

# Pauta Mini Control VII

## Estadística para la Gestión II

**Profesor:** Miguel Ángel Zapata Pérez

**Ayudante:** Rubén Ananías Martínez

**Primavera 2024**

### Instrucciones:

- Tienen 48 horas desde que el mini control es subido a la web.
- Sus respuestas deben ser subido a U-Cursos.
- Deben enviar un script en R con sus comandos, y el siguiente nombre: `RUT_C7.r`, y un archivo word con sus respuestas con el siguiente nombre: `RUT_C7.docx`

### Pregunta I - 3 puntos

Un municipio está interesado en entender los factores que afectan el porcentaje de participación ciudadana en elecciones municipales. Se cuenta con los siguientes datos recolectados de 50 comunas:

- $Y$ : Porcentaje de participación ciudadana en las elecciones municipales.
- $X_1$ : Porcentaje de hogares con acceso a internet.
- $X_2$ : Gasto en campañas de sensibilización por habitante (en miles de pesos).

### Preguntas:

1. Escribe la ecuación de regresión lineal múltiple que podría utilizarse para modelar este problema. Define el significado de cada coeficiente.

### Respuesta:

La ecuación de regresión lineal múltiple que modela este problema es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Donde:

- $Y$ : Porcentaje de participación ciudadana en elecciones municipales.
- $\beta_0$ : Intercepto del modelo, representa el porcentaje de participación ciudadana esperado cuando  $X_1$  y  $X_2$  son iguales a cero.
- $\beta_1$ : Coeficiente asociado al porcentaje de hogares con acceso a internet ( $X_1$ ), representa el cambio en el porcentaje de participación ciudadana por cada incremento de un punto porcentual en el acceso a internet, manteniendo constante  $X_2$ .
- $\beta_2$ : Coeficiente asociado al gasto en campañas de sensibilización por habitante ( $X_2$ ), representa el cambio en el porcentaje de participación ciudadana por cada incremento de una unidad en el gasto (en miles de pesos), manteniendo constante  $X_1$ .
- $\varepsilon$ : Término de error, captura la variabilidad en  $Y$  no explicada por  $X_1$  y  $X_2$ .

2. Supón que el coeficiente asociado a  $X_1$  es  $\beta_1 = 0,25$  y el asociado a  $X_2$  es  $\beta_2 = 1,5$ . Interpreta estos valores en el contexto del problema.

**Respuesta:**

- Coeficiente  $\beta_1 = 0,25$ : Por cada aumento de un punto porcentual en el acceso a internet en los hogares ( $X_1$ ), se espera que el porcentaje de participación ciudadana ( $Y$ ) aumente en 0.25 puntos porcentuales, manteniendo constante el gasto en campañas de sensibilización ( $X_2$ ).
- Coeficiente  $\beta_2 = 1,5$ : Por cada aumento de 1,000 pesos en el gasto en campañas de sensibilización por habitante ( $X_2$ ), se espera que el porcentaje de participación ciudadana ( $Y$ ) aumente en 1.5 puntos porcentuales, manteniendo constante el acceso a internet ( $X_1$ ).

3. Con base en la ecuación, describe cómo interpretar el término de error  $\varepsilon$  y explica su importancia en este modelo.

**Respuesta:**

El término de error  $\varepsilon$  representa la variabilidad en el porcentaje de participación ciudadana ( $Y$ ) que no es explicada por las variables incluidas en el modelo ( $X_1$  y  $X_2$ ). Esto incluye:

- Factores no observados o no medidos que pueden influir en  $Y$ , como características sociodemográficas, cultura cívica o condiciones climáticas el día de las elecciones.
- Error de medición en las variables  $Y$ ,  $X_1$  o  $X_2$ .
- Cualquier interacción o relación no lineal entre las variables que el modelo no captura.

## Pregunta II - 4 puntos

Un estudio está interesado en analizar los factores que pueden influir en el consumo de combustible de diferentes modelos de automóviles. Los investigadores cuentan con datos de una muestra de vehículos, incluyendo variables como el peso del vehículo, la potencia del motor y la aceleración. Para este ejercicio, utilizaremos el conjunto de datos `mtcars` de R.

- `mpg`: Consumo de combustible (millas por galón), la variable dependiente.
- `wt`: Peso del vehículo (en miles de libras), una variable independiente.
- `hp`: Potencia del motor (caballos de fuerza), una variable independiente.
- `qsec`: Tiempo en recorrer un cuarto de milla (en segundos), una variable independiente.

**Instrucciones:** Con los datos proporcionados, responde las siguientes preguntas.

1. **Análisis de Correlación:** Calcula las correlaciones entre las variables `mpg`, `wt`, `hp` y `qsec`. ¿Qué relaciones observas entre el consumo de combustible y las otras variables?

### Respuesta:

- El consumo de combustible (`mpg`) está negativamente correlacionado con el peso (`wt`) y la potencia (`hp`), lo que indica que a mayor peso o potencia, menor es el consumo.
- Existe una correlación positiva moderada entre el consumo de combustible (`mpg`) y el tiempo en recorrer un cuarto de milla (`qsec`).

2. **Modelo de Regresión Lineal:** Estima un modelo de regresión lineal múltiple donde el consumo de combustible (`mpg`) sea la variable dependiente, y el peso, la potencia y el tiempo en recorrer un cuarto de milla (`wt`, `hp`, y `qsec`) sean las variables independientes. Interpreta los coeficientes obtenidos en el modelo.

### Respuesta:

La ecuación estimada del modelo de regresión lineal múltiple es:

$$\text{mpg} = \beta_0 + \beta_1 \cdot \text{wt} + \beta_2 \cdot \text{hp} + \beta_3 \cdot \text{qsec} + \varepsilon$$

Los coeficientes estimados son:

- $\beta_0$ : Intercepto, representa el consumo de combustible promedio cuando todas las variables independientes son iguales a cero.
- $\beta_1$ : Coeficiente asociado al peso (`wt`), indica el cambio en el consumo de combustible por cada incremento de 1,000 libras en el peso del vehículo.
- $\beta_2$ : Coeficiente asociado a la potencia (`hp`), representa el cambio en el consumo por cada aumento de 1 caballo de fuerza.

- $\beta_3$ : Coeficiente asociado al tiempo (qsec), representa el cambio en el consumo por cada segundo adicional en el tiempo en recorrer un cuarto de milla.
- $\varepsilon$ : Término de error, captura la variabilidad en  $Y$  no explicada por  $X_1$  y  $X_2$ .

3. **Predicción:** Utiliza el modelo para predecir el consumo de combustible para un automóvil con 3,000 libras de peso, 150 caballos de fuerza y un tiempo de 18 segundos en recorrer un cuarto de milla. Reporta la predicción junto con un intervalo de confianza del 95 %.

**Respuesta:**

Para un automóvil con las siguientes características:

- Peso (wt) = 3.000 libras
- Potencia (hp) = 150 caballos de fuerza
- Tiempo (qsec) = 18 segundos

La predicción del consumo de combustible (mpg) es de 21.05 millas por galón con un intervalo de confianza del 95 % entre 19.98 y 22.11.

4. **Multicolinealidad:** Evalúa la presencia de multicolinealidad en el modelo utilizando el Factor de Inflación de la Varianza (VIF). ¿Consideras que alguna variable presenta problemas de multicolinealidad?

**Respuesta:**

El análisis de multicolinealidad basado en el Factor de Inflación de la Varianza (VIF) mostró que ninguna variable tiene un VIF superior a 10, lo que sugiere que no hay problemas significativos de multicolinealidad en el modelo.