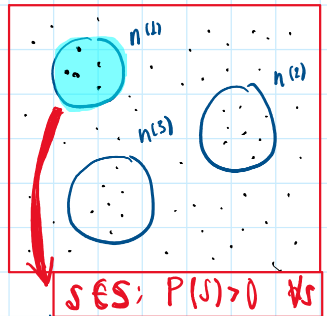
Primer Taller

jueves, 10 de julio de 2025 1:59 p. m.

1. Determine el valor de verdad de las siguientes afirmaciones.

- (a) En el muestreo probabilístico se puede definir el número de muestras posibles que pueden seleccionarse de una población con una probabilidad $P(s) > 0; \forall s$ preestablecida mediante un mecanismo adecuado. Verdadero
- (b) El diseño de una investigación por muestreo implica un proceso ordenado de tal forma que se cumple (1). Determinación de las mediciones; (2). Determinación del diseño muestral; (3). Planeación análisis estadístico; (4). Planeación trabajo de campo. Falso
- (c) Un marco muestral debe listar todas las unidades de la población objetivo sin incluir unidades ajenas. Incluye un código único de identificación, organización sistemática y con información relevante. Verdadera
- (d) Una buena muestra es una **muestra representativa**, que reproduce las características de interés que existen en la población de la forma más cercana posible. Verdadera



U = N (Población objetivo)

Orden lógico:
(1). Determinación mediciones
(2). Determinación del diseño muestral
(3). Planeación trabajo campo.
(4). Planeación análisis estadístico.

DIRECTORIO	SECUENCIA_P	ORDEN.x	CAR	SEC	CONS	REC	I_HOGAR
7910114	1	1	1	1	1	1	29795[illegible]3
7910115	1	1	1	0	1	0	189185[illegible]
7910119	1	1	1	0	1	1	1372884
7910120	1	1	1	0	1	1	1288338
7910121	1	1	1	0	1	1	1608499
7910122	1	1	1	0	1	0	2617299
7910123	1	1	1	0	1	1	1785715
7910125	1	1	0	1	1	0	2948595
7910126	1	1	1	0	1	1	1465174

Homogeneidad de la población

Respecto a una característica interés

- (e) En el muestreo aleatorio simple sin reemplazo, la probabilidad de seleccionar cualquier unidad es n/N, donde, n, el tamaño muestral, no influye en la precisión de las estimaciones. Falsa
- (f) Las muestras de conveniencia carecen de sesgo de selección, esto es, cuando alguna parte de la población objetivo no está representada en la muestra. Falsa
- (g) La varianza poblacional σ^2 puede ser estimada a partir del estimador insesgado S^2, que considera un factor de corrección por población finita.

n: Tamaño muestral $\quad \hat{\theta} \to \theta$

- (1). Tamaño muestral (n): $n \to \infty$
- (2). Nivel de confianza $(1-\alpha)$
- (3). Margen error admisible.

- Muestra conveniencia: Muestra sesgada, escogida a conveniencia: Se pueden seleccionar individuos, por facilidad de acceso o alcance.

- Varianza muestral: $S^2 = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{n-1}$; $\bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n}$

- $\sigma^2 = Var[\hat{\theta}]$: $\hat{\sigma}^2 = \widehat{Var}[\hat{\theta}]$: Factor de corrección por población finita: $\left(1 - \frac{n}{N}\right) : \left(\frac{N-n}{N}\right)$

- $\hat{\sigma}^2 = \left(\frac{N-n}{N}\right) S^2$

(1). Observación: ¿Es homogénea la muestra?

- Sí, en esta ocasión es homogénea, de forma que puede aplicarse un MAS, puesto que los hogares comparten características similares (ingreso).

2. Estime el ingreso total de los hogares τ, así como el ingreso promedio percibido μ. Calcule los correspondientes intervalos de confianza. Interprete estos resultados.

$\tau = \sum_{i=1}^{N} Y_i$; Y_i: Ingreso del i-ésimo hogar $\quad \hat{\tau} \to \tau$: $\hat{\tau} = \sum_{i=1}^{n} Y_i$

$\mu = \sum_{i=1}^{N} \frac{Y_i}{N}$; $\hat{\mu} = \bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n}$ ∴ Estructura general: $\hat{\theta} \pm t_{\alpha/2, n-1} \cdot \sqrt{\widehat{Var}[\hat{\theta}]}$: $\hat{\theta} \pm Z_{\alpha/2} \cdot \sqrt{\widehat{Var}[\hat{\theta}]}$

- $\bar{Y} \pm Z_{\alpha/2} \cdot \sqrt{\widehat{Var}[\bar{Y}]}$

- $\tau = N\mu$: $\hat{\tau} = N\bar{Y}$

Parámetro a estimar	Varianza estimada del estimador	Intervalo de confianza del $(1-\alpha)100\%$
μ	$\hat{V}(\bar{Y}) = \left(\frac{N-n}{N}\right)\frac{S^2}{n}$	$\bar{Y} \pm t_{\alpha/2,n-1}\sqrt{\hat{V}(\bar{Y})}$

```
##################################
#    Función para estimación
##################################
estimar_parametro <- function(muestra, N, parametro, intervalo = FALSE, confianza = 0.95){
```

lo que se quiere

"tau", "mu"

μ	$\widehat{V}\left(\bar{Y}\right) = \left(\frac{N-n}{N}\right)\frac{S^2}{n}$	$\bar{Y} \pm t_{\alpha/2,n-1}\sqrt{\widehat{V}\left(\bar{Y}\right)}$
τ	$\widehat{V}\left(\hat{\tau}\right) = N^2\widehat{V}\left(\bar{Y}\right)$	$N\bar{Y} \pm t_{\alpha/2,n-1}N\sqrt{\widehat{V}\left(\bar{Y}\right)}$
p	$\widehat{V}\left(\hat{p}\right) = \left(\frac{N-n}{N}\right)\frac{\hat{p}\,\hat{q}}{n-1}$	$\hat{p} \pm t_{\alpha/2,n-1}\sqrt{\widehat{V}\left(\hat{p}\right)}$
A	$\widehat{V}\left(\hat{A}\right) = N^2\widehat{V}\left(\hat{p}\right)$	$N\hat{p} \pm t_{\alpha/2,n-1}N\sqrt{\widehat{V}\left(\hat{p}\right)}$

lo que se quiere muestrear

Tamaño población

"tau", "mu" "p", "A"

TRUE

se puede cambiar

Clase objeto: Almacenar resultados

```
> obj_tau <- estimar_parametro(ingreso_muestra, N, parametro = "tau", intervalo = TRUE, confianza = 0.95)
> obj_tau
$Estimador
[1] 170952118135

$Varianza
[1] 3.585226e+17

$IC
[1] 169778271008 172125965261
```

$\hat{\tau}$ $\widehat{Var}[\hat{\tau}]$ (1) (2)

• IC: Con una confianza del 95%, se puede determinar que el ingreso total de los hogares comprendidos en la ECV del año 2023 está dado entre (1) y (2) en COP.

```
> obj_mu <- estimar_parametro(ingreso_muestra, N, parametro = "mu", intervalo = TRUE, confianza = 0.95)
> obj_mu
$Estimador
[1] 2085774

$Varianza
[1] 53370599

$IC
[1] 2071452 2100096
```

$\bar{Y}$ $\widehat{Var}[\bar{Y}]$ (1) (2)

• Con una confianza del 95%, se puede determinar que el ingreso promedio de los hogares comprendidos en la ECV del año 2023 está dado entre (1) y (2) en COP.

$Y_i = \begin{cases} 1, \\ 0, \end{cases}$

$Y_i \sim Ber(p)$: $Var[Y_i] = p(1-p)$, $E[Y_i] = p$

3. Estime el total de hogares que poseen automovil A, así como la proporción de hogares que se perciben como seguros p. Calcule los correspondientes intervalos de confianza. Interprete estos resultados.

$\hat{p} = \sum_{i=1}^{n} \frac{Y_i}{n}$; $\hat{A} = \sum_{i=1}^{n} Y_i$

$\hat{\theta} \pm t_{\alpha/2,n-1}\cdot\sqrt{Var[\hat{\theta}]}$: $\hat{\theta} \pm Z_{\alpha/2}\cdot\sqrt{Var[\hat{\theta}]}$

```
> obj_p
$Estimador
[1] 0.1436

$Varianza
[1] 2.309997e-05

$IC
[1] 0.1341777 0.1530223
```

```
> obj_A
$Estimador
[1] 74994.32

$Varianza
[1] 98137.46

$IC
[1] 74380.17 75608.46
```

```
> carro_poblacion
# A tibble: 2 × 2
    CAR Total
  <int> <int>
1     0  6903
2     1 75058
```

```
> seguro_poblacion
# A tibble: 2 × 3
    SEC Total proporcion
  <int> <int>      <dbl>
1     0 69440      0.847
2     1 12521      0.153
```

4. Determine el tamaño de muestra mínimo necesario para estimar la proporción de personas que poseen propiedad recreativa con un límite para el error de estimación del 2% y una confianza del 95%.

Parámetro	Tamaño de muestra
p	$n = \frac{Np(1-p)}{(N-1)D+p(1-p)}$ $D = \frac{\delta^2}{Z^2_{\alpha/2}}$ Para N muy grande: $n_0 = \frac{p(1-p)}{D}$ $n_{\max} = \frac{1}{4D}$ (cuando $p = 1/2$)

Probabilidad

Margen error

Distribución normal estándar

• Vamos a suponer que p viene de una estimación de la población.

```
> # ------------------------------------------
> #        CUARTO PUNTO
> # ------------------------------------------
> objeto_prec <- estimar_parametro(muestra$REC, 81961, parametro = "p")
> p_rec <- objeto_prec$Estimador
> N <- 81961
> D <- ((0.02)^2/qnorm(0.05, lower.tail = FALSE)^2)
> n <- (N*p_rec * (1 - p_rec))/((N - 1) * D + p_rec * (1- p_rec))
> n
[1] 530.4565
```

n ; Nrel: 531

• El tamaño mínimo obtenido para la estimación de la proporción de hogares que poseen propiedad recreativa, con un 95% de confianza y un error de [illegible] es de 531 hogares

que poseen propiedad recreativa, con
un 95% de confianza y un
error de estimación del 2% es de 531 hogares.