MUESTREO EN POBLACIONES FINITAS (1)

- Dos aspectos básicos de la inferencia estadística, no vistos aún:
    - Proceso de selección de la muestra → Métodos de muestreo
    - Tamaño adecuado en poblaciones finitas → Fiabilidad y coste

- ETAPAS EN UN ESTUDIO DE MUESTREO:

1. Definir la información que se necesita → fundamental *versus* accesorio
2. Determinar correctamente la población objeto del estudio → listado
3. Método de muestreo a seguir y tamaño de la muestra:
    3.1 El método depende del problema y de los recursos disponibles
    3.2 El tamaño depende de la fiabilidad requerida y del coste
4. Diseño adecuado de la forma de obtener la información. Objeto:
    4.2 Evitar falta de respuesta → forma encuesta, nº preguntas
    4.3 Respuestas honestas y precisas → cuestionario y entrevista
5. Uso de la muestra para hacer inferencia
6. Obtener conclusiones acerca de la población

MUESTREO EN POBLACIONES FINITAS (2)

- TIPOS DE ERRORES:
  - Debidos al muestreo → incertidumbre (nivel significación, etc.)
  - Ajenos al muestreo:
    1. Definición incorrecta de la población
    2. Respuestas falsas o imprecisas
    3. Falta de respuesta → posible sesgo
    4. Sesgo en la selección elementos muestrales
    5. Errores de manipulación, tabulación y cálculo

No hay un criterio general para evitarlos y/o analizarlos → minimizarlos

MUESTREO EN POBLACIONES FINITAS (3)

- MÉTODOS DE MUESTREO:

  - Muestreo aleatorio:

    a) unidad muestral elemental:
       a.1) muestreo aleatorio simple
       a.2) muestreo aleatorio sistemático
       a.3) muestreo aleatorio estratificado

    b) unidad muestral grupo:
       b.1) muestreo por áreas y conglomerados
       b.2) muestreo por etapas

  - Muestreo no aleatorio y semialeatorio (en general, no “científico”; no estudia precisión):
    - por cuotas
    - opinático o de intención

MÉTODOS DE MUESTREO (1)

- MUESTREO ALEATORIO SIMPLE:

  - Sirve de base a los demás métodos
  - Es el más sencillo desde el punto de vista teórico

  - Todos los elementos muestrales se tratan como iguales y se identifican mediante un número (tarjeta, bola, números aleatorios, etc...)

Elemento muestral	Identificador
A	1
B	2
..	..

  - La selección es sin reposición
  - Todas las muestras posibles son igualmente probables

  - Cuando N es muy grande su coste es muy alto

MÉTODOS DE MUESTREO (2)

- MUESTREO ALEATORIO SISTEMÁTICO:
  - Se necesita un listado ordenado de los elementos de la población
  - El orden no debe ser un factor distorsionante de la aleatoriedad:
    - No distorsionante: listas de clase para notas (no sesgo)
    - Sí puede generar sesgo: producción mensual empresa
  - Se selecciona al azar el primer elemento muestral (k) menor que $p=N/n$
  - Elegido este, los demás se obtienen sumándole p al anterior: $k+p$, $k+2p$,...
  - El método garantiza que aparezcan elementos de todas las clases, por lo que puede generar muestras más representativas que el muestreo aleatorio simple

MÉTODOS DE MUESTREO (3)

- MUESTREO ESTRATIFICADO:
  - En ocasiones es indispensable agrupar los elementos de la población en clases o estratos (homogeneidad de sus elementos; heterogeneidad entre estratos) → mejor información, reduce errores y costes.
  - Dentro de cada estrato se aplicará un muestreo aleatorio simple o sistemático

- MUESTREO POR CONGLOMERADOS:
  - Conglomerado: es un grupo de elementos de la población (familias, hogares, casas, edificios, municipios, provincias, empresas, etc.)
  - La unidad de muestreo es el conglomerado → a veces, áreas geográficas
  - Se seleccionan aleatoriamente cierto número de conglomerados y se investigan, a continuación, todos los elementos pertenecientes a ellos (EPA, Encuesta Presupuestos Familiares,..)
  - Características: homogeneidad entre conglomerados; heterogeneidad dentro de cada conglomerado → representar las clases de la población
  - Se reduce problema de listado, no es necesario saber tamaño población, entrevistas dentro del grupo (conglomerado)→ menos costoso

MÉTODOS DE MUESTREO (4)

- MUESTREO ALEATORIO POR ETAPAS:
  - Generalización del muestreo por conglomerados
  - Suele hacerse descendiendo de conglomerados más grandes a más pequeños:

    Provincia → Municipio → Barrio → Edificio → Familia (listados)
  - En cada etapa se aplica el muestreo aleatorio, sistemático o estratificado
  - Objetivo: Reducir al mínimo el coste

INFERENCIA CON MUESTREO ALEATORIO SIMPLE (1)

- INFERENCIA SOBRE LA MEDIA (μ):

  - Estimación por puntos: $\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$

  - Estimación por intervalos: $\boxed{\mu \in (\bar{x} \pm z_{\alpha/2}\hat{\sigma}_{\bar{x}})}$, por desconocerse $\sigma_{\bar{x}}^2$

    1. En poblaciones finitas: $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ → factor de corrección
    2. Como σ^2 es desconocida se estima mediante su estimador insesgado que es $\hat{s}^2 \frac{N-1}{N}$ y, por tanto, $\boxed{\hat{\sigma}_{\bar{x}}^2 = \frac{\hat{s}^2}{n}\frac{N-n}{N}}$
    3. Para utilizar la normal, n será suficientemente grande.
    4. Si n es pequeña y se supone normalidad → t de *Student*

INFERENCIA CON MUESTREO ALEATORIO SIMPLE (2)

- INFERENCIA SOBRE EL TOTAL ($N\mu$):
  - Estimación por puntos: $\boxed{N\bar{x}}$
  - Estimación por intervalos: $\boxed{N\mu \in (N\bar{x} \pm z_{\alpha/2} N\hat{\sigma}_{\bar{x}})}$

$$1.\ Var(N\bar{x}) = N^2\sigma^2_{\bar{x}} \rightarrow N^2\hat{\sigma}^2_{\bar{x}} = N^2 \frac{\hat{s}^2}{n}\frac{N-n}{N} = \frac{\hat{s}^2}{n} N(N-n)$$

- INFERENCIA SOBRE LA PROPORCIÓN (p):
  - Estimación por puntos: $\boxed{\hat{p} = \frac{x}{n}}$, x=nº observaciones característica en n
  - Estimación por intervalos: $\boxed{p \in (\hat{p} \pm z_{\alpha/2}\hat{\sigma}_{\hat{p}})}$

$$1.\ \boxed{\hat{\sigma}^2_{\hat{p}} = \frac{\hat{p}(1-\hat{p})}{n-1}\frac{N-n}{N}}$$

TAMAÑO MUESTRAL CON MUESTREO ALEATORIO SIMPLE (1)

- Para la estimación de la MEDIA:

  - Al dar $\bar{x}$ por μ, el error máximo permitido, para un nivel de confianza del 100(1-α)%, será: $\boxed{\varepsilon = |z_{\alpha/2}|\sigma_{\bar{x}}}$ |<-----ε-----$\boxed{\mu}$-----ε----->|
  - Fijado este error y el nivel de significación, se fija, también, la varianza máxima del estadístico muestral: $\boxed{\sigma_{\bar{x}} = \dfrac{\varepsilon}{|z_{\alpha/2}|}}$ → Recordemos: $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n}\dfrac{N-n}{N-1}$
  - Despejando de esta última expresión (o del cuadrado de la primera):

$$\boxed{n = \frac{N\sigma^2}{(N-1)\sigma_{\bar{x}}^2 + \sigma^2} = \frac{Nz_{\alpha/2}^2\sigma^2}{(N-1)\varepsilon^2 + z_{\alpha/2}^2\sigma^2}}$$ → σ por encuesta piloto o anterior

TAMAÑO MUESTRAL CON MUESTREO ALEATORIO SIMPLE (2)

- Para la estimación del TOTAL (va a ser igual que para la media):
  - Recordemos que $Var(N\bar{x}) = N^2\sigma^2_{\bar{x}}$ → $N^2\sigma^2_{\bar{x}} = N^2 \frac{\sigma^2}{n}\frac{N-n}{N-1}$. Se llega al mismo resultado: $\boxed{n = \frac{N\sigma^2}{(N-1)\sigma^2_{\bar{x}} + \sigma^2} = \frac{N z^2_{\alpha/2}\sigma^2}{(N-1)\varepsilon^2 + z^2_{\alpha/2}\sigma^2}}$
- Para la estimación de la PROPORCIÓN:
  - En poblaciones finitas: $\boxed{\sigma^2_{\hat{p}} = \frac{pq}{n}\frac{N-n}{N-1}}$. Despejando, se obtiene: $\boxed{n = \frac{Npq}{(N-1)\sigma^2_{\hat{p}} + pq} = \frac{N z^2_{\alpha/2} pq}{(N-1)\varepsilon^2 + z^2_{\alpha/2} pq}}$. Como p no se conoce, se estima o se calcula el tamaño muestral máximo → $\boxed{n_{max} = \frac{0{,}25N}{(N-1)\sigma^2_{\hat{p}} + 0{,}25}, \text{ con } \sigma^2_{\hat{p}} = \frac{\varepsilon^2}{z^2_{\alpha/2}}}$

INFERENCIA CON MUESTREO ALEATORIO ESTRATIFICADO (1)

- INFERENCIA SOBRE LA MEDIA:
  - Población dividida en K estratos: $N_1 + N_2 + \ldots\ldots + N_K = N$
  - Tamaños muestrales de los estratos: $n_1 + n_2 + \ldots\ldots + n_K = n$
  - Medias poblacionales de los estratos: $\mu_1 \quad \mu_2 \;\ldots\ldots\ldots\; \mu_K$
  - Medias muestrales de los estratos: $\bar{x}_1 \quad \bar{x}_2 \quad \ldots\ldots \quad \bar{x}_K$
- Puesto que en cada estrato se hace un muestreo aleatorio simple:
  - Estimadores insesgados de las medias poblacionales (μ_i): $\bar{x}_i$
  - Estimadores insesgados de la variancia de $\bar{x}_i$: $\hat{\sigma}^2_{\bar{x}_i} = \dfrac{\hat{s}_i^2}{n_i}\dfrac{N_i - n_i}{N_i}$
- Estimación por puntos de $\mu = \dfrac{1}{N}\sum_{i=1}^{K} N_i \mu_i \rightarrow \boxed{\bar{x} = \dfrac{1}{N}\sum_{i=1}^{K} N_i \bar{x}_i}$
- Estimación por intervalos: $\boxed{\mu \in (\bar{x} \pm z_{\alpha/2}\hat{\sigma}_{\bar{x}})}$, con $\boxed{\hat{\sigma}^2_{\bar{x}} = \dfrac{1}{N^2}\sum_{i=1}^{K} N_i^2 \hat{\sigma}^2_{\bar{x}_i}}$

INFERENCIA CON MUESTREO ALEATORIO ESTRATIFICADO (2)

- INFERENCIA SOBRE EL TOTAL:
  - Estimación por puntos de $N\mu = \sum_{i=1}^{K} N_i \mu_i$: $\boxed{N\bar{x} = \sum_{i=1}^{K} N_i \bar{x}_i}$
  - Estimación por intervalos: $\boxed{N\mu \in (N\bar{x} \pm z_{\alpha/2} \hat{\sigma}_{N\bar{x}})}$

    Con $N\bar{x} = \sum_{i=1}^{K} N_i \bar{x}_i$ y $\hat{\sigma}^2_{N\bar{x}} = \sum_{i=1}^{K} N_i^2 \hat{\sigma}^2_{\bar{x}_i}$
- INFERENCIA SOBRE LA PROPORCIÓN:
  - Proporciones poblacionales de los estratos: $p_1 \quad p_2 \ \ldots\ldots\ldots \quad p_k$
  - Proporciones muestrales de los estratos: $\hat{p}_1 \quad \hat{p}_2 \ \ldots\ldots \quad \hat{p}_K$
  - Estimación por puntos de $p = \frac{1}{N} \sum_{i=1}^{K} N_i p_i$: $\boxed{\hat{p} = \frac{1}{N} \sum_{i=1}^{K} N_i \hat{p}_i}$
  - Estimación por intervalos: $\boxed{p \in (\hat{p} \pm z_{\alpha/2} \hat{\sigma}_{\hat{p}})}$

    Con $\hat{\sigma}^2_{\hat{p}} = \frac{1}{N^2} \sum_{i=1}^{K} N_i^2 \hat{\sigma}^2_{\hat{p}_i}$ y $\hat{\sigma}^2_{\hat{p}_i} = \frac{\hat{p}_i (1 - \hat{p}_i)}{n_i - 1} \frac{N_i - n_i}{N_i}$

INFERENCIA CON MUESTREO ALEATORIO ESTRATIFICADO (3)

- DISTRIBUCIÓN DE LA MUESTRA ENTRE ESTRATOS:
  - No hay una respuesta única; depende de los objetivos de la encuesta
  - Criterios de asignación (*afijación*):
    1. Uniforme: todos igual; poco sentido real.
    2. Proporcional: La proporción de elementos de la población en cada estrato se aplica a la muestra:

$$\frac{N_i}{N} = \frac{n_i}{n} \rightarrow \boxed{n_i = \frac{N_i}{N} n}$$

    3. Óptima: Pondera el criterio anterior con las varianzas de los respectivos estratos, asignando más observaciones a los estratos con mayor variancia poblacional. Es el más deseable si el objetivo único es la precisión en la estimación:

Media y Total: $\boxed{n_i = \dfrac{N_i \sigma_i}{\sum_{i=1}^{k} N_i \sigma_i} n}$; Proporción: $\boxed{n_i = \dfrac{N_i \sqrt{p_i q_i}}{\sum_{i=1}^{k} N_i \sqrt{p_i q_i}} n}$

(al ser σ desconocida, muestreo preliminar y n_{max})

TAMAÑO MUESTRAL CON MUESTREO ALEATORIO ESTRATIFICADO (1)

- MEDIA Y TOTAL:
  - Asig. Proporcional ($n_i = \frac{N_i}{N} n$):

$$\boxed{n = \frac{\sum_{i=1}^{K} N_i \sigma_i^2}{N\sigma_{\bar{x}}^2 + \frac{1}{N}\sum_{i=1}^{K} N_i \sigma_i^2} \text{; con } \sigma_{\bar{x}}^2 = \frac{\varepsilon^2}{z_{\alpha/2}^2}}$$

  - Asig. Óptima ($n_i = \frac{N_i \sigma_i}{\sum_{i=1}^{k} N_i \sigma_i} n$):

$$\boxed{n = \frac{\frac{1}{N}\left(\sum_{i=1}^{K} N_i \sigma_i\right)^2}{N\sigma_{\bar{x}}^2 + \frac{1}{N}\sum_{i=1}^{K} N_i \sigma_i^2} \text{; con } \sigma_{\bar{x}}^2 = \frac{\varepsilon^2}{z_{\alpha/2}^2}}$$

TAMAÑO MUESTRAL CON MUESTREO ALEATORIO ESTRATIFICADO (2)

- PROPORCIÓN:
  - Asig. Proporcional ($n_i = \frac{N_i}{N} n$):

$$\boxed{n = \frac{\sum_{i=1}^{K} N_i p_i q_i}{N\sigma_{\hat{p}}^2 + \frac{1}{N}\sum_{i=1}^{K} N_i p_i q_i}\text{; con } \sigma_{\hat{p}}^2 = \frac{\varepsilon^2}{z_{\alpha/2}^2}}$$

  - Asig. Óptima ($n_i = \frac{N_i \sigma_i}{\sum_{i=1}^{k} N_i \sigma_i} n$):

$$\boxed{n = \frac{\frac{1}{N}\left(\sum_{i=1}^{K} N_i \sqrt{p_i q_i}\right)^2}{N\sigma_{\hat{p}}^2 + \frac{1}{N}\sum_{i=1}^{K} N_i p_i q}\text{; con } \sigma_{\hat{p}}^2 = \frac{\varepsilon^2}{z_{\alpha/2}^2}}$$