7 ELEMENTOS DE MUESTREO

CONTENIDOS:

7.1. Muestreo aleatorio simple.
7.2 Muestreo aleatorio estratificado.
7.3 Muestreo aleatorio de conglomerados.
7.4 Estimación del tamaño poblacional.

OBJETIVOS:

- Determinar el diseño de muestreo adecuado para un problema específico.
- Determinar el tamaño de muestra para un diseño de muestreo específico en un problema práctico.

7.1 MUESTREO ALEATORIO SIMPLE.

CONCEPTOS CLAVES: Parámetro. Estimador. Varianza del estimador. Error de estimación. Nivel de confianza. Tamaño de muestra.

RESUMEN DE CONCEPTOS Y PROPIEDADES:

Población: Colección de elementos acerca de los cuales se desea hacer inferencia.
Unidades de muestreo: Son elementos o colecciones de elementos de la población con posibilidades de entregar información.
Marco Muestral: Es una lista de unidades muestrales.
Muestra: Es una colección de elementos seleccionados de un marco muestral.
Diseño Muestral: Método de selección de la muestra.

Si $\hat{\theta}$ es el estimador de un parámetro θ y se supone que $\hat{\theta} \sim N(\theta, V(\hat{\theta}))$ entonces si se desea cometer un error de estimación no mayor a d_0 con una confianza $1-\alpha$ la inecuación base a resolver para determinar el tamaño de muestra necesario es $z_{1-\frac{\alpha}{2}}\sqrt{V(\hat{\theta})} \leq d_0$, donde la varianza del estimador $V(\hat{\theta})$ cambia según el diseño muestral usado y el parámetro que se desea estimar.

1.- Tamaño de muestra para estimar una media de la población usando un diseño MAS (Muestreo Aleatorio Simple)

$$n \geq \begin{cases} n_0 = \left(\dfrac{z_{1-\frac{\alpha}{2}}}{d_0}\right)^2 S_X^2 & si \quad \dfrac{n_0}{N} \leq 0.05 \\[2ex] n' = \dfrac{n_0}{1+\dfrac{n_0}{N}} & si \quad \dfrac{n_0}{N} > 0.05 \end{cases}$$

2.- Tamaño de muestra para estimar una proporción de la población usando un diseño MAS (Muestreo Aleatorio Simple)

$$n \geq \begin{cases} n_0 = \left(\dfrac{z_{1-\frac{\alpha}{2}}}{d_0}\right)^2 \hat{p}(1-\hat{p}) & si \quad \dfrac{n_0}{N} \leq 0.05 \\[2ex] n' = \dfrac{n_0}{1+\dfrac{n_0}{N}} & si \quad \dfrac{n_0}{N} > 0.05 \end{cases}$$

EJERCICIO RESUELTO, PASO A PASO:

Ejercicio 1: (Aplicación en Ciencias de la salud)

Para analizar el crecimiento de ratas de laboratorio se elige una muestra piloto de 13 ratas y se miden obteniendo una talla promedio de la muestra de 5.3 centímetros y una varianza muestral de 19.2

Pregunta: Un investigador desea determinar el tamaño de muestra mínimo usando un diseño MAS para estimar la talla promedio de las ratas en la población con una confianza de 95% y un error de estimación no mayor a 2 centímetros si en la población hay 200 ratas. ¿Cuántas ratas se debe elegir en la muestra?

Esquema de solución

Paso 1: Leer cuidadosamente el enunciado del problema.

Paso 2: Identificar la variable en estudio y los parámetros involucrados.

Sea X = Talla de las ratas (en centímetros). En este caso se debe suponer que $X \sim N(\mu,\sigma^2)$ y los parámetro involucrados son μ y σ^2 donde μ es la talla promedio de las ratas en la población y σ es la desviación estándar de la talla de las ratas.

Paso 3: Estimar los parámetros. En este caso se tiene, del enunciado del problema, que

$\bar{X} = 5.3$ es un estimador de μ y $S_x^2 = 19.2$ es un estimador de σ^2

Paso 4: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido. Para responder la pregunta se debe usar un diseño MAS para determinar el tamaño de muestra mínimo para estimar una media poblacional.

Paso 5: Determinación del tamaño de muestra.

$$n \geq n_0 = \left(\frac{z_{0.975}}{d_0}\right)^2 S_X^2 = \left(\frac{1.96}{2}\right)^2 19.2 = 18.4397 \approx 19$$

$$\frac{n_0}{N} = \frac{19}{200} = 0.095 > 0.05 \Rightarrow n \geq n' = \frac{n_0}{1+\frac{n_0}{N}} = \frac{19}{1+0.095} = 17.3516 \approx 18 .$$

Paso 6: Redactar una respuesta.

Para estimar la talla media de las ratas en la población usando un diseño MAS con una confianza de 95% y un error de estimación no mayor a 2 centímetros el investigador debe elegir a lo menos 18 ratas en la muestra

EJERCICIO RESUELTO, PASO A PASO:

EJERCICIO 2: (APLICACIÓN EN CIENCIAS DE LA INGENIERÍA)

En una fábrica de artículos electrónicos generalmente el 10% de los artículos presenta algún defecto de fabricación. Para analizar la calidad del producto se desea estimar la proporción de artículos electrónicos defectuosos de un lote de 2000 artículos listo para ser embarcado.
Pregunta: ¿Cuántos artículos deben ser elegidos del lote si se desea una confianza de 95% y un error de estimación no mayor a 0.05?

Esquema de solución

Paso 1: Leer cuidadosamente el enunciado del problema.

Paso 2: Identificar la característica en estudio y los parámetros involucrados.
Sea A = {Artículos electrónicos de la fábrica con algún defecto}.
En este caso el parámetro de interés es $p = P(A) =$ proporción de artículos electrónicos defectuosos en la fábrica.

Paso 3: Estimar los parámetros. En este caso se tiene, del enunciado del problema, que
$\hat{p} = 0.1$ (10%) es un estimador de p

Paso 4: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido. Para responder la pregunta se debe usar un diseño MAS para determinar el tamaño de muestra mínimo para estimar una proporción poblacional.

Paso 5: Determinación del tamaño de muestra.

$$n \geq n_0 = \left(\frac{z_{0.975}}{d_0}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.05}\right)^2 0.1 * 0.9 = 138.2976 \approx 139$$

$$\frac{n_0}{N} = \frac{139}{2000} = 0.695 > 0.05 \Rightarrow n \geq n' = \frac{n_0}{1+\frac{n_0}{N}} = \frac{139}{1+0.695} = 82.0059 \approx 83 .$$

Paso 6: Redactar una respuesta.

Para estimar la proporción de artículos electrónicos en el lote usando un diseño MAS con una confianza de 95% y un error de estimación no mayor a 0.05 se debe elegir a lo menos 83 artículos del lote en la muestra

EJERCICIOS PROPUESTOS:

1. (**Aplicación en Ciencias de la Salud**) Con el propósito de verificar la efectividad de un tratamiento basado en ejercicios para el aumento de la talla en niños de 10 años (en cms.), se elige una muestra piloto de 29 niños de un colegio que tiene 300 niños y se realizó el experimento aplicando el tratamiento a 13 niños y dejando 16 como controles, el que arrojó los siguientes resultados:

   TRATADOS $n = 13$, $\overline{X} = 138.6$, $S_X^2 = 29.16$

   CONTROLES $m = 16$, $\overline{Y} = 125.8$, $S_Y^2 = 20.7$

   i) Si se desea estimar la talla promedio de los niños a los cuales se les debe aplicar el tratamiento con una confianza de 95% y un error no superior a 2 cms.. ¿Cuántos niños se debe elegir como mínimo del colegio para aplicarles el tratamiento?
   ii) Si se desea estimar la talla promedio de los niños a los cuales se les debe dejar como controles con una confianza de 95% y un error no superior a 2.3 cms.. ¿Cuántos niños se debe elegir del colegio para dejarlos como controles?

2. (**Aplicación en Ciencias de la Salud**) Los siguientes datos representan el aumento de peso (en Kilos.) mediante dos dietas alternativas en una muestra piloto de niños atendidos en el consultorio norte que atiende 500 niños en la ciudad de Antofagasta:

Dieta 1	0.98	0.62	0.53	0.76	0.43	0.36	0.51	0.24	0.24
Dieta 2	0.40	0.14	0.16	0.21	0.09	0.02	0.02	0.01	

   a) Si se desea estimar el aumento de peso promedio en los niños con la dieta 1 con una confiabilidad de 95% y un error no mayor a 0.1 Kilos.. ¿Cuántos niños se debe elegir del consultorio?
   b) Si se desea estimar el aumento de peso promedio en los niños con la dieta 2 con una confiabilidad de 95% y un error no mayor a 0.12 Kilos.. ¿Cuántos niños se debe elegir del consultorio?

3. **(Aplicación en Ciencias del Mar)** La siguiente tabla muestra la distribución de una muestra aleatoria de 400 truchas cafés de un gran río., según la longitud y el sector donde fueron extraídas.

	SECTOR DEL RIO		
LONGITUD	Alto	Centro	Bajo
Bajo el promedio	67	64	25
Promedio	42	76	56
Sobre el promedio	10	23	37

   a) Si se desea estimar la proporción de truchas de longitud bajo el promedio en el sector alto del río, usando una confianza de 95% y un error no mayor a 0.05 ¿Cuántas truchas se debe extraer del sector alto del río?.
   b) Si se desea estimar la proporción de truchas del sector alto del río cuya longitud está bajo el promedio con una confianza de 95% y un error de estimación no mayor a 0.05 ¿Cuántas truchas con longitud bajo el promedio se debe elegir?

7.2 MUESTREO ALEATORIO ESTRATIFICADO.

CONCEPTOS CLAVES: Parámetro. Estimador. Varianza del estimador. Error de estimación. Nivel de confianza. Tamaño de muestra.

RESUMEN DE CONCEPTOS Y PROPIEDADES:

Si una población de tamaño N está particionada en h subpoblaciones o estratos de tamaño $N_1, N_2, ..., N_h$ y se elige una muestra aleatoria simple de cada estrato de tamaño $n_1, n_2, ..., n_h$ tal que $\sum_{i=1}^{h} n_i = n$, entonces el diseño muestral así estructurado se denomina **Muestreo Aleatorio Estratificado (MAE).**

Una vez determinado el tamaño de muestra *n* este puede ser asignado a cada estrato por uno de los siguientes métodos

1.- Asignación Proporcional: $n_i = n\, W_i$ donde $W_i = \frac{N_i}{N} \quad \forall i = 1, 2, ..., h$

2.- Asignación Óptima:

$$n_i = n \frac{N_i S_i}{\sum_{j=1}^{h} N_j S_j} \quad \forall i = 1, 2, ..., h$$

1.1 Tamaño de muestra para estimar una media de la población usando un diseño MAE con asignación proporcional

$$n \geq \begin{cases} n_0 = \left(\dfrac{z_{1-\frac{\alpha}{2}}}{d_0} \right)^2 \displaystyle\sum_{i=1}^{h} W_i S_i^2 & si \quad \dfrac{n_0}{N} \leq 0.05 \\ n' = \dfrac{n_0}{1 + \dfrac{n_0}{N}} & si \quad \dfrac{n_0}{N} > 0.05 \end{cases}$$

1.2 Tamaño de muestra para estimar una media de la población usando un diseño MAE con asignación óptima

$$n \geq \begin{cases} n_0 = \left(\dfrac{z_{1-\frac{\alpha}{2}}}{d_0} \right)^2 \left(\displaystyle\sum_{i=1}^{h} W_i S_i \right)^2 & si \quad \dfrac{n_0}{N} \leq 0.05 \\ n' = \dfrac{n_0}{1 + \dfrac{n_0}{N}} & si \quad \dfrac{n_0}{N} > 0.05 \end{cases}$$

2.1 Tamaño de muestra para estimar una proporción de la población usando un diseño MAE con asignación proporcional

$$n \geq \begin{cases} n_0 = \left(\dfrac{z_{1-\frac{\alpha}{2}}}{d_0} \right)^2 \displaystyle\sum_{i=1}^{h} W_i \hat{p}_i (1 - \hat{p}_i) & si \quad \dfrac{n_0}{N} \leq 0.05 \\ n' = \dfrac{n_0}{1 + \dfrac{n_0}{N}} & si \quad \dfrac{n_0}{N} > 0.05 \end{cases}$$

EJERCICIO RESUELTO, PASO A PASO:

Ejercicio 1: (Aplicación en Ciencias del mar)

Se desea estimar la talla media de salmones y la proporción de salmones que cumplen la norma para el consumo en un cultivo de 310 salmones distribuidos en tres estanques con la siguiente información.

Estanque	N_i	S_i^2	$\hat{p}_i$
1	155	25	0.80
2	62	225	0.25
3	93	100	0.50
Total	310		

a) ¿Cuántos salmones se debe elegir en la muestra del cultivo y por estanque para estimar la talla media si se desea una confianza de 95% y un error no mayor a 2 centímetros? Si usamos un diseño MAE con i) Asignación proporcional ii) Asignación óptima

b) ¿Cuántos salmones se debe elegir en la muestra del cultivo y por estanque para estimar la proporción de salmones que cumple la norma si se desea una confianza de 95% y un error no mayor a 0.05? Si usamos un diseño MAE con Asignación proporcional

Esquema de solución

a)

Paso 1: Leer cuidadosamente el enunciado del problema.

Paso 2: Identificar la variable en estudio y los parámetros involucrados.

Sea X = Talla de los salmones (en centímetros). En este caso se debe suponer que $X \sim N(\mu, \sigma^2)$ y el parámetro involucrado es μ donde μ es la talla media de los salmones en el cultivo

Paso 3: Estimar los parámetros. En este caso se tiene, del enunciado del problema, que $S_1^2 = 25 \,; S_2^2 = 225 \,; S_3^2 = 100$ son los estimadores de las varianzas por estrato

Paso 4: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido. Para responder la pregunta a) se debe usar un diseño MAE en la determinación del tamaño de muestra mínimo para estimar una media poblacional usando asignación proporcional y óptima

Paso 5: Determinación del tamaño de muestra.

i) Usando asignación proporcional

$$n \geq n_0 = \left(\frac{1.96}{2}\right)^2 \left(\frac{155}{310}25 + \frac{62}{310}225 + \frac{93}{310}100\right) = 84.035 \approx 85$$

$$\frac{n_0}{N} = \frac{85}{310} = 0.2742 > 0.05 \Rightarrow n \geq n' = \frac{n_0}{1+\frac{n_0}{N}} = \frac{85}{1+0.2742} = 66.7085 \approx 67$$

$$n_1 = 67 * \frac{155}{310} = 33.5 \approx 34;\ n_2 = 67 * \frac{62}{310} = 13.4 \approx 13;\ n_3 = 67 * \frac{93}{310} = 20.1 \approx 20\,.$$

ii) Usando asignación óptima

$$n \geq n_0 = \left(\frac{1.96}{2}\right)^2 \left(\frac{155}{310}5 + \frac{62}{310}15 + \frac{93}{310}10\right)^2 = 69.3889 \approx 70$$

$$\frac{n_0}{N}=\frac{70}{310}=0.2258>0.05\Rightarrow n\geq n'=\frac{n_0}{1+\frac{n_0}{N}}=\frac{70}{1+0.22258}=57.105\approx 58$$

$$n_1=58*\frac{155*5}{2635}=17.06\approx 17;\, n_2=58*\frac{62*15}{2635}=20.47\approx 21;\, n_3=67*\frac{93*10}{2635}=20.47\approx 20\,.$$

Paso 6: Redactar una respuesta.

a) i) Para estimar la talla media de los salmones en el cultivo usando un diseño MAE con asignación proporcional con una confianza de 95% y un error de estimación no mayor a 2 centímetros se debe elegir a lo menos 67 salmones en la muestra de los cuales 34 deben ser elegidos del estanque 1, 13 del estanque 2 y 20 del estanque 3

a) ii) Para estimar la talla media de los salmones en el cultivo usando un diseño MAE con asignación óptima con una confianza de 95% y un error de estimación no mayor a 2 centímetros se debe elegir a lo menos 58 salmones en la muestra de los cuales 17 deben ser elegidos del estanque 1, 21 del estanque 2 y 20 del estanque 3

b)

Paso 1: Leer cuidadosamente el enunciado del problema.

Paso 2: Identificar la característica en estudio y los parámetros involucrados.

Sea A = {Salmones del cultivo que cumplen la norma para el consumo}.

En este caso el parámetro de interés es $p=P(A)$ = proporción de salmones que cumplen la norma para el consumo.

Paso 3: Estimar los parámetros. En este caso se tiene, del enunciado del problema, que $\hat{p}_1=0.80;\quad \hat{p}_2=0.25;\quad \hat{p}_3=0.50$ son las proporciones estimadas por estanque

Paso 4: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido. Para responder la pregunta se debe usar un diseño MAE para determinar el tamaño de muestra mínimo para estimar una proporción poblacional.

Paso 5: Determinación del tamaño de muestra.

Usando asignación proporcional

$$n\geq n_0=\left(\frac{1.96}{0.05}\right)^2\left(\frac{155}{310}0.8*0.2+\frac{62}{310}0.25*0.75+\frac{93}{310}0.5*0.5\right)=295.8032\approx 296$$

$$\frac{n_0}{N}=\frac{296}{310}=0.9548>0.05\Rightarrow n\geq n'=\frac{n_0}{1+\frac{n_0}{N}}=\frac{296}{1+0.9548}=151.4221\approx 152$$

$$n_1=152*\frac{155}{310}=76;\, n_2=152*\frac{62}{310}=30.4\approx 30;\, n_3=152*\frac{93}{310}=45.6\approx 46\,.$$

. **Paso 6: Redactar una respuesta.**

Para estimar la proporción de salmones que cumplen la norma para el consumo en el cultivo, usando un diseño MAE con asignación proporcional con una confianza de 95% y un error de estimación no mayor a 0.05 se debe elegir a lo menos 152 salmones en la muestra de los cuales 76 deben ser elegidos del estanque 1, 30 del estanque 2 y 46 del estanque 3

EJERCICIOS PROPUESTOS:

1. (**Aplicación en Ciencias de la Salud**) En la ciudad de Antofagasta se desea estimar la presión sistólica media de los pacientes atendidos en cuatro consultorios de los cuales se obtuvo la siguiente información:

Consultorio	N_i	s_i^2
1	1000	2410
2	3000	2938
3	2000	2047
4	1000	2214

a) Determine el tamaño de muestra adecuado para estimar la presión sistólica media en la ciudad con una confianza de 95% y un error no superior a 5 mm/Hg usando muestreo aleatorio estratificado con :
   i) Asignación proporcional
   ii) Asignación óptima

b) Determine el número de pacientes a elegir por consultorio en cada caso.

2. (**Aplicación en Ciencias del Mar**) Los siguientes datos corresponden a la varianza del peso de gónadas de ostiones obtenidos de un cultivo distribuido en tres sectores costeros

Sector	N_h	s_h^2
1	12473	1.96
2	35241	1.21
3	23178	3.24
Total	70892	

a) Si se desea estimar el peso promedio de las gónadas en el cultivo completo con una confianza de 95% y un error no mayor a 0.5 gramos. ¿Cuántas ostiones se deben seleccionar del cultivo si se usa:
   i) Muestreo aleatorio estratificado con asignación proporcional?
   ii) Muestreo aleatorio estratificado con asignación óptima?

b) Asigne el número de ostiones por sector para cada caso de la parte a).

3. **(Aplicación en Ciencias de la Salud)** Se desea estimar la proporción de sujetos infectados con cierto virus para lo cual se estratifica la población en tres regiones y los datos obtenidos fueron los siguientes:

	Región 1	Región 2	Región 3
Nº de sujetos	240	432	316
Tamaño muestra	50	80	60
Nº sujetos infectados	14	34	29

a) Determine el tamaño de muestra adecuado para estimar la proporción de sujetos infectados con una confiabilidad de 95% y un error de 0.05 usando muestreo aleatorio estratificado con asignación proporcional

b) Determine el número de sujetos a seleccionar por región en cada caso

7.3 MUESTREO ALEATORIO DE CONGLOMERADOS.

CONCEPTOS CLAVES: Parámetro. Estimador. Varianza del estimador. Error de estimación. Nivel de confianza. Tamaño de muestra. Conglomerado

RESUMEN DE CONCEPTOS Y PROPIEDADES:

Si una población de tamaño N está formada por colecciones de elementos o conglomerados, donde cada conglomerado es considerado una unidad muestral y se selecciona una muestra aleatoria de n conglomerados entonces el diseño muestral así estructurado se denomina **Muestreo Aleatorio de Conglomerados (MAC).**

Sea N = nº de conglomerados en la población

n = nº de conglomerados en la muestra

m_i = Tamaño del conglomerado i

$M = \sum_{i=1}^{N} m_i$ = nº de elementos en la población

$\bar{M} = \frac{M}{N}$ = Tamaño promedio de los conglomerados

x_{ij} = Dato del j-ésimo elemento en el conglomerado i

$x_i = \sum_{i=1}^{m_i} x_{ij}$ = Total de la variable en el conglomerado i

a_i = Total de elementos que poseen una cierta característica en el cong. i

$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} m_i}$ = Estimador de la media poblacional

$\hat{p} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}$ = Estimador de la proporción poblacional

1.- Tamaño de muestra para estimar una media usando un diseño MAC

$$n \geq \begin{cases} n_0 = \left(\dfrac{z_{1-\frac{\alpha}{2}}}{d_0} \right)^2 \dfrac{1}{\bar{M}^2} \dfrac{\sum_{i=1}^{n"} (x_i - \hat{\mu}\, m_i)^2}{n"-1} & si \quad \dfrac{n_0}{N} \leq 0.05 \\ n' = \dfrac{n_0}{1 + \dfrac{n_0}{N}} & si \quad \dfrac{n_0}{N} > 0.05 \end{cases}$$

Donde $\sum_{i=1}^{n"} (x_i - \hat{\mu}\, m_i)^2 = \sum_{i=1}^{n"} x_i^2 - 2\hat{\mu} \sum_{i=1}^{n"} x_i m_i + \hat{\mu}^2 \sum_{i=1}^{n"} m_i^2$ y n" = Tamaño muestra piloto

2.- Tamaño de muestra para estimar una proporción usando un diseño MAC

$$n \geq \begin{cases} n_0 = \left(\dfrac{z_{1-\frac{\alpha}{2}}}{d_0} \right)^2 \dfrac{1}{\bar{M}^2} \dfrac{\sum_{i=1}^{n"} (a_i - \hat{\mu}\, m_i)^2}{n"-1} & si \quad \dfrac{n_0}{N} \leq 0.05 \\ n' = \dfrac{n_0}{1 + \dfrac{n_0}{N}} & si \quad \dfrac{n_0}{N} > 0.05 \end{cases}$$

Donde $\sum_{i=1}^{n"} (a_i - \hat{p}\, m_i)^2 = \sum_{i=1}^{n"} a_i^2 - 2\hat{p} \sum_{i=1}^{n"} a_i m_i + \hat{p}^2 \sum_{i=1}^{n"} m_i^2$ y n" = Tamaño muestra piloto

EJERCICIO RESUELTO, PASO A PASO:

Ejercicio 1: (Aplicación en Ciencias del mar)

Se desea estimar el peso medio de jaibas y la proporción de jaibas que cumplen la norma para el consumo en un cultivo que consta de 50 jaulas. Una muestra piloto de 5 jaulas entrega la siguiente información.

Jaula	Nº de jaibas (m_i)	Peso total (x_i)	Jaibas que cumplen norma (a_i)
1	80	960	40
2	120	1210	70
3	40	420	10
4	50	650	30
5	60	520	30

a) ¿Cuántas jaulas se debe elegir en la muestra del cultivo para estimar el peso medio de las jaibas si se desea una confianza de 95% y un error no mayor a 0.5 gramos? Si usamos un diseño MAC

b) ¿Cuántas jaulas se debe elegir en la muestra del cultivo para estimar la proporción de jaibas que cumple la norma para el consumo si se desea una confianza de 95% y un error no mayor a 0.05? Si usamos un diseño MAC

Esquema de solución a)

Paso 1: Leer cuidadosamente el enunciado del problema.

Paso 2: Identificar la variable en estudio y los parámetros involucrados.

Sea X = Peso de las jaibas (en gramos). En este caso se debe suponer que $X \sim N(\mu,\sigma^2)$ y el parámetro involucrado es μ donde μ es el peso medio de las jaibas en el cultivo

Paso 3: Estimar los parámetros. En este caso se debe ingresar los datos (x, y) en la calculadora considerando $x = m_i$ e $y = x_i$ obteniendo los siguientes resultados

$$\sum m_i = 350; \quad \sum m_i^2 = 28500; \quad \sum x_i = 3760; \quad \sum x_i^2 = 3255000; \quad \sum m_i x_i = 302500$$

luego $\hat{\mu} = \dfrac{\sum x_i}{\sum m_i} = \dfrac{3760}{35} = 107.429$ es el estimador del peso medio y además se tiene que

$$\sum_{i=1}^{n"} (x_i - \hat{\mu}\, m_i)^2 = \sum_{i=1}^{n"} x_i^2 - 2\hat{\mu}\sum_{i=1}^{n"} x_i m_i + \hat{\mu}^2 \sum_{i=1}^{n"} m_i^2$$

$$= 3255000 - 2*107.420*302500 + (107.420)^2 * 28500 = 44727.66$$

Paso 4: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido. Para responder la pregunta a) se debe usar un diseño MAE en la determinación del tamaño de muestra mínimo para estimar una media poblacional usando asignación proporcional y óptima

Paso 5: Determinación del tamaño de muestra.

$$n \ge n_0 = \left(\frac{1.96}{0.5}\right)^2 \frac{1}{4900} \frac{44727.66}{4} = 35.066 \approx 36$$

$$\frac{n_0}{N} = \frac{36}{50} = 0.72 > 0.05 \Rightarrow n \ge n' = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{36}{1 + 0.72} = 20.93 \approx 21$$

. **Paso 6: Redactar una respuesta.**

Para estimar el peso medio de las jaibas en el cultivo usando un diseño MAC con una confianza de 95% y un error de estimación no mayor a 0.5 gramos se debe elegir a lo menos 21 jaulas en la muestra.

Esquema de solución pregunta b)

Paso 1: Leer cuidadosamente el enunciado del problema.

Paso 2: Identificar la variable en estudio y los parámetros involucrados.
Sea A = {Jaibas del cultivo que cumplen la norma para el consumo}.
En este caso el parámetro de interés es $p = P(A) =$ proporción de jaibas que cumplen la norma para el consumo.

Paso 3: Estimar los parámetros. En este caso se debe ingresar los datos (x, y) en la calculadora considerando $x = m_i$ e $y = a_i$ obteniendo los siguientes resultados $\sum m_i = 350$; $\sum m_i^2 = 28500$; $\sum a_i = 180$; $\sum a_i^2 = 8400$; $\sum m_i a_i = 15300$

luego $\hat{p} = \frac{\sum a_i}{\sum m_i} = \frac{180}{350} = 0.5143$ es el estimador del peso medio y además se tiene que

$$\sum_{i=1}^{n"}\left(a_i - \hat{p}\, m_i\right)^2 = \sum_{i=1}^{n"} a_i^2 - 2\hat{p}\sum_{i=1}^{n"} a_i m_i + \hat{p}^2 \sum_{i=1}^{n"} m_i^2$$
$$= 8400 - 2 * 0.5143 * 15300 + (0.5143)^2 * 28500 = 200.798$$

Paso 4: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido. Para responder la pregunta a) se debe usar un diseño MAE en la determinación del tamaño de muestra mínimo para estimar una media poblacional usando asignación proporcional y óptima

Paso 5: Determinación del tamaño de muestra.

$$n \geq n_0 = \left(\frac{1.96}{0.05}\right)^2 \frac{1}{4900} \frac{200.798}{4} = 15.7425 \approx 16$$

$$\frac{n_0}{N} = \frac{16}{50} = 0.32 > 0.05 \Rightarrow n \geq n' = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{16}{1 + 0.32} = 12.12 \approx 13$$

. **Paso 6: Redactar una respuesta.**
Para estimar la proporción de jaibas que cumple la norma para el consumo en el cultivo usando un diseño MAC con una confianza de 95% y un error de estimación no mayor a 0.05 se debe elegir a lo menos 13 jaulas en la muestra.

EJERCICIOS PROPUESTOS:

1. (**Aplicación en Ciencias del mar**) En un cultivo de pulpos dividido en 50 estanques, se selecciona una muestra preliminar de 5 estanques y se obtiene la siguiente información respecto a la talla total de los pulpos por estanque y el número de pulpos con una cierta bacteria:

Estanque	N° de pulpos	Talla Total (cms.)	Nº pulpos con bacteria
1	42	248	10
2	51	295.8	20
3	49	328.3	15
4	55	269.5	25
5	60	378	18

a) Si se desea estimar la talla promedio de los pulpos en el cultivo con una confianza de 95% y un error no mayor a 0.5 centímetros. ¿Cuántos estanques se debe seleccionar en la muestra?

b) Si se desea estimar la proporción de pulpos con bacteria en el cultivo con una confianza de 95% y un error no mayor a 0.05 ¿Cuántos estanques se debe seleccionar en la muestra?

2. (**Aplicación en Ciencias de la salud)** En una zona residencial existen 60 manzanas de las cuales se selecciona una muestra preliminar de 5 y se obtiene la siguiente información respecto al ingreso total por manzana y la cantidad de residentes con una edad mayor o igual a 65 años

Manzana	N° de residentes	Ingreso Total (Miles de pesos.)	Edad ≥ 65 años
1	90	960	10
2	32	1210	20
3	47	420	15
4	25	650	25
5	16	520	18

a) Si se desea estimar ingreso promedio de los residentes con una confianza de 95% y un error no mayor a 0.5. ¿Cuántas manzanas se debe seleccionar en la muestra?

b) Si se desea estimar la proporción de residentes con una edad mayor o igual a 65 años con una confianza de 95% y un error no mayor a 0.05 ¿Cuántas manzanas se debe seleccionar en la muestra?

3. (**Aplicación en Ciencias del mar**) Se diseña un muestreo para estimar la cantidad promedio gastada en alimentos por jaula en un criadero de jaibas. Ya que no se encuentra identificadas las jaulas, se usa un muestreo de conglomerados considerando una hectárea como conglomerado. Para lo cual se selecciona una muestra piloto de 6 hectáreas del criadero de un total de 60 y se obtiene el costo en alimentos de cada jaula dentro de la hectárea seleccionada. Los costos totales y el número de jaulas por hectárea se muestran en la siguiente tabla :

Hectárea	**Nº de jaulas**	**Gasto total en alimentos (Dólares)**
1	55	2210
2	63	2430
3	78	3110
4	52	1990
5	64	2470
6	75	2870

Determine el número de hectáreas necesarias a seleccionar en la muestra para estimar el gasto promedio en alimentos para las jaulas del criadero si se desea una confiabilidad de 95% y un error en la estimación no superior a 0.4 dólares.

7. 4 ESTIMACIÓN DEL TAMAÑO POBLACIONAL.

CONCEPTOS CLAVES: Parámetro. Estimador. Varianza del estimador. Error de estimación. Nivel de confianza. Tamaño de muestra

RESUMEN DE CONCEPTOS Y PROPIEDADES:

1.- Muestreo Directo

Se tiene una población de tamaño N desconocido y para estimarlo se elige una muestra aleatoria de *t* elementos, se marcan y se devuelven a la población.

De esta forma $p = \frac{t}{N}$ es la proporción de marcados en la población, luego

Parámetro: $N = \frac{t}{p}$

Estimador: Se elige una segunda muestra de tamaño *n* en la cual se encuentran *s* marcados, luego $\hat{p} = \frac{s}{n}$ es la proporción estimada de marcados en la segunda muestra, así un estimador puntual del tamaño de la población es $\hat{N} = \frac{t}{\hat{p}} = \frac{nt}{s}$

Varianza estimada: $V(\hat{N}) = \frac{t^2 n(n-s)}{s^3}$

Un intervalo de confianza de nivel $(1-\alpha)$ para el tamaño de la población está dado por

$$\left[\hat{N} - z_{1-\frac{\alpha}{2}}\sqrt{V(\hat{N})};\ \hat{N} + z_{1-\frac{\alpha}{2}}\sqrt{V(\hat{N})}\right]$$

2.- Muestreo Inverso

Se selecciona una muestra de *t* elementos de la población, se marcan y se regresan a la población. Luego se realiza una selección aleatoria hasta obtener exactamente *s* elementos marcados. Si la segunda muestra es necesario seleccionar *n* elementos para obtener los *s* marcados, entonces $\hat{p} = \frac{s}{n}$ es una estimación de la proporción de marcados en la población

(Note que en este caso *s* es fijo y *n* es aleatorio)

Estimador: $\hat{N} = \frac{t}{\hat{p}} = \frac{nt}{s}$

Varianza estimada: $V(\hat{N}) = \frac{t^2 n(n-s)}{s^2(s+1)}$

En este caso un intervalo de confianza de nivel $(1-\alpha)$ para el tamaño de la población está dado por

$$\left[\hat{N} - z_{1-\frac{\alpha}{2}}\sqrt{V(\hat{N})};\ \hat{N} + z_{1-\frac{\alpha}{2}}\sqrt{V(\hat{N})}\right]$$

EJERCICIO RESUELTO, PASO A PASO:

Ejercicio 1: Antes de abrir la temporada de caza se desea estimar la población de venados, para lo cual se captura una muestra de 300 venados, se marcan y se regresan al bosque. Dos semanas después se eligen 200 venados, de los cuales 62 venían marcados. Estime el total de venados en el bosque y determine un intervalo de 95% de confianza para dicho parámetro.

Paso 1: Leer cuidadosamente el enunciado del problema.

Paso 2: Identificar los parámetros involucrados

En este caso el parámetro a estimar es el número de venados en la población

Paso 3: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido

Para responder la pregunta se debe usar muestreo directo

Paso 4: Estimar los parámetros

En este caso se tiene $t = 300$; $n = 200$; $s = 62$

Luego el número estimado de venados en el bosque es

$$\widehat{N} = \frac{200*300}{62} = 967.74 \approx 968 \qquad V(\widehat{N}) = \frac{(300)^2(200)(138)}{(62)^3} = 10422.61086$$

$$d = 1.96\sqrt{10422.61086} = 200.098 \approx 200$$

Luego un intervalo de 95% de confianza para el total de venados en el bosque es $[968-200;968+200] = [768;1168]$

Paso 5: Redactar una respuesta a la pregunta:

Con 95% de confianza el número de venados en la población se encuentra entre 768 y 1168 venados

Ejercicio 2: En una gran reservación de animales es de interés estimar el número total de pájaros de una determinada especie que allí habitan, para lo cual se atrapa una muestra inicial de 150 pájaros, se marcan y luego se dejan libres. En el mismo mes se capturan pájaros hasta encontrar 35 marcados, logrando capturar 100 pájaros para lograr esto. Estime el número total de pájaros en la reservación y un intervalo de 95% de confianza para dicho total.

Paso 1: Leer cuidadosamente el enunciado del problema.

Paso 2: Identificar los parámetros involucrados

En este caso el parámetro a estimar es el número de pájaros en la población

Paso 3: Leer la pregunta y revisar cual de los conceptos se debe usar para obtener lo pedido

Para responder la pregunta se debe usar muestreo inverso

Paso 4: Estimar los parámetros

En este caso $t = 150$; $n = 100$ y $s = 35$

Luego el número estimado de pájaros en la reservación es

$$\widehat{N} = \frac{100*150}{35} = 428.57 \approx 429 \qquad V(\widehat{N}) = \frac{(150)^2(100)(65)}{(35)^2 36} = 3316.3265$$

$$d = 1.96\sqrt{3316.3265} = 112.87 \approx 113$$

Luego un intervalo de 95% de confianza para el total de pájaros en la reservación es

$$[429-113;429+113] = [316;542]$$

Paso 5: Redactar una respuesta a la pregunta:

Con 95% de confianza el número de pájaros en la población se encuentra entre 316 y 542 pájaros.

EJERCICIOS PROPUESTOS:

1. Ciertos biólogos de poblaciones salvajes desean estimar el tamaño total de la población de codorniz común en una región del sur del país. Se usa una serie de 50 trampas. En la primera muestra se atrapan 320 codornices. Después de ser capturadas, cada ave es retirada de la trampa y marcada con una banda de metal en su pata izquierda. Luego se sueltan las aves. Varios meses más tarde se obtiene una segunda muestra de 515 codornices de las cuales 91 estaban marcadas. Estime el número total de codornices en la región y determine con una confiabilidad de 95% entre qué valores se encuentra el verdadero total.

2. Los expertos en recursos animales de cierta reservación desean conocer la población de conejos cuya disminución es evidente. En un estudio realizado hace dos años, el tamaño de la población resultó ser de 2500 conejos. Supóngase qué éste sigue siendo de la misma magnitud. Usando el cuadro correspondiente determine los tamaños de muestras de captura y recaptura para estimar el número total de conejos en la población con un límite para el error igual a 356 conejos.

3. Una zoóloga desea estimar el tamaño de la población de tortugas en una determinada área geográfica. Ella cree que el tamaño de la población está entre 500 y 1000 tortugas, por lo que una muestra inicial de 100 parece ser suficiente. Las 100 tortugas son capturadas, marcadas y liberadas. Toma una segunda muestra un mes después y decide seguir muestreando hasta obtener 15 tortugas marcadas, lo que se logra al recapturar 160 tortugas. Estime el número total de tortugas en el área y establezca con un 99% de confianza entre qué valores se encuentra el verdadero total.