

**Instituto Tecnológico de las Américas ITLA**  
Estadística Avanzada  
Parcial Final  
Esloboan Mora

**Valor total: 20 puntos**

*Temas evaluados:*

- *Regresión Lineal simple*
- *Regresión Lineal múltiple*
- *Contraste de Hipótesis para una muestra.*

**Análisis estadístico del Huracán Erin (simulado): regresión múltiple y contraste de hipótesis**

**Contexto**

El huracán Erin (simulado) afectó indirectamente la región del Caribe durante 5 días, registrándose variables meteorológicas (viento, lluvia, presión, distancia al centro urbano) y consecuencias como cortes de energía. A partir de este conjunto de datos ficticio (archivo hurricane\_erin\_case\_study.csv), los estudiantes deberán realizar un **estudio aplicado de regresión múltiple** y un **contraste de hipótesis para una muestra**.

**Objetivos de Aprendizaje**

1. Aplicar técnicas de **regresión múltiple** para explicar y predecir la variable *cortes de energía* usando variables climáticas.
2. Desarrollar un **contraste de hipótesis para una media poblacional** con base en una de las variables (ej. velocidad del viento o lluvia).
3. Interpretar resultados de manera clara y fundamentada en la teoría estadística.
4. Usar software estadístico (Python, R o Excel) de manera reproducible y presentar el análisis en un informe.

## Actividades

### Regresión múltiple (12 pts.)

- Seleccionar al menos **tres variables independientes** (ej.: max\_wind\_kts, rainfall\_mm\_per\_hr, distance\_to\_SantoDomingo\_km).
- Ajustar un modelo de regresión múltiple para explicar power\_outages.
- Reportar:

- Ecuación estimada.

$$\text{power\_outages} = -46.32 + 0.66 \cdot \text{max\_wind\_kts} + 0.67 \cdot \text{rainfall\_mm\_per\_hr} + 0.024 \cdot \text{distance\_to\_SantoDomingo\_km}$$

- Significancia global del modelo (ANOVA / F-test).  
 $F = 139.37, p < 0.001 \rightarrow$  el modelo es estadísticamente significativo.
- Significancia individual de coeficientes (valores p).

```
const          7.604518e-15
max_wind_kts    6.115267e-22
rainfall_mm_per_hr  4.511585e-08
distance_to_SantoDomingo_km  9.106257e-08
dtype: float64
```

- Interpretación de resultados (qué variables influyen más).

**Las variables climáticas (max\_wind\_kts y rainfall\_mm\_per\_hr) son las que más afectan los cortes de energía.**

La distancia al centro urbano tiene efecto positivo pero mucho menor.

El modelo confirma que la intensidad meteorológica es el principal determinante de apagones durante el huracán Erin.

- $R^2$  y conclusiones sobre la calidad del modelo.
  - $R^2 = 0.783 \rightarrow$  el modelo explica ~78% de la variabilidad de apagones.
  - Ajustado  $R^2 = 0.777 \rightarrow$  modelo robusto incluso considerando las variables.

Instituto Tecnológico de las Américas ITLA  
Estadística Avanzada  
Parcial Final  
Esloboan Mora

**Contraste de hipótesis para una muestra (8 pts.)**

- Formular una hipótesis de investigación sobre la media de una variable. Ejemplos:
  - “La velocidad máxima del viento superó en promedio los **50 kt** durante el evento.”
  - “La precipitación media fue mayor a **15 mm/hr.**”
- Especificar:

- Hipótesis nula y alternativa.  
Queremos comprobar si **la velocidad máxima del viento superó en promedio los 50 kt** durante el evento:

$H_0: \mu \leq 50$  (la velocidad media es  $\leq 50$  kt)

$H_0: \mu \leq 50$  (la velocidad media es  $\leq 50$  kt)

**$H_0$  (nula):** La velocidad media máxima del viento es menor o igual a 50 kt.

**$H_1$  (alternativa):** La velocidad media máxima del viento es mayor a 50 kt.

Nivel de significancia:  $\alpha = 0.05$

- Estadístico de prueba (t de Student para una muestra).

Se utiliza la **t de Student para una muestra** para comparar la media de la velocidad máxima del viento con el valor de referencia (50 kt).

Fórmula:

$$t = (\text{media muestral} - \text{valor hipotético}) / (\text{desviación estándar} / \sqrt{n})$$

Valores:

- Media muestral ( $\bar{x}$ ) = 60.37 kt
- Valor bajo la hipótesis nula ( $\mu_0$ ) = 50 kt
- Desviación estándar muestral (s) = 28.53 kt
- Tamaño de la muestra (n) = 120

Cálculo:

$$t = (60.37 - 50) / (28.53 / \sqrt{120})$$

$$t \approx 10.37 / 2.604$$

$$t \approx 3.98$$

**Instituto Tecnológico de las Américas ITLA**

Estadística Avanzada

Parcial Final

Esloboan Mora

- Valor p e interpretación final.

**Valor p y decisión**

- Como es una prueba **unilateral** (mayor que), el valor p se obtiene del área a la derecha de  $t = 3.98$ .
- Con 119 grados de libertad ( $n-1$ ),  $p \approx 0.0001$  (muy menor a 0.05).

**Decisión:** Rechazamos  $H_0$ .

**Interpretación final**

- Hay **evidencia estadística suficiente** para afirmar que la velocidad máxima promedio del viento **superó los 50 kt** durante el huracán Erin.
- Esto indica que los vientos fueron suficientemente fuertes como para generar un impacto importante en la región.

**Informe final** (documento PDF o Word):

- Resumen ejecutivo.
- Metodología (datos, modelo, prueba de hipótesis).
- Resultados y gráficos.
- Conclusiones.

**Adjuntar** los documentos de trabajo en Python (.py o notebook), R (.R) o Excel (.xlsx) junto con el CSV original.