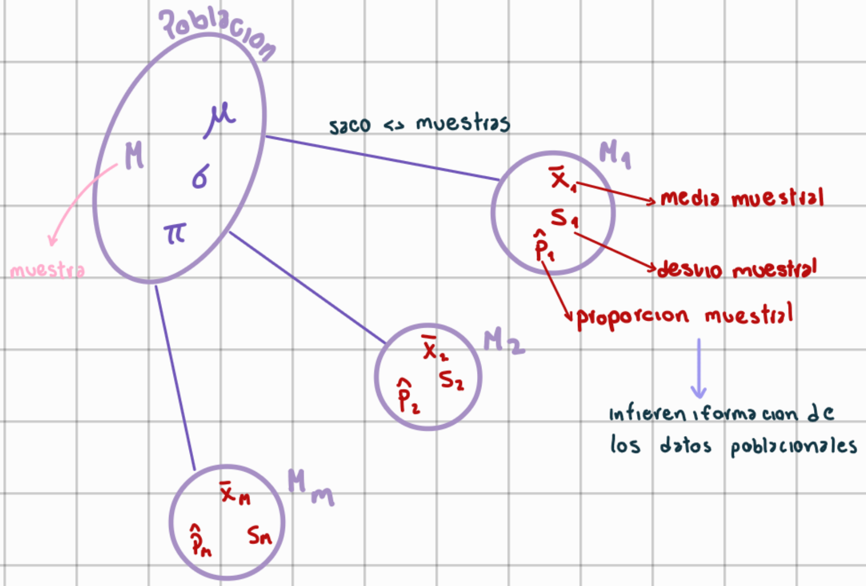
Muestreo

→ subconjunto de la poblacion

(media) μ
(desvio) σ
π
} Parametros Poblacionales

π → proporcion poblacional



Metodos de Muestreo

- Muestreo Aleatorio Simple: de una poblacion tomamos una muestra al azar
- Muestreo Aleatorio Sistematico: se toma una muestra de inicio y luego eligo de acuerdo a una regla, arranco con un número aleatorio y luego lo sistematizo
- Muestreo Aleatorio Estratificado: la poblacion se divide en subgrupos y se toma una muestra al azar de cada estrato
- Muestreo Aleatorio por Conglomerado: poblacion se divide en conglomerados de acuerdo a limites naturales (geograficos o de otra clase)

Definicion de Estadistica

Es una funcion real de la muestra. Sea $(x_1, x_2, \ldots x_n)$ una muestra aleatoria de tamaño n;

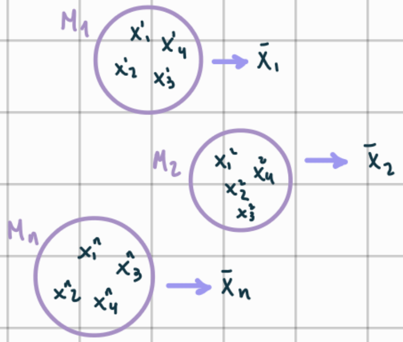
$$T = H(x_1, x_2, \ldots x_n)$$

Ej: $\bar{x}_1 = \dfrac{x_1 + x_2 + \ldots x_n}{n}$

Cada media de la distribucion muestral sigue una distribucion normal

① La media de las medias de las muestras es (=) a la media de la poblacion

$$\mu_{\bar{x}} = \mu$$



$$\bar{x} = \sum_{i=1}^{n} \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_n}{n}$$

② La dispersión de la distribucion muestral de la media, es (+) estrecha que la distribucion poblacional

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \therefore \quad \sigma_{\bar{x}} < \sigma$$

Error estandar de la media $= \frac{\sigma}{\sqrt{n}}$

③ La distribución muestral de la media se aproxima a la normal

TEOREMA CENTRAL DEL LIMITE (TCL)

- Si la población es normal, la distribucion muestral de medias también será normal, independientemente del tamaño de la muestra

- Si la población no es normal (pero al menos es simétrica). La forma de la dist. muestral de medias es aproximadamente normal, aún para valores de la muestra menores a 30 ejemplares
- Si la población tiene una distribucion sesgada o con colas gruesas, la distribución muestral de medias es normal para tamaños de muestra ≥ 30 ejemplares

de analisis empírico, no hay formula para demostrarlo

Ⓩ estandariza los valores de la normal

$$z = \frac{x - \mu}{\sigma} \rightarrow z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Recordamos que...

$$z = \frac{x - \mu}{\sigma} = \frac{x \overset{\pm 1/2}{} - np}{\sqrt{npq}} = \frac{\frac{x - np}{n} \overset{\pm 1/2}{}}{\frac{\sqrt{npq}}{n}} = \frac{x/n \overset{\pm 1/2}{} - p}{\sqrt{\frac{npq}{n^2}}} = \frac{\hat{p} - \pi}{\sqrt{\frac{\cancel{n}\pi(1-\pi)}{n^{\cancel{2}}}}} = \frac{\hat{p} \overset{\pm 1/2}{} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \rightarrow$$

mi p de la población

lo escribo como π porque hablo de mi poblacion

$$\rightarrow \boxed{z = \frac{\hat{p} \pm 1/2 - \pi}{\sigma_{\hat{p}}}} \; ; \; \sigma_{\hat{p}} = \frac{\sqrt{\pi(1-\pi)}}{n}$$

FCC

Factor de correccion por continuidad

La distribucion de muestreo seguirá la distribucion normal con 2 condiciones

* Cuando se sepa que las muestras se toman de una población con distribucion normal
* Cuando se desconoce la distribución de la población (o se sabe que no es normal) pero la muestra contiene POR LO MENOS 30 observaciones

Aplicando el TCL a la variable z que estandariza los valores de la normal, se obtiene

$$z = \frac{(x-\mu)}{\sigma} \xrightarrow{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}} z = \frac{(\bar{x}-\mu)}{\sigma/\sqrt{n}}$$

Distribucion muestral de proporciones

Sabiendo que la distribución binomial puede aproximarse por la normal con: $z = \frac{(x-\mu)}{\sigma} = \frac{(x-np)}{\sqrt{npq}}$

% y x por n ...
$$z = \frac{\left(\frac{x}{n} - \frac{\cancel{n}p}{\cancel{n}}\right)}{\frac{\sqrt{npq}}{n}}$$

$\frac{x}{n}$ ES LA PROPORCION MUESTRAL

lo llamaremos $p = \frac{x}{n}$

y π la proporcion poblacional

$$z = \frac{(p-\pi)}{\sqrt{\frac{\cancel{n}\pi(1-\pi)}{n^{\cancel{2}}}}} = \frac{(p-\pi)}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

$p = \pi$
$q = (1-p) = (1-\pi)$

Situación ACTUAL :)

Cuando trabajamos con proporciones, hay 2 correcciones!

Factor de correccion por continuidad (FCC)

Porque usamos una normal (continua) para aproximar algo discreto (la proporcion)

$$FCC = \frac{1}{2n}$$

Factor de corrección por poblacion finita (FCPF)

Cuando el muestreo es sin reposición, $E(p) \neq \pi$, entonces afecta al desvío

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

- el FCPF tiende a 1 cuando n es pequeño comparado con un N muy grande. Ej $n = 10$ y $N = 1.000.000$
- y es + grande si n se acerca a N

$$z = \frac{\left(p \pm \frac{1}{2n} - \pi\right)}{\sigma_p} \qquad \text{donde} \qquad \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

el signo del FCC...

* Si el valor de z se calcula del EXTREMO IZQUIERDO del area barrida, se resta el FCC

* Si el valor de z se calcula del EXTREMO DERECHO del area barrida, se suma el FCC