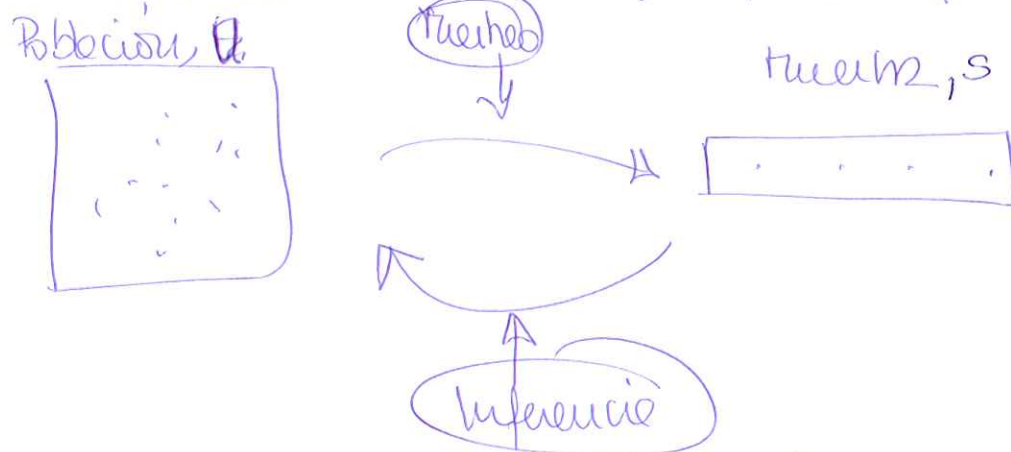


0 - Introducción

CONCEPTO DE MUESTRA

Para dejar claro el concepto de muestreo hay que empezar conociendo los fundamentos de la Inferencia Estadística.



1- Tenemos un problema en la población (sobre el valor de una característica poblac. o cualquier otra cosa),

• Población \rightarrow Grp. ^{elementos} ~~individuos~~ que se pretende estudiar.
y por falta de dinero, tiempo, etc. nos limitamos a estudiar una parte de dicha población, muestra,

• Muestra \rightarrow Grp. de elementos ~~que~~ ^{que} fue realmente estudiados

con el objetivo de generalizar o inferir los resultados obtenidos en la muestra a toda la población.

Al proceso de selección de los elementos poblac. que forman parte de la muestra se le denomina muestreo, mientras que al proceso de generalización de resultados se le denomina inferencia.

1- CONCEPTO DE MUESTREO PROBABILÍSTICO

Notese la importancia de la muestra en la inferencia, en la población, puesto que de ella dependen las conclusiones sobre la población.

Por tanto, lo más importante que hay que pedirle a una muestra es que sea representativa de la población a la que pertenece, es decir, que sea representativa (No solamente m.p. inferir en inferencia referida a la muestra).

- Representativa: que represente bien, que refleje las características propias y la heterogeneidad presente en la población de partida, (que sea una poble. a escala pequeña)

En general, la manera de seleccionar la muestra se puede dividir en dos grandes grupos:

Muestreo	<	no probabilístico	→	de inferir	<ul style="list-style-type: none"> • Probab. desconocida • No error
		probabilístico	→	de inferir	

En el muestreo no probabilístico la probab. de que una unidad sea seleccionada en el muestreo es desconocida, por lo que no se puede medir el grado de ~~de~~ error ni se puede dar la distrib. de frec. de las estimaciones (¿las estimaciones?).

En el muestreo probabilístico puede establecerse la probab. de obtener cada uno de las muestras posibles, ya que se considera la selección de muestras un fenómeno aleatorio probabilizable, y por tanto se puede medir el error de muestreo. → TIRAR PAP SOTE.

⊗ En el proceso inferencial los resultados obtenidos de la muestra difieren de los verdaderos valores poblacionales.

El muestreo probabilístico garantiza que las diferencias se deben exclusivamente al azar (haber tomado una muestra en vez de otra) y se puede cuantificar el ^(error de muestreo) error cometido a partir de estudiar la probab. de extracción de la muestra.

Muestreo no probabilístico:

- Opinitico \rightarrow expertos
- aplicando criterio \rightarrow criterio personal
- cuotas \rightarrow libertad, pero respetando las cuotas
- sin cuotas \rightarrow a voleo

Muestreo probabilístico:

< con probab. =
≠

< con reemplaz
sin reemplaz.

Muestreo probabilístico: Puede establecerse la probab. de obtener cada uno de las muestras que sea posible seleccionar (elementos de S , espacio muestral), mediante un procedimiento de muestreo dado, cuando la selección de la muestra sea un fenómeno aleatorio probabilístico. La incertidumbre derivada es susceptible de medirse, por lo que se puede medir/valorar el error de muestreo.

Proce

Inferencia estadística: Metodología consistente en inferir resultados, predicciones y generalizaciones sobre la población, basándose en la información contenida en las muestras representativas previamente elegidas por mt. formal de muestreo. Está basado en la Tª de la Probabilidad & puede ser ~~consta~~

Métodos de muestreo: Conjunto de técnicas estadísticas que estudian la forma de seleccionar una muestra lo sufic. representativa de la población, cuya información permite inferir las propiedades o características de la población cometiendo un error medible y acotable.

~~(A) partir de~~

Formalmente, el ~~espacio~~ muestreo probabilístico se describe de la siguiente manera:

Sea $U = \{u_1, \dots, u_N\}$ población (finita)

" $S = \{s_1, \dots, s_n\}$ cito de sucesos elementales, espacio muestral
 $s_i = \text{muestra } i$.

Sobre S se define

$$s_i = \{u_{i1}, \dots, u_{in}\} \begin{cases} i \rightarrow \text{muestra } i \\ j \rightarrow \text{poblac. } j \end{cases}$$

Sobre S se define una medida de probab P q:

$$P: S \rightarrow [0, 1]$$

$$P(s_i) \geq 0$$

$$\sum_{s \in S} P(s_i) = 1$$

$$P(s_i) = P(u_{i1}) \cdot P(u_{i2}/u_{i1}) \cdot \dots \cdot P(u_{in}/u_{i1} \dots u_{i(n-1)})$$

La terna así definida constituye un espacio de probabilidad

$$(U, S, P)$$

2 - DISTRIB. de un ESTIMADOR en el MUESTREO

El muestreo probabilístico garantiza que la muestra obtenida por el mt. de muestreo elegido sea una variable aleatoria, de la cual se puede deducir su distrib. de probabilidad (siempre y cuando la selección sea un fenómeno aleatorio probabilizable).

A partir de la información muestral se construyen estimadores, que son estadísticos ^{que no dependen de ningún parámetro del} (funciones de la muestra) cuya misión es dar un valor aproximado del parámetro poblacional desconocido.

Los estimadores, por tanto, son v.a., por haber variabilidad entre muestras. El muestreo probabilístico garantiza que se puede medir el error de muestreo.

En general, ~~para~~ un parámetro poblacional de valor desconocido θ , se estima mediante una función muestral denominada estimador, $\hat{\theta}$. Al valor concreto que toma $\hat{\theta}$ por cada muestra concreta se le denomina estimación.

La distrib. de probabilidad de la v.a. $\hat{\theta}$ se obtiene asignando probabilidades a todos los valores posibles del estimador, de manera que conocida la probab. asociada a cada uno de estos valores \neq queda constituida su distrib. de probab.

Sea T el qto de todos los valores posibles del estimador,

$$T = \{t \in \mathbb{R} / \exists (x_1, \dots, x_n) \in S(x) \text{ t.q. } \hat{\theta}(x_1, \dots, x_n) = t\}.$$

definimos su probab. asociada como la suma de las probab. de que ocurran todas las muestras que den el valor concreto del estimador.

$$P(\hat{\theta}(x_1, \dots, x_n) = t) = \sum_{\{S_i / \hat{\theta}(S_i) = t\}} P(S_i)$$

Al par $\{T, P^T\}$ se le denomina distrib. del estimador en el muestreo.

$$\begin{aligned} \hat{\theta}: S(x) &\longrightarrow \mathbb{R} \\ s &\longmapsto \hat{\theta}(x_1, \dots, x_n) = t \end{aligned}$$

$$P(\hat{\theta} = t) = \sum_{\{S_i / \hat{\theta}(S_i) = t\}} P(S_i)$$

Para cada elemento poblacional, podemos observar una característica cuantitativa X ó bien una característica cualitativa

$$U_i \begin{cases} \text{cuantit. } X & \left| \begin{aligned} \theta &= X \\ \theta &= \bar{X} = \frac{\sum_{i=1}^N X_i}{N} \end{aligned} \right. \\ \text{cualit. } A & \left| \begin{aligned} \theta &= A \\ \theta &= P = \frac{\sum_{i=1}^N A_i}{N} \end{aligned} \right. \end{cases}$$

3. ERROR CUADRÁTICO MEDIO Y SUS COMPONENTES

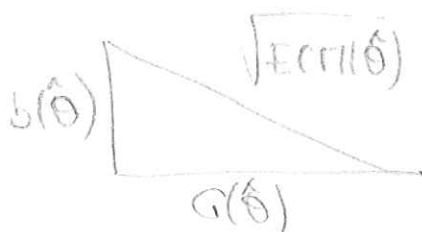
La precisión de un estimador se analiza ~~fundamentalmente~~ ^{esencialmente} en función de los conceptos de:

- error de muestreo $\rightarrow \sigma(\hat{\theta}) = +\sqrt{V(\hat{\theta})} \equiv$ concentración de las estimaciones alrededor de su valor medio.
- acuracidad $\rightarrow ECM(\hat{\theta}) = E[\hat{\theta} - \theta]^2 \equiv$ concentración de las estimaciones alrededor del verdadero valor del parámetro.
- sesgo $\rightarrow b(\hat{\theta}) = E[\hat{\theta}] - \theta \equiv$ distancia entre el valor esperado del estimador y el verdadero valor del parámetro.

Estas magnitudes pueden relacionarse a partir de la descomposición del $ECM(\hat{\theta}) = Var(\hat{\theta}) + \text{sesgo}^2$

$$\begin{aligned} ECM(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] \xrightarrow{\text{desarr. el cuadrado}} E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] \xrightarrow{\text{tomando exp.}} E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 \\ &= \underbrace{(E[\hat{\theta}^2] - E[\hat{\theta}]^2)}_{\substack{\text{varianza y reducida} \\ d_2 - d_1^2 \\ Var''(\hat{\theta})}} + \underbrace{(E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2)}_{\substack{\text{cuadrado diferencia} \\ (\theta - E[\hat{\theta}])^2 \\ \text{al revés}}} = Var(\hat{\theta}) + b(\hat{\theta})^2 \end{aligned}$$

Geométricamente,



la contribución del sesgo y de la derivación típica al ECM viene dado por la

$$\left| \tan \alpha = \frac{b(\hat{\theta})}{\sigma(\hat{\theta})} \right|$$

Desearse $\rightarrow |\tan \alpha| < 1/10$, pero

pero si $b(\hat{\theta}) \downarrow$, $\sigma(\hat{\theta}) \uparrow$, por lo que se acepta se admite estimadores siempre si $\frac{|b(\hat{\theta})|}{|\sigma(\hat{\theta})|} \leq 0.10$

4 - MT's de SELECCIÓN, PROBABILIDAD de la UNIDAD de PERTENECER a la muestra y PROPIEDADES

Métodos de selección \rightarrow Son cuatro, según dos criterios.

Reposición $\begin{cases} \text{CON} \\ \text{SIN} \end{cases}$

Probabilidad de E muestra $\begin{cases} \text{IGUALES} \\ \text{DISTINTAS} \end{cases}$

Además, se puede considerar la distinción

ORDEN $\begin{cases} \text{NO importa} \\ \text{SÍ importa} \end{cases}$

SIN reposición \rightarrow las muestras con algún elemento repetido son imposibles.

Todos los elementos de la muestra son \neq

SR $\begin{cases} \text{NO importa orden, } \#S = C_{N,n} = \binom{N}{n} \\ \text{SÍ importa orden, } \#S = V_{N,n} = \binom{N}{n} n! \end{cases}$

CON reposición \rightarrow las muestras con algún elemento repetido SÍ son posibles. la muestra puede tener uno, varios o todos los elementos repetidos.

CR $\begin{cases} \text{NO importa orden, } \#S = CR_{N,n} = \binom{N+n-1}{n} \quad \textcircled{X} \\ \text{SÍ importa orden, } \#S = VR_{N,n} = N^n \end{cases}$ $N+n-1$

$\#S$	NO importa orden	SÍ importa orden
SIN reposición	$C_{N,n} = \binom{N}{n}$	$V_{N,n} = \binom{N}{n} n!$
CON reposición	$CR_{N,n} = \binom{N+n-1}{n}$	$VR_{N,n} = N^n$

(usualmente, no importa el orden)

Muestras SIN reposición y Probab. DESIGUALES

Dada la población $U = \{u_1, \dots, u_N\}$ y una muestra de tamaño n obtenida sin reposición.

$$u_i \rightarrow e_i = \begin{cases} 1 & \text{si } u_i \in \text{Muestra, con probab. } \pi_i \\ 0 & \text{si } u_i \notin \text{Muestra, con probab. } 1 - \pi_i \end{cases}$$

$$e_i \text{ v.a. discreta} \Rightarrow E[e_i] = 1 \cdot \pi_i + 0 \cdot (1 - \pi_i) = \pi_i$$

$$E[e_i^2] = 1^2 \cdot \pi_i + 0^2 \cdot (1 - \pi_i) = \pi_i$$

$$V[e_i] = E[e_i^2] - E[e_i]^2 = \pi_i - \pi_i^2 = \pi_i(1 - \pi_i)$$

$\{e_i \rightarrow B(1, p) ?$

Para la ocurrencia conjunta, definimos:

$$\underline{e_i \cdot e_j} = \begin{cases} 1 & \text{si } (u_i, u_j) \in \text{Muestra, con probab. } \pi_{ij} \\ 0 & \text{si } (u_i, u_j) \notin \text{Muestra, con probab. } 1 - \pi_{ij} \end{cases}$$

$$E[e_i \cdot e_j] = 1 \cdot \pi_{ij} + 0 \cdot (1 - \pi_{ij}) = \pi_{ij}$$

$$\text{cov}[e_i, e_j] = E[e_i \cdot e_j] - E[e_i]E[e_j] = \pi_{ij} - \pi_i \pi_j$$

Propiedades:

$$P1 \rightarrow \sum_{i=1}^N \pi_i = n$$

$$P2 \rightarrow \sum_{\substack{i=1 \\ i \neq j}}^N \pi_i = n - \pi_j$$

$$P3 \rightarrow \sum_{\substack{i=1 \\ i \neq j}}^N \pi_{ij} = (n-1)\pi_j \quad (\text{para } j \neq 0)$$

$$P4 \rightarrow \sum_{\substack{i=1 \\ i \neq j}}^N (\pi_{ij} - \pi_i \pi_j) = -\pi_j (1 - \pi_j)$$

¿dem?

Muestras SIN reposición y Probab. IGUALES

$$\pi_i = P(u_i \in \text{Muestra}) = \frac{CF}{CP} = \frac{N^{\circ} \text{muestras con } u_i}{N^{\circ} \text{total muestras}} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

$$\pi_{ij} = P((u_i, u_j) \in \text{Muestra}) = \frac{CF}{CP} = \dots = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

¿propiedades? las anteriores se demuestran con facilidad

Muestras con reposición y Probab. DESIGUALES

Dada la población $U = \{u_1 \dots u_N\}$ y una muestra de tamaño n obtenida con reposición.

Para cada u_i , defino

$e_i =$ "nº veces que aparece u_i en la muestra"

$e_i \rightarrow B(n, P_i)$ donde $P_i = P(u_i \in \text{muestra en cada extrac})$

$$E[e_i] = n P_i$$

$$V[e_i] = n P_i (1 - P_i)$$

$$\rightarrow P(e_1 = t_1, \dots, e_N = t_N)$$

$$\text{Muestra} \rightarrow \text{Multinomial}, P(\text{Muestra}) = \frac{n!}{t_1! \dots t_N!} P_1^{t_1} \dots P_N^{t_N}$$

donde t_i es el nº de veces que aparece u_i en la muestra.

Para la ocurrencia conjunta,

$e_i \cdot e_j =$ "nº veces que aparecen (u_i, u_j) en la muestra"

$$E[e_i \cdot e_j] = n(n-1) P_i P_j$$

$$\begin{aligned} \text{cov}[e_i \cdot e_j] &= E[e_i \cdot e_j] - E[e_i] E[e_j] = n(n-1) P_i P_j - n P_i n P_j = \\ &= \cancel{n P_i n P_j} - n P_i P_j - \cancel{n P_i n P_j} = -n P_i P_j \end{aligned}$$

Muestras con reposición y Probab. IGUALES

Igual que el anterior, pero más sencillo.

$e_i =$ nº veces que u_i está en la muestra $e_i \rightarrow B(n, p = P_i = \frac{1}{N})$

$$E[e_i] = n/N$$

$$V[e_i] = n \frac{1}{N} \left(1 - \frac{1}{N}\right) = \frac{n(N-1)}{N^2}$$

$$\text{cov}[e_i, e_j] = -n \frac{1}{N} \cdot \frac{1}{N} = -\frac{n}{N^2}$$

5. MÉTODOS DE SELECCIÓN CON PROBABILIDADES PROPORCIONALES AL TAMAÑO de la MUESTRA

Son procedimientos de muestreo que asignan diferentes probabilidades de inclusión en la muestra a las unidades de la población, sobre todo cuando se trabaja con unidades compuestas (unidad = grupo unid. elemental).

Podemos distinguir entre mt. sin reposición y mt. con reposición.

SIN reposición

Modelo polinomial o extremo de una generalización

$$\pi_i = \#u_i \begin{cases} \text{conj.} \rightarrow \pi_i = \#u_i \\ \text{unid. simple} \rightarrow \pi_i = w_i \end{cases}$$

En la poblecc. hay N unidades de muestreo compuestas que contienen un total de M unidades elementales

$$M = \sum_{i=1}^N \pi_i$$

$$\begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_N \end{bmatrix}$$

Este modelo es equivalente a considerar un modelo de urne generalizado con M bolas, π_i de cada color. Si en una extracción sale bola roja, se retirarán todas las bolas rojas antes de la sigte. extracción.

Para $n=2$: $(P_i = \pi_i / \pi)$

$$\begin{aligned} \pi_i &= P(u_i \in \pi_{\text{muestra}}) = P(u_i \in 1^c) + P(u_i \in 2^c \cap u_i \notin 1^c) = \\ &= P(u_i \in 1^c) + P(u_i \in 2^c / u_{j \neq i} \in 1^c) \cdot P(u_{j \neq i} \notin 1^c) = \\ &= P_i + \sum_{j \neq i} \frac{\pi_j}{M - \pi_j} \cdot \frac{\pi_j}{M} = P_i + \frac{\pi_i}{M} \sum_{j \neq i} \frac{\pi_j}{\pi - \pi_j} = \dots = P_i * k = \pi_i * k \end{aligned}$$

$\rightarrow \pi_i$ es proporcional a P_i y a π_i .

$$\begin{aligned} \pi_{ij} &= P((u_i, u_j) \in \pi_{\text{muestra}}) = P(u_i \in 1^c \text{ y } u_j \in 2^c) + P(u_j \in 1^c \cap u_i \in 2^c) = \dots \\ &= P_i P_j \left(\frac{1}{1 - P_i} + \frac{1}{1 - P_j} \right) \end{aligned}$$

Modelo de Ikeda

La 1ª unidad se obtiene sin reposición con probab. P_i proporcional a su tamaño M_i , y las $n-1$ unidades restantes de la muestra se seleccionan sin reposición y con probab. iguales.

Modelo de Brewer

Para $n=2$ $\left\{ \begin{array}{l} 1^a \rightarrow \text{Sin reposición, con probab. proporcional a } \\ k_i = P_i \frac{(1-P_i)}{1-2P_i} \quad / P_i < 1/2 \\ 2^a \rightarrow \text{Sin reposición, con probab. proporcional a } P_i. \end{array} \right.$

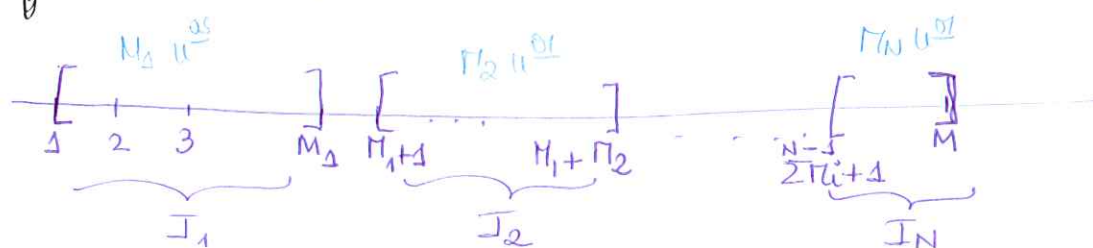
Con reposición

Modelo polinomial: muestreo con reposición y probabilidades independientes y proporcionales a los tamaños

$$\left. \begin{array}{l} \pi_i = \# u_i \text{ ó ponderación de } u_i \\ \sum \pi_i = M \end{array} \right\} P_i = \frac{M_i}{M}$$

Método: 1- Dividir el intervalo de u 's entre $[1, M]$ en N subintervalos con π_i unidades.

Repetir n veces \Rightarrow 2- Elegir al azar $\delta \in \mathbb{Z} / \delta \in [1, M]$ con probab. iguales



Si δ_i está en $J_i \Rightarrow u_i \in \text{Muestra}$.