pizza? ¿Consumen más pizza los hombres que las mujeres? ¿Depende el consumo de pizza del nivel educativo del cliente? ¿Tienen el mismo nivel de consumo los clientes con educación secundaria y universitaria?

¿Qué es inferencia?

Instrumentos para la inferencia.

a) **Intervalos de confianza:**

Procedimiento para establecer un rango de valores, denominado intervalo de confianza, en el que es probable que se encuentren los valores desconocidos de los parámetros.
El hecho de contar con un rango de valores para los parámetros (en vez de con un sólo valor o estimación) proporciona información no sólo sbore cuál puede ser el valor del parámetro, sino también sobre la precisión con la que ha sido estimado.

b) **Contrastes de hipótesis**:

Procedimientos que permiten comparar las conjeturas (hipótesis) que podamos hacer sobre los valores poblacionales de los parámetros y las estimaciones obtenidas de la muestra. Los contrastes de hipótesis nos permiten decidir si los datos de la muestra son compatibles o no con las conjeturas (hipótesis) planteadas.

¿Qué es inferencia?

La inferencia trata de validar conjeturas sobre los parámetros poblacionales.

Ejemplo. Consumo de pizza.

$$pizza_i = \beta_1 + \beta_2 renta_i + \beta_3 edad_i + \beta_4 M_i + \beta_5 B_i + \beta_6 U_i + \beta_7 P_i + u_i$$

¿Afecta la renta al consumo de pizza? $\Rightarrow$ $\beta_2 = 0$

¿Consumen más pizza las mujeres que los hombres? $\Rightarrow$ $\beta_4 > 0$

¿Depende el consumo del nivel educativo del cliente? $\Rightarrow$ $\beta_5 = \beta_6 = \beta_7 = 0$

¿Tienen el mismo nivel de consumo los clientes con educación secundaria y universitaria? $\Rightarrow$ $\beta_5 = \beta_6$

Contenido

Distribución del estimador MCO.

Los procedimientos que se utilizan para estimar los intervalos de confianza o para construir los contrastes de hipótesis dependen de la distribución del estimador MCO.

$$\boxed{\hat{\beta} = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'u}$$

Sabemos que bajo los supuestos del modelo de regresión lineal S1-S6,

$$\begin{aligned} E(Y|X) &= X\beta \\ V(Y|X) &= E(Y|X - E(Y|X))(Y|X - E(Y|X))' = E_X(uu') = \sigma^2 \\ Y|X &\sim \text{distribución normal} \end{aligned}$$

$$\Rightarrow \quad \boxed{Y|X \sim N(X\beta, \sigma^2)}$$

Distribución del estimador MCO.

- Condicionado a los valores de X, $\hat{\beta}$ es lineal en Y por lo que:

$$\hat{\beta}|X \;\sim\; \text{Normal}$$

- Media de la distribución:

$$E(\hat{\beta}|X) = \beta + E[(X'X)^{-1}X'u|X] = \beta \qquad (E_X(\hat{\beta}_j) = \beta_j,\; j = 1, 2, \ldots, k)$$

- Varianza de la distribución

$$V(\hat{\beta}|X) = \sigma^2(X'X)^{-1} \qquad (Var_X(\hat{\beta}_j) = \sigma^2_{\hat{\beta}_j} = \sigma^2\, a_{jj}\;\; j = 1, 2, \ldots, k)$$

La varianza de los estimadores MCO proporciona información sobre la precisión de los estimadores $\hat{\beta}$. Como los estimadores MCO son insesgados, cuanto menor sea su varianza, mayor será la probabilidad del que obtengamos estimaciones "cerca" de los verdaderos valores de los parámetros.

$$\boxed{\hat{\beta}|X \;\sim\; N[\beta, V(\hat{\beta})]} \qquad \boxed{\hat{\beta}_j|X \;\sim\; N[\beta_j, V(\hat{\beta}_j)] \quad j = 1, 2, \ldots, k}$$

Contenido

Estimación de los coeficientes

A. Estimación por punto: $\hat{\beta} = (X'X)^{-1}X'Y$

B. Estimación por intervalo:

Intervalo de confianza proporciona un rango de valores probables para el parámetro poblacional β_j.

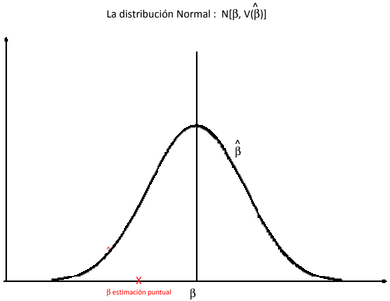
$$[A \quad B] \quad \text{tal que} \quad P[A < \beta_j < B] = (1-\alpha)\,\%$$

Interpretación:

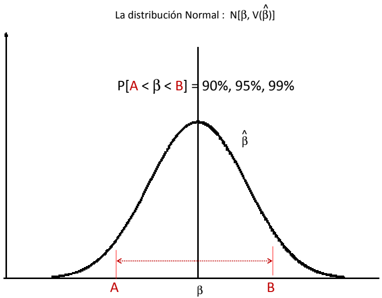
Si obtuviéramos muestras aleatorias repetidamente y, para cada una de ellas, calculáramos A y B, entonces el parámetro poblacional desconocido β_j se encontraría en el intervalo $[A \;\; B]$ para $(1-\alpha)\,\%$ de las muestras.

Estimación por intervalo.

Estimación por punto



Estimación por intervalo



Estimación por intervalo.

$$\hat{\beta}_j \sim N[\beta_j, V(\hat{\beta}_j)] \qquad \longrightarrow_{Estandarizando} \quad \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N[0,1]$$

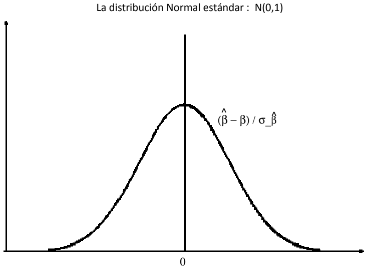
Elegimos un nivel de probabilidad: $(1-\alpha)\,\%$

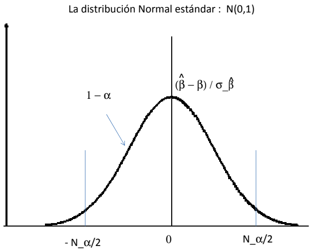
$$P\left[-N_{\alpha/2} \;\leq\; \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \leq N_{\alpha/2}\right] = (1-\alpha)\,\%$$

$$P\left[\hat{\beta}_j - N_{\alpha/2}\;\sigma_{\hat{\beta}_j} \;\leq\; \beta_j \;\leq\; \hat{\beta}_j + N_{\alpha/2}\;\sigma_{\hat{\beta}_j}\right] = (1-\alpha)\,\%$$

$$IC(\beta_j)_{(1-\alpha)\,\%} = \left[\hat{\beta}_j - N_{\alpha/2}\;\sigma_{\hat{\beta}_j} \qquad \hat{\beta}_j + N_{\alpha/2}\;\sigma_{\hat{\beta}_j}\right]$$

Estimación por intervalo.





$P[- N\_\alpha/2 < (\hat{\beta} - \beta) / \sigma\_\hat{\beta} < N\_\alpha/2] = (1-\alpha)\%$

Estimación por intervalo.

PROBLEMA:

$$V(\hat{\beta}_j) = \sigma^2_{\hat{\beta}_j} = \sigma^2 \, a_{jj}$$ es habitualmente desconocida.

Se puede estimar: $$\hat{V}(\hat{\beta}_j) = \hat{\sigma}^2_{\hat{\beta}_j} = \hat{\sigma}^2 \, a_{jj}$$

¿Cuál es la distribución de: $$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim ????$$

Se puede derivar sabiendo que:

$$\frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N[0,1] \qquad \text{y} \qquad \frac{\sum \hat{u}_t^2}{\sigma^2} \sim \chi^2(T-k)$$

Y que se puede demostrar que estas dos variables aleatorias son independientes. .

Estimación por intervalo.

Aplicando el resultado siguiente:

$$\frac{N(0,1)}{\sqrt{\frac{\chi^2(T-k)}{T-k}}} \;=\; t(T-k)$$

se obtiene que:

$$\frac{\frac{\hat{\beta}_j-\beta_j}{\sigma_{\hat{\beta}_j}}}{\sqrt{\frac{\frac{\sum \hat{u}_t^2}{\sigma^2}}{T-k}}} \;=\; \frac{\frac{\hat{\beta}_j-\beta_j}{\sigma\sqrt{a_{jj}}}}{\sqrt{\frac{\frac{\sum \hat{u}_t^2}{\sigma^2}}{T-k}}} \;=\; \frac{\hat{\beta}_j-\beta_j}{\hat{\sigma}\sqrt{a_{jj}}} \;=\; \frac{\hat{\beta}_j-\beta_j}{\hat{\sigma}_{\hat{\beta}_j}}$$

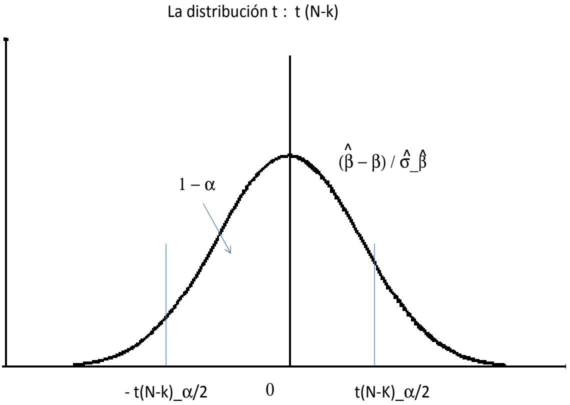
$$\frac{\hat{\beta}_j-\beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \;\sim\; t(T-k)$$

Recuérdese que la distribución t de Student está centrada en cero pero tiene las colas más anchas que la distribución normal estándar.

Estimación por intervalo.

Los intervalos de confianza se construirán habitualmente basados en:

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t(T-k)$$



P[- t(N-k)_α/2 < $(\hat{\beta} - \beta) / \hat{\sigma}\_\hat{\beta}$ < t(N-k)_α/2] = (1-α)%

Estimación por intervalo.

Intervalo de confianza.

$$P\left[-t_{\alpha/2}(T-k) \quad \leq \quad \frac{\hat{\beta}_j-\beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \quad \leq \quad t_{\alpha/2}(T-k)\right] = (1-\alpha)\,\%$$

$$P\left[\hat{\beta}_j - t_{\alpha/2}(T-k)\,\hat{\sigma}_{\hat{\beta}_j} \quad \leq \quad \beta_j \quad \leq \quad \hat{\beta}_j + t_{\alpha/2}(T-k)\,\hat{\sigma}_{\hat{\beta}_j}\right] = (1-\alpha)\,\%$$

$$IC(\beta_j)_{(1-\alpha)\,\%} = \left[\hat{\beta}_j - t_{\alpha/2}(T-k)\,\hat{\sigma}_{\hat{\beta}_j} \qquad \hat{\beta}_j + t_{\alpha/2}(T-k)\,\hat{\sigma}_{\hat{\beta}_j}\right]$$

Interpretación: Si este estimador por intervalo se usa en muchas muestras de la población, entonces $(1-\alpha)\,\%$ de los $IC(\beta_j)_{(1-\alpha)\,\%}$ contendrán el verdadero valor de β_j.

Aplicación en el **Ejemplo 6.1.**

Contenido

Contrastes de hipótesis.

Sobre los coeficientes del modelo de regresión.

Etapas de realización de un contraste:

Paso 1. Formular la hipótesis en términos de los coeficientes de regresión.
$\rightarrow$ Determinar la hipótesis nula y la alternativa.

Paso 2. Especificar el estadístico para contraste y su distribución cuando la hipótesis nula es cierta.

Paso 3. Calcular el valor del estadístico para la muestra disponible.

Paso 4. Regla de decisión.

Aplicaciones de los distintos contrastes en los **Ejemplo 6.2 y 6.3.**

Ejemplo. Consumo de pizza.

$$pizza_i = \beta_1 + \beta_2 renta_i + \beta_3 edad_i + \beta_4 M_i + \beta_5 B_i + \beta_6 U_i + \beta_7 P_i + u_i$$

Caso A. ¿Afecta la renta al consumo de pizza? $\Rightarrow \ \beta_2 = 0$

Caso B. ¿Consumen más pizza las mujeres que los hombres? $\Rightarrow \ \beta_4 > 0$

Caso C. ¿Depende el consumo del nivel educativo del cliente?
$\Rightarrow \ \beta_5 = \beta_6 = \beta_7 = 0$

Caso D. ¿Tienen el mismo nivel de consumo los clientes con educación secundaria y universitaria? $\Rightarrow \ \beta_5 = \beta_6$

Caso E. ¿Afecta el conjunto de variables explicativas incluidas en el modelo al consumo de pizza $\Rightarrow \ \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$?

Contrastes de hipótesis.

Contrastes sobre un coeficiente. [Caso A]

Paso 1.
$$\begin{cases} H_0 : \quad \beta_j = c \\ H_a : \quad \beta_j \neq c \end{cases}$$

Paso 2. Estadístico t de contraste
$$t = \frac{\hat{\beta}_j - c}{\hat{\sigma}_{\hat{\beta}_j}} \overset{H_0}{\sim} t(T-k)$$

Paso 4. Elige un nivel de significación (α %).

$$\alpha = \text{P[Rechazar la hipótesis nula} \mid H_0 \text{ es cierta]}$$

Rechazar la hipótesis nula H_0 en favor de la alternativa H_a al nivel de significación elegido si:

$$\left.\begin{array}{l} t > t_{\alpha/2}(T-k) \\ t < -t_{\alpha/2}(T-k) \end{array}\right\} \Rightarrow \quad |t| > t_{\alpha/2}(T-k)$$

Contraste de varias restricciones lineales. [Caso C]

Paso 1.
$$\begin{cases} H_0: & \beta_5 = 0,\ \beta_6 = 0,\ \beta_7 = 0 \quad (\beta_5 = \beta_6 = \beta_7 = 0) \\ H_a: & \beta_5 \neq 0 \quad \text{y/o} \quad \beta_6 \neq 0 \quad \text{y/o} \quad \beta_7 \neq 0 \end{cases}$$

Modelo No Restringido:

$$pizza_i = \beta_1 + \beta_2 renta_i + \beta_3 edad_i + \beta_4 M_i + \beta_5 B_i + \beta_6 U_i + \beta_7 P_i + u_i \qquad (1)$$

Modelo Restringido:

$$pizza_i = \beta_1 + \beta_2 renta_i + \beta_3 edad_i + \beta_4 M_i + u_i \qquad (2)$$

Contrastes de hipótesis.

Contraste de varias restricciones lineales. [Caso C]

Paso 2. Estadístico de contraste F

$$F = \frac{(SCR_R \; - \; SCR_{NR})/q}{SCR_{NR}/(T-k)} \overset{H_0}{\sim} \mathcal{F}(q, T-k)$$

donde:

SCR_R: suma de cuadrados de residuos del modelo restringido y
SCR_{NR}: suma cuadrados de residuos del modelo no restringido.

Paso 4. Se rechaza la hipótesis nula H_0 en favor de H_a al nivel de significación α % si:

$$F > \mathcal{F}_\alpha(q, T-k)$$

Contrastes de hipótesis.

Contraste de significación conjunta. [Caso E]

Efecto conjunto de las variables explicativas

Paso 1.

$$\begin{cases} H_0: \quad \beta_2 = \beta_3 = \beta_4 = \ldots = \beta_7 = 0 \\ H_a: \quad \beta_2 \neq 0 \quad \text{y/o} \quad \beta_3 \neq 0 \text{ y/o} \quad \beta_4 \neq 0 \ \ldots \ \beta_7 \neq 0 \end{cases}$$

Modelo No Restringido:

$$pizza_i = \beta_1 + \beta_2 renta_i + \beta_3 edad_i + \beta_4 M_i + \beta_5 B_i + \beta_6 U_i + \beta_7 P_i + u_i \qquad (3)$$

Modelo Restringido:

$$pizza_i = \beta_1 + u_i \qquad (4)$$

Contraste de significación conjunta. [Caso E]

Paso 2. Estadístico F para este contraste en particular:

$$F = \frac{(SCR_R - SCR_{NR})/q}{SCR_{NR}/(T-k)} = \frac{(SCR_R/SCT - SCR_{NR}/SCT)/q}{SCR_{NR}/SCT/(T-k)} =$$

$$= \frac{[(1-R_R^2) - (1-R_{NR}^2)]/q}{(1-R_{NR}^2)/(T-k)} = \frac{R_{UR}^2/q}{(1-R_{NR}^2)/(T-k)}$$

$$F = \frac{R^2/q}{(1-R^2)/(T-k)} \overset{H_0}{\sim} \mathcal{F}(q, T-k)$$

Paso 4. Se rechaza H_0 en favor de H_a al nivel de significación α % si:

$$F > \mathcal{F}_\alpha(q, T-k)$$

Contrastes de hipótesis.

Contrastes a una cola. [Caso B]

¿Consumen más pizza las mujeres que los hombres?

Paso 1.

$$\begin{cases} H_0: & \beta_4 = 0 \ (\beta_4 \leq 0) \\ H_a: & \beta_4 > 0 \end{cases}$$

Paso 2. Estadístico de contraste t $\quad t = \dfrac{\hat{\beta}_4 - 0}{\hat{\sigma}_{\hat{\beta}_4}} \overset{H_0}{\sim} t(T-k)$

Paso 4. Se rechaza la hipótesis nula H_0 en favor de H_a al nivel de significación de α % si:

$$t > t_\alpha(T-k)$$

Contenido

Predicción.

¿Qué es predecir?

Sea el modelo de regresión lineal:

$$Y = X\beta + u \qquad u|X \sim N(0, \sigma^2(X'X)^{-1})$$

$$\boxed{\text{Predicción:} \quad \text{Estimar} \quad Y_p \quad \text{para valores dados de} \quad X_p}$$

Y_p es una variable aleatoria. Se proponen dos tipos de predicción:

- Predicción por punto, $\hat{Y}_p$: un único valor de la distribución de Y_p.
- Predicción por intervalo: proporciona una medida de la incertidumbre de la predicción por punto, $\hat{Y}_p$, mediante la construcción de un intervalo de confianza.

Predicción.

Predicción por punto.

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t \qquad t = 1, 2, \ldots, T$$

$$Y_t = [1 \; X_{2t} \; X_{3t} \; \ldots \; X_{kt}] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + u_t \qquad\qquad Y_t = X_t' \beta + u_t$$

Supuestos para la predicción:

▷ La relación entre Y y las variables explicativas X es válida fuera de la muestra.

▷ $E(u_p|X) = 0, \quad V(u_p|X) = \sigma^2 \quad cov(u_p u_i|X) = 0, i = 1, 2, \ldots T.$

Predicción.

De acuerdo con el MRLG:

$$Y_p \quad = \quad X_p'\,\beta + u_p$$

Dados los valores de las variables explicativas, X_p, la predicción por punto viene dada por:

$$\hat{Y}_p = X_p'\,\hat{\beta}$$

Error de predicción:

$$e_p \;=\; Y_p - \hat{Y}_p \;=\; Y_p - X_p'\,\hat{\beta} \;=\; X_p'\,\beta + u_p - X_p'\,\hat{\beta} \;=\; X_p'\,(\beta - \hat{\beta}) + u_p$$

La predicción por intervalo se deriva a partir de la distribución del error de predicción.

Predicción.

Distribución del error de predicción:

$$e_p = Y_p - \hat{Y}_p \sim N(0, \sigma_e^2)$$

1. Distribución normal.

2. Media: $E(e_p) = E(X_p'\,(\beta - \hat{\beta}) + u_p) = X_p'\,(\beta - E(\hat{\beta})) + E(u_p) = 0$

3. Varianza:

$$\begin{aligned}
V(e_p) &= \sigma_e^2 = E[e_p - E(e_p)]^2 = E[(X_p'\,(\beta - \hat{\beta}) + u_p)\;(X_p'\,(\beta - \hat{\beta}) + u_p)'] = \\
&= E[(X_p'\,(\beta - \hat{\beta})(\beta - \hat{\beta})' X_p + u_p\,(\beta - \hat{\beta})' X_p + X_p'\,(\beta - \hat{\beta})\,u_p' + u_p u_p'] = \\
&= X_p'\,E[(\beta - \hat{\beta})(\beta - \hat{\beta})']\,X_p + E(u_p u_p') = X_p'\,V(\hat{\beta})\,X_p + \sigma_u^2 = \\
&= \sigma_u^2 + X_p'\,\sigma_u^2 (X'X)^{-1}\,X_p = \sigma_u^2\;[1 + X_p'\,(X'X)^{-1}\,X_p]
\end{aligned}$$

Dado que: $E[X_p'\,(\beta - \hat{\beta})\,u_p'] = X_p'\,E[(\beta - \hat{\beta})\,u_p'] = X_p'\,E[(X'X)^{-1}\,X'u\,u_p'] = 0$

Predicción por intervalo.

Se parte de la distribución del error de predicción

$$e_p = Y_p - \hat{Y}_p \sim N(0, \sigma_e^2)$$

- Estandarizando:

$$\frac{e_p}{\sigma_e} \sim N(0, 1)$$

- La varianza del error de predicción normalmente es desconocida:

$$\sigma_e^2 = \sigma^2 \left[1 + X_p' \left(X'X\right)^{-1} X_p\right]$$

Tiene que ser estimada: $\hat{\sigma}_e^2 = \hat{\sigma}^2 \left[1 + X_p' \left(X'X\right)^{-1} X_p\right]$

Predicción.

- Dado que :

$$(T-k)\ \frac{\hat{\sigma}^2}{\sigma^2} = \frac{\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2(T-k)$$

$$\frac{e_p}{\hat{\sigma}_e} \sim t(T-k) \qquad \Rightarrow \qquad \frac{Y_p - \hat{Y}_p}{\hat{\sigma}_e} \sim t(T-k)$$

Intervalo de Predicción (1- α) %

$$P\left[-t_{\alpha/2}(T-k) \leq \frac{Y_p - \hat{Y}_p}{\hat{\sigma}_e} \leq t_{\alpha/2}(T-k)\right] = 1-\alpha$$

$$P\left[\hat{Y}_p - t_{\alpha/2}(T-k)\ \hat{\sigma}_e \leq Y_p \leq \hat{Y}_p + t_{\alpha/2}(T-k)\ \hat{\sigma}_e\right] = 1-\alpha$$

$$IC(Y_p)_{(1-\alpha)\,\%} = \left[\hat{Y}_p \pm t_{\alpha/2}(T-k)\ \hat{\sigma}_e\right]$$

Aplicación en el **Ejemplo 6.4.**

Contenido

Contenido