

PARCIAL_24-I

Asencios Menacho Luz

2024-10-08

UNIVERSIDAD NACIONAL AGRARIA LA MOLINA
DPTO. DE ESTADÍSTICA E INFORMÁTICA
PRIMERA EVALUACIÓN DE ESTRATEGIA DE MUESTREO

Instrucciones:

- Lea detenidamente cada pregunta y responda lo que se solicita.
- Responda todas las preguntas de manera clara y concisa.
- Muestre sus cálculos y razonamiento para cada problema.
- El uso de calculadora está permitido.

Parte I: Seleccione la respuesta correcta en cada una de las preguntas adjuntas: (1 pto c/u)

1) Una población está conformada por 500 estudiantes de una universidad. Se desea seleccionar una muestra aleatoria simple de 50 estudiantes para estimar la media de la cantidad de horas de estudio por semana. ¿Cuál es la probabilidad de que un estudiante específico llamado Ana sea seleccionado en la muestra?

- a) $1/10$
- b) $1/50$
- c) $1/100$
- d) $50/500$

Respuesta: b

2) ¿En qué situación sería más adecuado utilizar el muestreo estratificado que el muestreo aleatorio simple?

- a) Cuando la población es homogénea en cuanto a las variables de interés.
- b) Cuando la población es grande y diversa.
- c) Cuando no se dispone de un marco de muestreo completo.
- d) Cuando se desea obtener resultados más precisos para subgrupos específicos de la población.

Respuesta: d

Población homogénea: Elementos muy similares entre sí, con poca variabilidad.

Población heterogénea: Elementos muy distintos entre sí, con mucha variabilidad.

Este concepto es clave al seleccionar métodos de muestreo, ya que en poblaciones heterogéneas se suele preferir técnicas de muestreo estratificado para garantizar que se representen adecuadamente todas las subpoblaciones.

3) ¿En qué consiste el muestreo estratificado?

- a) Dividir la población en grupos homogéneos (estratos) y seleccionar aleatoriamente muestras de cada uno.
- b) Seleccionar una muestra aleatoria de toda la población sin tener en cuenta subgrupos.
- c) Utilizar un método no probabilístico para seleccionar una muestra de la población.
- d) Enfocarse en un solo subgrupo de la población para realizar el estudio.

Respuesta: a

4) ¿Cuáles son las desventajas del muestreo sistemático?

- a) Puede ser menos preciso que el muestreo aleatorio simple si la población no está ordenada aleatoriamente.
- b) Puede ser susceptible al sesgo si el marco de muestreo no es completo o está actualizado.
- c) Puede ser más difícil de implementar que el muestreo aleatorio simple si la población es grande.
- d) Todas las anteriores.

Respuesta: d

Ventajas del Muestreo Sistemático:

- Es más fácil de implementar en campo que un muestreo aleatorio simple, ya que se siguen intervalos regulares.
- Permite una cobertura más uniforme del área de estudio.

Posibles Limitaciones:

- Si hay algún patrón cíclico en la disposición de la contaminación (por ejemplo, si las fuentes de contaminación afectan a ciertas parcelas con un patrón regular), el muestreo sistemático podría no representar bien la variabilidad real de la población.
Este diseño de muestreo sistemático, con un punto de inicio seleccionado aleatoriamente, asegura que cada parcela tiene una igual probabilidad de ser seleccionada, lo que conserva las características de un muestreo aleatorio simple.

5) ¿En qué situación se recomienda utilizar el muestreo sistemático?

- a) Cuando la población se encuentra ordenada en una lista o secuencia.
- b) Cuando la población es muy grande y heterogénea.
- c) Cuando se necesite una muestra muy precisa.
- d) Cuando se necesita estudiar un subgrupo específico de la población.

Respuesta: a

El muestreo sistemático este método es eficiente cuando la población tiene un cierto orden y es particularmente útil para grandes poblaciones, aunque no es adecuado si la secuencia tiene patrones cíclicos, ya que podría sesgar los resultados.

Parte II: Marque V si la considera verdadera o F si la considera falsa

- 1) (F) El muestreo aleatorio simple siempre es más eficiente que el muestreo estratificado en términos de reducir la varianza de la estimación.

Respuesta: El muestreo estratificado puede ser más eficiente que el aleatorio simple en términos de reducir la varianza, especialmente cuando hay heterogeneidad dentro de la población.

- 2) (V) En el muestreo estratificado, la población se divide en grupos homogéneos llamados estratos antes de seleccionar la muestra.

- 3) (F) Un estimador sesgado es preferible a un estimador insesgado porque no afecta la precisión del estimador.

Respuesta: Un estimador insesgado es preferible, ya que proporciona una estimación no sesgada del parámetro poblacional.

- 4) (F) El tamaño de la muestra de un muestreo con probabilidades simples no afecta la precisión de los resultados.

- 5) (F) En el muestreo con probabilidades desiguales, cada unidad de la población tiene la misma probabilidad de ser seleccionada.

- 6) (V) El algoritmo de selección de Horvitz-Thompson puede ser utilizado para estimar totales y medias en muestras con probabilidades desiguales. Respuesta: Un estimador insesgado es preferible, ya que proporciona una estimación no sesgada del parámetro poblacional.

Parte III: Casos para desarrollar y analizar

- 1) Una empresa de cosméticos desea estimar la proporción de clientes que prefieren un nuevo producto de maquillaje. La empresa tiene una lista de 10 000 clientes.

- a) Calcule el tamaño de muestra necesaria para obtener una estimación con un nivel de confianza del 95% y un margen de error de $\pm 5\%$

$$n = \frac{Z^2 * p * (1 - p)}{E^2}$$

Donde:

n: Tamaño de la muestra.

Z: Valor crítico de la distribución normal estándar asociado al nivel de confianza.

p: Proporción esperada de éxito (si no se conoce, se usa 0.5 para máxima variabilidad).

q=(1-p)

E: Margen de error.

```
Z <- 1.96
p <- 0.5
q <- 0.5
E <- 0.05

n <- ((Z^2)*p*q)/(E^2)
n
```

```
## [1] 384.16
```

Resultado del tamaño de muestra: Para un nivel de confianza del 95% y un margen de error de $\pm 5\%$, la empresa necesitaría una muestra de al menos 385 clientes (siempre se redondea hacia arriba).

- b) Si la empresa selecciona una muestra aleatoria simple de 300 clientes, cual es la probabilidad de que la proporción muestral este dentro del margen de error de $\pm 5\%$ de la proporción poblacional?

$$\sigma_p = \sqrt{\frac{p * q}{n}}$$

```
p <- 0.5
q <- 0.5
n <- 300
sigma <- sqrt((p*q)/n)
sigma
```

```
## [1] 0.02886751
```

Ahora, calculamos el intervalo alrededor de la proporción poblacional: Para un margen de error de $\pm 5\%$ ($E = 0.05$), debemos calcular la probabilidad de que la proporción muestral esté en el intervalo $p \pm E$.

El margen de error estándar para esta muestra sería:

```
E <- 0.05
Z <- E / sigma
Z
```

```
## [1] 1.732051
```

Este valor de $Z = 1.732$ se puede buscar en una tabla de la distribución normal estándar para obtener la probabilidad acumulada. El valor correspondiente a $Z = 1.732$ es aproximadamente 0.9582.

Resultado: La probabilidad de que la proporción muestral esté dentro del margen de error de $\pm 5\%$ de la proporción poblacional es aproximadamente 95.82%.

CASO:

Una empresa de cosméticos desea realizar un estudio de mercado para conocer las preferencias de sus clientes respecto a una nueva línea de productos. La empresa cuenta con una base de datos de 10,000 clientes, distribuidos en diferentes regiones del país:

Región Norte: 30%

Región Centro: 40%

Región Sur: 30%

Se desea obtener una muestra representativa de 300 clientes para realizar el estudio.

- a) Indique qué tipo de muestreo probabilístico utilizaría para este caso y explique por qué.
Se utiliza el muestreo estratificado. El metodo es ideal para garantizar que la muestra sea representativa de las difrentes regiones del pais. Las regiones tienen diferentes proporciones de la población total.
- b) Describa los pasos que seguiría para seleccionar la muestra utilizando el método elegido.
Dividir la población: Serian las tres regiones.
Determinar el tamaño de la muestra para cada estrato.
Dentro de cada estrato, seleccionar MAS de clientes para garantizar que cada cliente tenga la misma posibilidad de ser slccionada.

c) Calcule el tamaño de la muestra para cada región, considerando la proporcionalidad de la población.

$$n_n = n * f$$

Se distribuye de forma proporcional según el porcentaje de clientes en cada región:

```
# Norte 30 %  
N <- 10000  
  
n <- 300  
rn <- 0.3  
A <- n*rn  
A
```

```
## [1] 90
```

```
# Centro 40%  
rc <- 0.4  
B <- n*rc  
B
```

```
## [1] 120
```

```
# Sur 30%  
rs <- 0.3  
C <- n*rs  
C
```

```
## [1] 90
```

Resultado del tamaño de la muestra:

90 clientes de la Región Norte.

120 clientes de la Región Centro.

90 clientes de la Región Sur.

d) Explique cómo seleccionaría los clientes de cada región para formar parte de la muestra.

- Dentro de cada estrato (región), se seleccionará una muestra aleatoria simple.
- Para realizar esto, se puede asignar un número a cada cliente dentro de cada región y luego utilizar una herramienta de selección aleatoria (como una tabla de números aleatorios o software como Excel o RStudio) para seleccionar los clientes.
- Repetir este proceso para las tres regiones, asegurando que se seleccionen 90 clientes de la Región Norte, 120 clientes de la Región Centro, y 90 clientes de la Región Sur.

e) Mencione algunas consideraciones adicionales que se deben tomar en cuenta al realizar el muestreo para este estudio.

- **Actualización de la base de datos:** Verificar que la base de datos de 10,000 clientes esté actualizada y que no contenga duplicados o datos erróneos.
- **No respuesta:** Tener en cuenta la posibilidad de no respuestas en la muestra seleccionada. Sería recomendable aumentar el tamaño de la muestra inicial para compensar posibles no respuestas.
- **Representatividad de subgrupos:** Asegurarse de que el muestreo también considere subgrupos importantes dentro de cada región, como edad, género, o preferencias pasadas de productos.
- **Costos y tiempos:** Considerar que realizar un muestreo estratificado puede ser más costoso y llevar más tiempo que un muestreo aleatorio simple, ya que implica dividir la población en estratos y realizar un proceso separado de selección en cada estrato.
- **Comunicaciones y logística:** Establecer un buen plan de contacto para los clientes seleccionados, asegurando que todos sean alcanzables y que la muestra final no tenga sesgo por inaccesibilidad de ciertos clientes.