Distribuciones de muestreo para población

Población $\mathcal{N}(\mu, \sigma^2)$ $\longrightarrow$ $\bar{X}, s^2$ Muestra

$$\boxed{X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)}$$

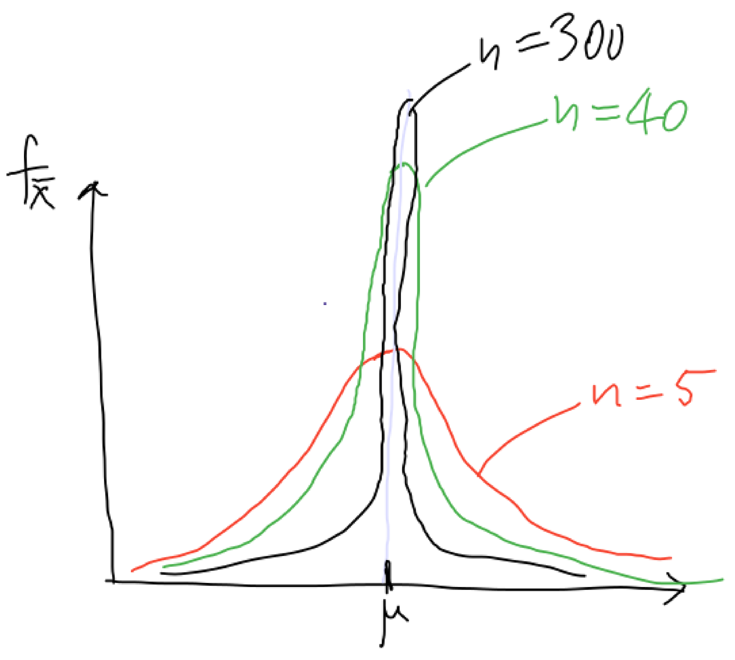
Caso 1: Queremos estimar μ conociendo el valor de σ^2.

Usaremos $T(\underline{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ como estadístico para estimar μ.

Como $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ es una combinación lineal de normales, sabemos que $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

$$\rightarrow E[\bar{X}] = E\left[\tfrac{1}{n}\sum X_i\right] = \tfrac{1}{n} E\left[\sum X_i\right] = \tfrac{1}{n}\sum E[X_i] = \tfrac{1}{n}\sum_{i=1}^{n}\mu = \tfrac{1}{\cancel{n}}\cdot\cancel{n}\mu = \mu$$

$$\rightarrow V(\bar{X}) = V\left(\tfrac{1}{n}\sum X_i\right) = \left(\tfrac{1}{n}\right)^2 V\left(\sum X_i\right) = \tfrac{1}{n^2}\sum V(X_i) = \tfrac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \tfrac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$



$\frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$

$\mathcal{N}(\mu, 0)$

$$\therefore \bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Nota: $$Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$$

Ejemplo

Una cadena de supermercados desea comprar una compañía de alimentos que tiene 80 tiendas en el país. Antes de cerrar el trato, el dueño (de la cadena) analizará los registros financieros de 9 de las tiendas de la empresa que va a adquirir. El director de éstas afirma que las utilidades de cada sucursal siguen una distribución normal con la misma media y con desviación de 25,000 pesos.

(a) ¿Cuál es la probabilidad de que la media muestral de las utilidades se aleje a lo más 6,000 pesos de la media real (o poblacional)?

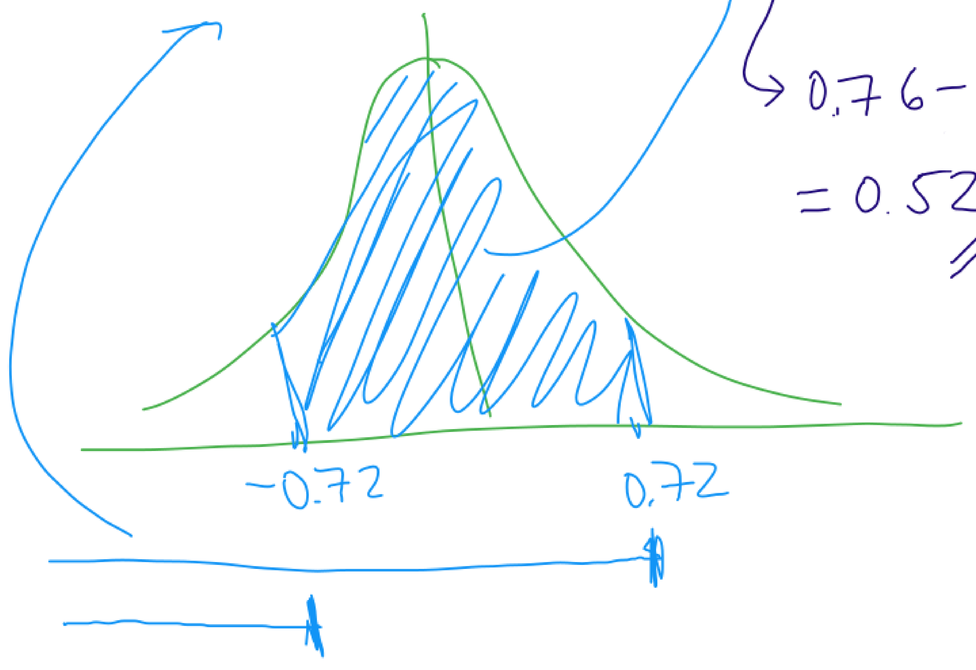
(b) ¿Cuántas tiendas deben incluirse para asegurar, con más del 95% de probabilidad, que la media muestral se alejará a lo más en 6,000 pesos de la media real?

a) $X_1, \dots, X_9 \overset{iid}{\sim} \mathcal{N}(\mu, 25000^2)$

$\bar{X} \sim \mathcal{N}\left(\mu, \frac{25000^2}{9}\right)$

$$P(|\bar{X}-\mu| \le 6000) = P(-6000 \le \bar{X}-\mu \le 6000)$$
$$= P\left(-\frac{3\times 6000}{25000} \le \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le \frac{3\times 6000}{25000}\right)$$
$$= P(-0.72 \le Z \le 0.72) \quad \text{donde } Z \sim \mathcal{N}(0,1)$$
$$= \Phi(0.72) - \Phi(-0.72) = 0.76 - 0.24 = 0.52$$

$$\boxed{\Phi(x) = F_X(x) \text{ donde } X \sim \mathcal{N}(0,1)}$$



b) $$0.95 = P\left(-\frac{\sqrt{n}\cdot 6000}{25000} \le \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le \frac{\sqrt{n}\cdot 6000}{25000}\right)$$
$$= P(-0.24\sqrt{n} \le Z \le 0.24\sqrt{n}) \quad \text{donde } Z \sim \mathcal{N}(0,1)$$

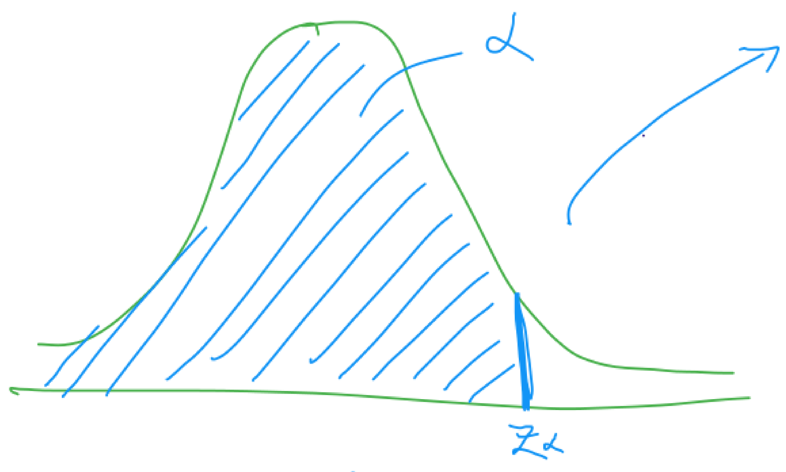
$\Rightarrow \Phi(0.24\sqrt{n}) = 0.975$

$\Rightarrow 0.24\sqrt{n} = \Phi^{-1}(0.975) = 1.96$

$\Rightarrow \sqrt{n} = 8.17$

$\Rightarrow n = 66.75$

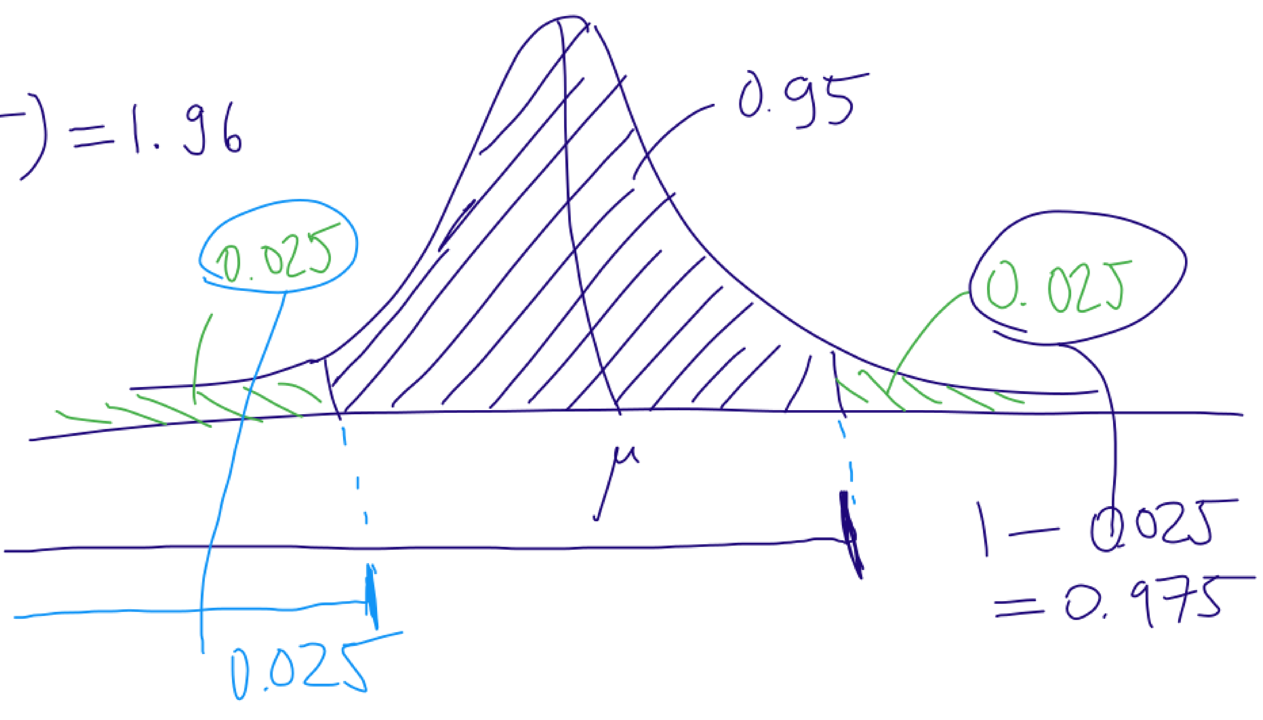
$\therefore n = 67$



$\Phi(z_\alpha) = \alpha$

$z_\alpha = \Phi^{-1}(\alpha)$

$\Phi(-0.24\sqrt{n}) = 0.025$



$1 - 0.025 = 0.975$

Caso 2: Queremos estimar σ^2 (no importa si μ es conocida o no)

Usaremos $T(\underline{X}) = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$$\Rightarrow (n-1)s^2 = \sum (X_i - \bar{X})^2$$

$$\Rightarrow \frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum (X_i-\bar{X})^2 = \sum\left(\frac{X_i-\bar{X}}{\sigma}\right)^2$$

Se parece a $Z = \frac{X_i - \mu}{\sigma}$

$$\Rightarrow \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

Recordemos: $X_i \sim N(\mu, \sigma^2)$

$$Z_i = \frac{X_i - \mu}{\sigma^2} \sim N(0,1)$$

$$Z_i^2 \sim \chi^2_1$$

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$$

$\left.\right\} *$

Nota: en el caso $\frac{(n-1)s^2}{\sigma^2}$, vemos que es muy similar a (*). Sin embargo, en lugar de μ, estamos tomando el estimador de μ ($\bar{X}$). Esto ocasiona la pérdida de un grado de libertad en la χ^2

$W \sim \chi^2_K$ ✓

$$f_W(w) = \frac{1}{2^{K/2}\,\Gamma(\frac{K}{2})} \cdot w^{\frac{K}{2}-1} \cdot e^{-\frac{w}{2}} \cdot I_{\mathbb{R}^+}(w)$$

Una cosa es la distribución gamma y otra la función gamma:

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\,dx \qquad z \in \mathbb{R}$$

$\Gamma(n) = (n-1)!$ $\quad n \in \mathbb{N}$	$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Ejemplo

Si la distribución del puntaje obtenido por alumnos de prepa en el examen de admisión del ITAM es N(110, 20^2), ¿cuál es la probabilidad de que, al tomar una muestra de 20 alumnos, la desviación estándar de su puntaje sea menor 15 puntos?

$$\rightarrow P(s \leq 15) = P(s^2 \leq 225)$$

$$= P\left(\frac{(n-1)s^2}{\sigma^2} \leq \frac{19 \cdot 225}{400}\right)$$

$$= P(W \leq 10.69) \quad \text{donde } W \sim \chi^2_{19}$$

$$= 0.07$$

$$\approx 0.075 \text{ (visto en tablas)}$$

En el 10.117 se acumula el 5%

En el 11.651 se acumula el 10%

$$\frac{10.117 + 11.651}{2} = 10.88$$

Suficientemente parecido a 10.69