

# Universidad Panamericana

## Maestría en Ciencia de Datos

### Estadística

Enrique Ulises Báez Gómez Tagle      Joel Vázquez Anaya  
Luis Guillén      Roberto Requejo

18 de abril de 2025

## Tarea

Resuelve los siguientes ejercicios, incluye sus desarrollos en hojas aparte.

1. (2 puntos) Un desarrollador inmobiliario en la CdMx quiere comprar un terreno en la colonia Del Valle para construir unos edificios. Desea estimar el área  $A$  de dicho terreno. Cuando mide la longitud del terreno comete un error aleatorio de modo que la longitud observada  $X$  es una variable aleatoria con media  $\mu$  y varianza  $\sigma^2$ . Preocupado por su posible error, decide hacer dos mediciones independientes  $X_1$  y  $X_2$ . Para estimar el área  $A$  está en un dilema de cómo proceder por lo que propone dos estimadores:

$$\hat{A}_1 = \left( \frac{X_1 + X_2}{2} \right)^2 \quad \text{y} \quad \hat{A}_2 = \frac{X_1^2 + X_2^2}{2}$$

- a) Calcula el sesgo de cada estimador del área  $A = \mu^2$ .
- b) Determina cuál de los dos es mejor.

**Explicación:** El estimador  $\hat{A}_1$  presenta un sesgo de  $\sigma^2/2$ , menor que el sesgo de  $\hat{A}_2$  igual a  $\sigma^2$ ; por ello,  $\hat{A}_1$  se considera el mejor estimador (menor sesgo y menor MSE).

### Solución paso a paso

Sea  $X$  la medición individual con

$$E[X] = \mu, \quad \text{Var}(X) = \sigma^2, \quad X_1, X_2 \text{ i.i.d.}$$

El área real es  $A = \mu^2$ .

#### 1. Sesgo de cada estimador

$$\hat{A}_1 = \left( \frac{X_1 + X_2}{2} \right)^2 = \frac{1}{4}(X_1^2 + 2X_1X_2 + X_2^2)$$

$$\begin{aligned} E[\hat{A}_1] &= \frac{1}{4}(E[X_1^2] + 2E[X_1]E[X_2] + E[X_2^2]) \\ &= \frac{1}{4}(2(\sigma^2 + \mu^2) + 2\mu^2) = \mu^2 + \frac{\sigma^2}{2}. \end{aligned}$$

$$\boxed{\text{Bias}(\hat{A}_1) = E[\hat{A}_1] - A = \frac{\sigma^2}{2}}$$

$$\hat{A}_2 = \frac{X_1^2 + X_2^2}{2}, \quad E[\hat{A}_2] = \frac{1}{2}(E[X_1^2] + E[X_2^2]) = \mu^2 + \sigma^2.$$

$$\boxed{\text{Bias}(\hat{A}_2) = \sigma^2}$$

## 2. Comparación de MSE

Para variables normales  $N(\mu, \sigma^2)$ , los sesgos y varianzas son:

Cantidad	$\hat{A}_1$	$\hat{A}_2$
Bias	$\sigma^2/2$	$\sigma^2$
Varianza	$\sigma^2(\mu^2 + \frac{\sigma^2}{2})$	$2\sigma^2(\mu^2 + \frac{\sigma^2}{2})$
MSE = Var + Bias <sup>2</sup>	$\sigma^2(\mu^2 + \frac{\sigma^2}{2}) + \frac{\sigma^4}{4}$	$2\sigma^2(\mu^2 + \frac{\sigma^2}{2}) + \sigma^4$

Claramente el MSE de  $\hat{A}_1$  es menor, por lo que  $\hat{A}_1$  es el estimador preferido.

2. (2 puntos) Una persona saca del cajero automático \$3,300. El cajero le entrega dos billetes de \$50, seis de \$200 y cuatro de \$500; mismos que guarda en su cartera. Todos los billetes son indistinguibles.
  - a) Si esta persona saca al azar dos de estos billetes de su cartera sin reemplazo, sea  $T$  la variable aleatoria que denota el monto total de dinero que tiene cada muestra de tamaño 2. Obtenga la distribución de muestreo de  $T$ .
  - b) Calcule el valor esperado y la varianza de  $T$ .
3. (2 puntos) Un nuevo medicamento está siendo desarrollado en un laboratorio. Se toma una muestra de 10 pacientes similares para probar dicho medicamento. En cada paciente se midió el tiempo de recuperación en días obteniéndose los siguientes valores:

10, 10, 7, 12, 5, 7, 2, 1, 7, 2

- a) Obtenga un intervalo de confianza del 95 % para el tiempo promedio real de recuperación.

- b) El laboratorio establece que no lanzará el medicamento al mercado si la desviación estándar de los tiempos de recuperación es mayor a 2 días. Determina si el laboratorio lanzará o no el medicamento al mercado mediante la construcción de un intervalo de confianza al 98 %.
  - c) ¿Cuál es el tamaño de muestra necesario para afirmar, con una probabilidad de 0.9 que el tiempo promedio estimado de recuperación no dista del tiempo real de recuperación en más de 3 días?
- 4. (2 puntos) Por la temporada navideña una tienda departamental recibe un lote de chocolates en forma de Santa Claus. El gerente de compras desea estimar la proporción de chocolates rotos en el lote, para ello toma una muestra aleatoria de 100 chocolates, de los cuales 10 están rotos.
  - a) Construya un intervalo de confianza al 95 % para la verdadera proporción de chocolates rotos en el lote.
  - b) La fábrica de chocolates acepta la devolución si el lote contiene más del 5 % de chocolates rotos. Plantea el problema como una prueba de hipótesis.
  - c) ¿Cuál sería la recomendación al gerente de compras?
- 5. (1 punto) La cantidad de agua consumida por un adulto sano sigue una distribución Normal con media de 1.4 litros. Una campaña de salud promueve el consumo de cuando menos 2 litros diarios. Después de la campaña, una muestra de 10 adultos muestra el siguiente consumo en litros:

1,5, 1,6, 1,5, 1,4, 1,9, 1,4, 1,3, 1,9, 1,8, 1,7

- a) Con el nivel de significancia de 0.01, ¿se puede concluir que se ha incrementado el consumo de agua?
- b) Calcule e interprete el valor-p.

## Parte B: Opción múltiple (1 punto)

Defina si las siguientes aseveraciones son verdaderas (V) o falsas (F).

- 1. A mayor tamaño de muestra menor longitud del intervalo de confianza. V  
 El error estándar es inversamente proporcional al tamaño de muestra; al disminuir, reduce el margen de error y el ancho del intervalo.
- 2. Al disminuir el nivel de confianza, la longitud del intervalo disminuye. V  
 Un nivel de confianza menor implica un valor crítico más pequeño  $z$  o  $t$ , reduciendo el margen de error.

3. Si el coordinador de la Maestría dice que tiene una confianza del 95 % de que la media de las calificaciones de todos los estudiantes de Estadística está entre 7 y 10, ¿qué es lo que realmente está diciendo?

Si se repite muchas veces el muestreo y la construcción del intervalo, aproximadamente el 95 % de estos intervalos contendrá la verdadera media poblacional. El intervalo calculado ya está fijo; podría o no incluir la media.

4. En una prueba de hipótesis, la estadística de prueba sigue siempre una distribución Normal. F

No siempre es normal, puede ser  $t$ ,  $\chi^2$ ,  $F$ , binomial, etc.

5. En una prueba de hipótesis de dos colas, la zona de no rechazo es equivalente al intervalo de confianza para el parámetro de interés. V

En una prueba bilateral, los valores del parámetro que quedan en la zona de no rechazo son exactamente los que conforman el intervalo de confianza al mismo nivel  $1 - \alpha$ .