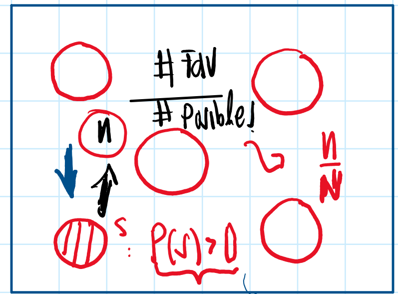
Primer taller

miércoles, 19 de febrero de 2025 1:57 p. m.

Parte teórica

De respuesta a las preguntas formuladas a continuación en base a la teoría tratada en clase. **Provea una interpretación de ser necesario.**

$N > n$



$U = N$

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. Verdadera ; s = muestra
   - (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. Falsa : (1). Diseño muestral; (2). Determinación med. (3). Trabajo campo; (4). Análisis est.
   - (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. Verdadero

     $\boxed{\text{Unidad: individuo u objeto de estudio}}$

DIRECTORIO	SECUENCIA_P	ORDEN.x	CAR	SEC	CONS	REC	I_HOGAR
7910114	1	1	1	1	1	1	2979593
7910115	1	1	1	0	1	0	1891852
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
7910127	1	1	1	1	1	1	2004751

   - (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. Verdadero
   - (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

     θ : $\boxed{\hat{\theta}\text{ (Máxima Verosimilitud)}}$

     ↳ $\boxed{\text{Tamaño muestral}}$
     - Nivel confianza
     - Error admisible
   - (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
     - Individuos más fáciles de muestrear
     - Gente Voluntario : sesgo de selección
   - (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. Falso

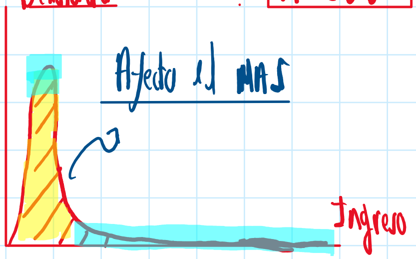
     σ^2 : ↳ $\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$: Añadir un factor de corrección por población finita

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

Parte práctica :

Densidad

$\boxed{N = 5000}$



Parámetro a estimar	Estimador puntual	Varianza del estimador
$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$	$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$	$V(\bar{Y}) = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}$
$\tau = \sum_{i=1}^{N} y_i = N\mu$	$\hat{\tau} = N\bar{Y}$	$V(\hat{\tau}) = N^2 V(\bar{Y})$
$p = \frac{1}{N}\sum_{i=1}^{N} y_i$	$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{a}{n}$	$V(\hat{p}) = \left(\frac{N-n}{N-1}\right)\frac{p(1-p)}{n}$
$A = \sum_{i=1}^{N} y_i = Np$	$\hat{A} = N\hat{p}$	$V(\hat{A}) = N^2 V(\hat{p})$

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

$_{i=1}^{n} y_i$ es el número de unidades en la muestra que ...buto de interés.

Donde $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2$ es la varianza muestral de la

- $\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$; y_i = Valor para el individuo i; $i = 1, \ldots, N$

  ↳ $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

"mu" ; "p" ; "A"

Donde $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2$ es la varianza muestral de la variable cuantitativa Y, y $t_{\alpha/2,n-1}$ es el cuantil $1-\alpha/2$ de la distribución t-Student con $n-1$ grados de libertad.

→ $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

"mu" ; "p" ; "A"

• estimar_parametro(muestra, N, parametro = "tau", ..., intervalo = TRUE, confianza = 0,95).

```
$Estimador
[1] 171650812280

$IC
[1] 170470649271 172830975289
```

$\hat{\tau}$ LI LS

$\hat{\tau} \pm t_{\alpha/2,n-1}\cdot\sqrt{\widehat{Var}[\hat{\tau}]}$ IC_1

• $\mu \to \hat{\mu} = \bar{Y}$; $\bar{Y} \pm t_{\alpha/2,n-1}\cdot\sqrt{Var[\bar{Y}]}$ IC_2

• $IC_1 = N\, IC_2$

• Con una confianza del 95%, se puede determinar que el ingreso total de las personas consideradas en la encuesta está comprendido entre LI y LS.

```
$Estimador
[1] 2094299

$IC
[1] 2079900 2108698
```

LI LS

$\hat{\mu} = \bar{Y}$

• Con una confianza del 95% se puede determinar que el ingreso promedio de los hogares considerados en la encuesta está comprendido entre LI, LS

```
> ingreso_promedio
[1] 2080171
```

μ ~ Directamente desde la encuesta (Población)

Variables cualitativas: $y_i = \begin{cases} 1, & \text{pertenece} \\ 0, & \text{no pertenece} \end{cases}$ → Ensayo Bernoulli

$Y_i \sim Ber(p)$; $E[Y_i] = p\ \forall i$; $Var[y_i] = p(1-p)\ \forall i$

↳ $\hat{p}$ ↳ $\hat{p}(1-\hat{p})$

; $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} y_i$ ~ Suma de 1 y 0

```
$Estimador
[1] 75256.59

$IC
[1] 74653.06 75860.12
```

$\hat{A}$ 75257

LI LS

CAR	Total	proporcion
0	6903	0.08422298
1	75058	0.91577702

↳ A

DIRECTORIO	SECUENCIA_P	ORDEN	CAR	SEC	CONS	REC	I_HOGAR
7910114	1	1	1	1	1	1	2979593
7910115	1	1	1	0	1	0	1891852
7910119	1	1	1	0	1	1	1372884
7910120	1	1	1	0	1	1	1288338
7910121	1	1	1	0	1	1	1608499

Go forward to the next source location (Ctrl+F10)

• Calcular proporción personas tienen carro:

$\hat{p} = \frac{1}{3}\cdot[1+1+1] = \frac{3}{3} = 1$ ~ 100% : $\underbrace{1-\hat{p} = 1-1 = 0}$

↳ Personas que no tienen carro

```
$Estimador
[1] 0.151

$IC
[1] 0.1413798 0.1606202
```

$\hat{p}$ → 15,1%

SEC	Total	proporcion
0	69440	0.8472322
1	12521	0.1527678

→ P 15,2%

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%.

$\hat{p} = 0,6$

N = 81961

Error admisible

$y_i \begin{cases} 1 \\ 0 \end{cases}$

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 $p = 1/2$)
A	$n = \frac{Np(1-p)}{(N-1)D+p(1-p)}$ $D = \frac{\delta^2}{Z^2_{\alpha/2}N^2}$

$\delta = 0,02$ $\alpha = 0,05$

$N = 81961$

$Z_{0,025} \approx 1,96$: qnorm(1 - α/2)

$\hat{p}$: Estudios anteriores

$Z_{0,05/2} \approx 1.96$: $\boxed{\text{qnorm}(1-\alpha/2)}$

A $\quad n = \frac{Np(1-p)}{(N-1)D+p(1-p)} \quad D = \frac{\delta^2}{Z_{\alpha/2}^2 N^2}$

$\hat{p}$:
- Estudios anteriores
- Opinión de expertos
- Otras fuentes.

: $\boxed{p = 1/2}$

```
> alpha <- 0.05
> D <- (0.02^2)/(qnorm(1- alpha/2)^2)
> n <- (N * 0.6 * (1 - 0.6))/((N - 1) * D + 0.6 * (1 - 0.6))
> n
[1] 2241.858
>
```

2242 individuos