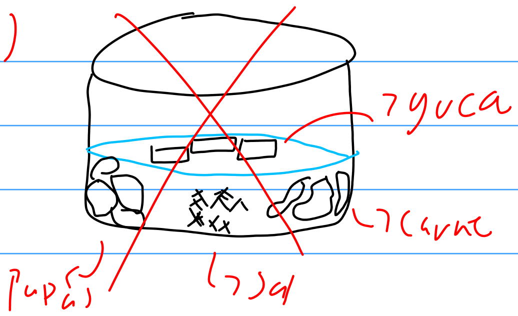
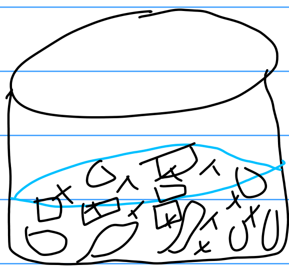
MUESTREAR ES COMO HACER UN BUEN

ESTAMOS HACIENDO UN SANCOCHO

Manera 1)



Manera 2)



hay de todo por todos lados

1)

a) Un censo consiste en seleccionar un subconjunto de la población, medir la característica de interés e inferir con esta información acerca de toda la población.

Falso, un censo es sobre toda la población.

b) La probabilidad de que una unidad cualesquiera esté presente en la muestra es $\frac{n}{N}$.

n: tamaño de muestra, N: tamaño de la población

$$\underset{1}{\_} \quad \underset{2}{\_} \quad \cdots \quad \underset{n-1}{\_} \quad \underset{n}{\_}$$

Una unidad cualesquiera tiene n casos favorables de ser seleccionada en la muestra

Los casos posibles son el tamaño de la poblacion

Entonces la probabilidad de que una unidad cualesquiera esté en una muestra es de n/N

c) Para N muy grande se tiene que $\mathbb{E}\left(S^2\right) \approx \sigma^2$.

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

Para una población finita $E(S^2) = \frac{N}{N-1} \sigma^2$

Si $N \to \infty$, $\frac{N}{N-1} \to 1$ Para N grande $\frac{N}{N-1} \simeq 1$

Para N grande $E(s^2) \approx \delta^2$

d) Suponga que se tiene un grupo de 11 personas númeradas del 1 al 11, de este grupo se desea extraer una muestra de tamaño 6. Para esto se decide tirar un par de dados 6 veces y en cada lanzamiento se resta una unidad al resultado de los dados, siendo dicho resultado final el número del individuo que va a ser incluido en la muestra. En caso de que se repita algún resultado los dados se arrojan nuevamente, por tanto como se quiere extraer una muestra y el resultado es aleatorio, se está ante el diseño muestral de muestreo aleatorio simple sin reemplazo.

Para el m.a.s cada unidad elemental tiene la misma probabilidad de ser seleccionada en la muestra

$P(\#1 \text{ sea seleccionado}) = \frac{1}{36}$ $\quad$ $P(\#6 \text{ sea seleccionado}) = \frac{6}{36}$

$\sum$ dados $= 2$
pasa si ambos dados caen en 1

$\sum$ dados $= 7$
pasa si
$(1,6);(6,1);(2,5);(5,2);$
$(3,4);(4,3)$

Como $P(\#1) \neq P(\#6)$ no estamos en m.a.s

2. Una m.a.s de 100 contadores de agua es controlada dentro de una comunidad para estimar el promedio de consumo de agua diario por casa durante un periodo seco. Realizado el estudio, se halló que la media y varianza muestrales fueron 12.5 y 1.252 galones respectivamente. Haga una estimación del promedio de consumo diario y calcule su respectivo intervalo de confianza, realice lo propio para el consumo total diario de la comunidad. Suponga que en la comunidad existe un total de 10000 casas. Determine cuantas unidades son necesarias para obtener un límite para el error de estimación de la media del consumo diario de 1 galón.

$\bar{Y} = 12.5$ y $S^2 = 1.252$
$N = 10000$ y $n = 100$

La estimación puntual del promedio es de 12.5 gal/día

Solo se conoce haciendo un censo

No se usa por lo general

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

Vamos a usar las estimaciones

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

Error estándar del estimador

lo que vamos a usar

I.C. μ: $\bar{y} \pm t_{1-\frac{\alpha}{2}, n-1}\sqrt{\hat{V}(\bar{y})}$

$$\hat{V}(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{S^2}{n} = \left(1-\underbrace{\frac{n}{N}}_{¿?}\right)\frac{S^2}{n} = \left(1-\frac{100}{10000}\right)\frac{1.252}{100} = 0.0123948$$

% unidades elementales no incluidas en la muestra

$\alpha = 0.05 \Rightarrow t_{1-\frac{\alpha}{2}, n-1} = 1.984217$

Así el IC va a ser $(12.27909, 12.72091)$

Para el total: basta con multiplicar todo por N

$\hat{\tau} = N\bar{y}$, $\hat{V}(\hat{\tau}) = \hat{V}(N\bar{y}) = N^2\hat{V}(\bar{y})$

IC τ: $\hat{\tau} \pm t_{1-\frac{\alpha}{2}, n-1}\hat{SE}(\hat{\tau}) = N\bar{y} \pm t_{1-\frac{\alpha}{2}, n-1}\sqrt{N^2\hat{V}(\bar{y})}$

$$= N\bar{y} \pm t_{1-\frac{\alpha}{2}, n-1} \cdot N \cdot \sqrt{\hat{V}(\bar{y})}$$

$$= N\underbrace{\left(\bar{y} \pm t_{1-\frac{\alpha}{2}, n-1}\sqrt{\hat{V}(\bar{y})}\right)}_{\text{IC para } \mu}$$

$\hat{\tau} = N \cdot \bar{Y} = 725\,000$ gal/día

$IC_\tau = N \, IC_\mu \Rightarrow (722790.9,\ 727209.1)$

Parámetro	Tamaño de muestra usando σ^2	Tamaño de muestra usando $\widetilde{\sigma}^2$
μ $\delta = \lvert\theta - \hat{\theta}\rvert$	$n = \frac{N\sigma^2}{(N-1)D+\sigma^2}$ $D = \frac{\delta^2}{Z^2_{\alpha/2}}$ Para N muy grande: $n_0 = \frac{\sigma^2}{D}$	$n = \frac{N\widetilde{\sigma}^2}{ND+\widetilde{\sigma}^2}$ $D = \frac{\delta^2}{Z^2_{\alpha/2}}$ Para N muy grande: $n_0 = \frac{\widetilde{\sigma}^2}{D}$
τ	$n = \frac{N\sigma^2}{(N-1)D+\sigma^2}$ $D = \frac{\delta^2}{Z^2_{\alpha/2}N^2}$	$n = \frac{N\widetilde{\sigma}^2}{ND+\widetilde{\sigma}^2}$ $D = \frac{\delta^2}{Z^2_{\alpha/2}N^2}$

Parámetro	Tamaño de muestra usando σ^2	Tamaño de muestra usando $\widetilde{\sigma}^2$
μ y τ $\varepsilon = \left\lvert\frac{\hat{\theta} - \theta}{\theta}\right\rvert$ $= \frac{\delta}{\lvert\theta\rvert}$	$n = \frac{N\,CV^2}{(N-1)D+CV^2}$ $D = \frac{\varepsilon^2}{Z^2_{\alpha/2}}$ Para N muy grande: $n_0 = \frac{CV^2}{D}$	$n = \frac{N\,\widetilde{CV}^2}{ND+\widetilde{CV}^2}$ $D = \frac{\varepsilon^2}{Z^2_{\alpha/2}}$ Para N muy grande: $n_0 = \frac{\widetilde{CV}^2}{D}$

Para ε: $n = \dfrac{N\,CV^2}{(N-1)D + CV^2}$; $CV = \dfrac{\sigma}{\mu}$, $D = \dfrac{\delta^2/\mu^2}{Z^2_{1-\frac{\alpha}{2}}}$

$$= \frac{N \cdot \sigma^2/\cancel{\mu^2}}{(N-1)\cdot\frac{\delta^2}{\cancel{\mu^2} Z^2_{1-\frac{\alpha}{2}}} + \sigma^2/\cancel{\mu^2}} = \frac{N\sigma^2}{(N-1)\frac{\delta^2}{Z^2_{1-\frac{\alpha}{2}}} + \sigma^2}$$

Observación: Todo es lo mismo, lo único que cambia es D

Como σ^2 es desconocido, se usa la estimación

$\hat{\sigma}^2 = \left(\frac{N-1}{N}\right) S^2 = 1.25787 \qquad z_{1-\alpha/2} \approx 1.96$

Reemplazando se tiene que $n \approx 4.08 \Rightarrow n = 5$

mínimo

3. París es una ciudad que recibe diariamente 1500 turistas. Se desea realizar un estudio y se ha seleccionado una muestra aleatoria simple de turistas, donde se les preguntó cuanto gastan diariamente y si eran extranjeros.

Cuadro 1: Datos de los turistas

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Gastos	48	41	34	25	32	25	36	31	30	38	31	19	26	27	22
Nacionalidad	0	1	1	0	1	0	1	1	0	1	1	0	1	1	1

a. Estime los gastos totales en consumo que realizan los turistas en París.
b. Estime el total de turistas que són extranjeros. Estime un intervalo de confianza.
c. Determine el tamaño de muestra mínimo necesario para estimar la propoción de extranjeros que visitan París en un día con un límite para el error de estimación de 2 % y una confianza de 95 %

a) Nos piden $\hat{\tau} = N\bar{Y}$ donde $N = 1500$ y $\bar{Y} = 31 \Rightarrow \hat{\tau} = 46500$ EUR

b) Vamos a estimar la proporción de turistas que son extranjeros

$$\hat{p} = \frac{a}{n} = \frac{10}{15} \approx 0.6667$$

Un intervalo de confianza para la proporción de extranjeros va a ser

$$\hat{p} \pm t_{1-\frac{\alpha}{2}, n-1} \underbrace{\widehat{SE}(\hat{p})}_{\sqrt{\hat{V}(\hat{p})}}, \quad \hat{V}(\hat{p}) = \left(1 - \frac{n}{N}\right) \frac{\hat{p}(1-\hat{p})}{n-1} \quad \text{y } \alpha = 0.05$$

Reemplazando se tiene que el IC es $(0.3978, 0.9355)$

Para A basta multiplicar por N

$\hat{A} = N\hat{p} = 1000$ y IC A es $(596, 1404)$

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

δ

Parámetro	Tamaño de muestra
p y A	$n = \frac{Np(1-p)}{(N-1)Dp^2+p(1-p)} = \frac{N\,CV^2}{(N-1)D+CV^2} \approx \frac{CV^2}{D+\frac{1}{N}CV^2}$ con: $D = \frac{\varepsilon^2}{Z^2_{\alpha/2}}$ y $CV = \sqrt{\frac{1-p}{p}}$ Para N muy grande: $n_0 = \frac{CV^2}{D}$

ε

$Var(X) = p(1-p)$ si $X \sim Bernoulli(p)$

Usar $\hat{p}$ o usar $p = 0.5$ (máxima variabilidad)

$$n = \frac{N\hat{p}(1-\hat{p})}{(N-1)D + \hat{p}(1-\hat{p})} \quad , \quad D = \frac{\delta^2}{z^2_{1-\alpha/2}}$$

$$\simeq 881.114 \Rightarrow n = 882$$