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

Caso 1. Estimar μ con $\bar{X}$ con σ^2 conocida

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

Caso 2 Estimar σ^2 con s^2 (sin importar μ)

$$W = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \qquad s^2 = \sum (X_i - \bar{X})^2$$

$\cancel{\mu}\ \bar{X}$

Caso 3: Estimar μ mediante $\bar{X}$, con σ^2 desconocida

$\cancel{\bar{X} \sim \mathcal{N}(\_,\_)}$

Habíamos visto que para σ^2 conocida teníamos el estadístico $\bar{X}$ tal que

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$$

Si usamos $s = \sqrt{s^2}$ para estimar σ, tenemos

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

$$\boxed{\frac{\sqrt{n}(\bar{X}-\mu)}{s}}^{\,1} \sim ?$$

$$T = \boxed{\frac{Z}{\sqrt{\frac{W}{K}}}}^{\,2} = \boxed{\frac{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{(n-1)s^2}{(n-1)\sigma^2}}}}^{\,3} = t_{n-1}$$

$\frac{W}{K}$

$Z \sim \mathcal{N}(0,1)$

$W \sim \chi^2_K$

Tomamos (1) y con retoques algebraicos que no perturban su valor encontramos una forma de reescribirla como (2), llegamos (3).

Paréntesis cultural

Si $X \sim t_K$, su f.d.p. es

$$f_X(x;k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\ \Gamma\left(\frac{k}{2}\right)} \cdot \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} \cdot I_{\mathbb{R}}(x)$$

Ejemplo

La duración de un viaje en Uber en la CDMX se distribuye normal con media y varianza desconocidas. En cierta empresa, los empleados pueden viajar en Uber pagado por la compañía. Con el fin de ahorrar costos, la compañía analizó todos los viajes realizados el mes anterior, y observó una media de 32.5 minutos y una desviación de 13 minutos en 20 viajes.

(a) Encuentra la probabilidad de que la media poblacional sea de una duración mayor a 30 minutos.

$$P(\mu > 30) = P(-\mu < -30) = P(\bar{X}-\mu < \bar{X}-30)$$
$$= P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{s} < \frac{\sqrt{20}(32.5-30)}{13}\right)$$
$$= P(T < 0.86) \quad \text{donde } T \sim t_{n-1}$$
$$= 0.8 \quad \text{(con calculadora) (y también con tablas)}$$

Sean ahora $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$

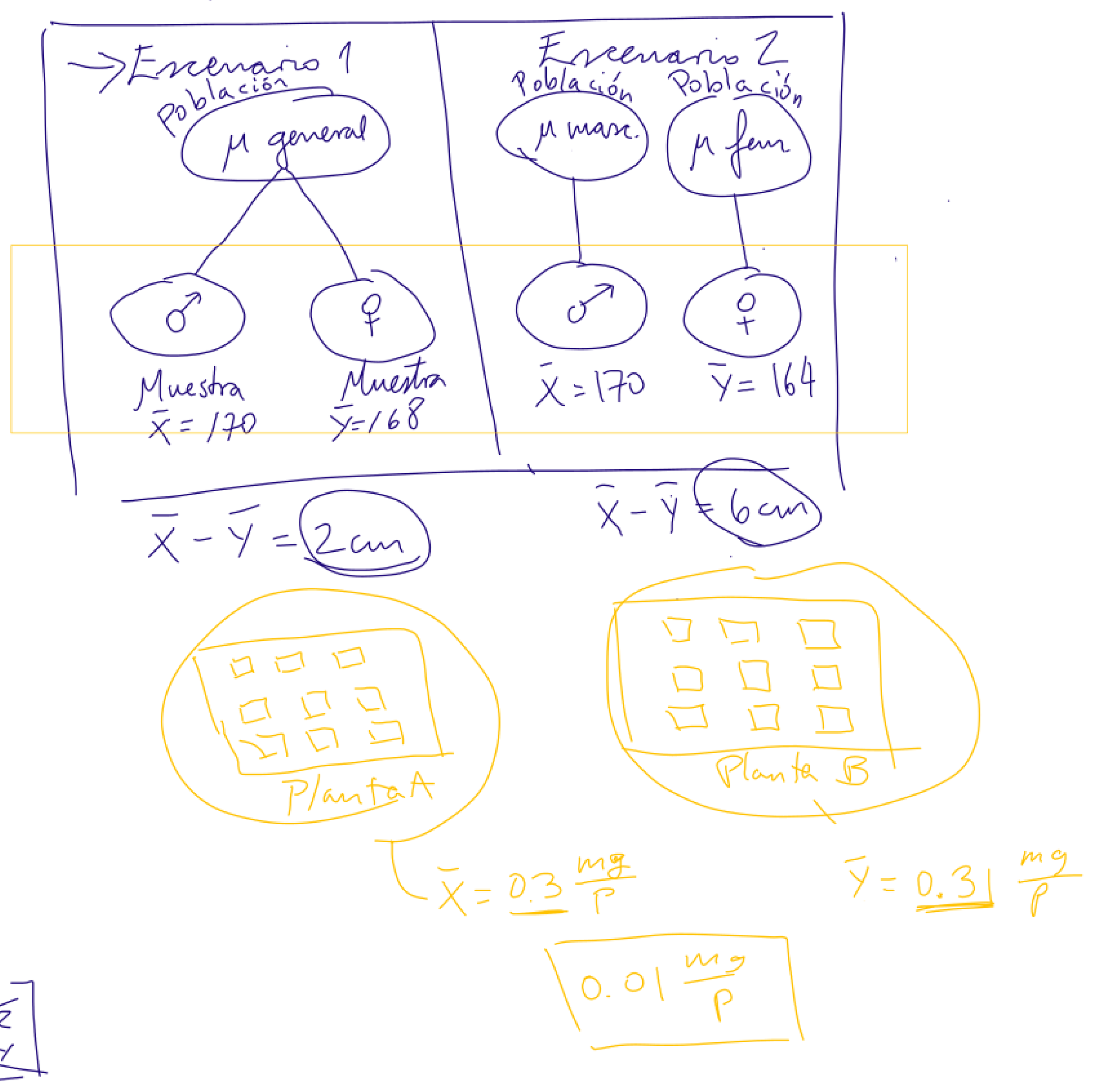
y $Y_1, \dots, Y_m \overset{iid}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$

Caso 4 Queremos estimar $\mu_X - \mu_Y$ mediante $\bar{X} - \bar{Y}$ y conociendo los valores de σ_X^2 y σ_Y^2 $(n \neq m)$

$$\bar{X} \sim \mathcal{N}\left(\mu_X, \frac{\sigma_X^2}{n}\right), \quad \bar{Y} \sim \mathcal{N}\left(\mu_Y, \frac{\sigma_Y^2}{m}\right)$$

$$\Rightarrow \boxed{\bar{X} - \bar{Y}} \sim \mathcal{N}\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

$$\Rightarrow \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim \mathcal{N}(0,1)$$



Caso 5 Queremos estimar $\boxed{\frac{\sigma_X^2}{\sigma_Y^2}}$ $\boxed{\cancel{\sigma_X^2 - \sigma_Y^2}}$

Para comparar las varianzas poblacionales de dos normales indep. con σ_X^2 y σ_Y^2, podríamos hacer $\sigma_X^2 - \sigma_Y^2$. Sin embargo, $\boxed{S_X^2 - S_Y^2}$ sería la diferencia de dos ji-cuadradas, la cual no sigue una distribución conocida. Por eso mejor usamos el cociente

$$\frac{\sigma_X^2}{\sigma_Y^2} \begin{cases} > 1 & \Rightarrow \sigma_X^2 > \sigma_Y^2 \\ = 1 & \Rightarrow \sigma_X^2 = \sigma_Y^2 \\ < 1 & \Rightarrow \sigma_X^2 < \sigma_Y^2 \end{cases}$$

Resulta que $\frac{S_X^2}{S_Y^2}$ se asocia con la F de Fisher.

Recordemos: $V \sim \chi^2_\alpha$ $\quad W \sim \chi^2_\beta$

$$\Rightarrow \frac{V/\alpha}{W/\beta} \sim F_{\alpha,\beta}$$

$$V = \frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$$

$$W = \frac{(m-1)S_Y^2}{\sigma_Y^2} \sim \chi^2_{m-1}$$

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1,m-1}$$

$$\frac{S_X^2}{S_Y^2} = \frac{\frac{\frac{(n-1)S_X^2}{\sigma_X^2}}{n-1} \cdot \sigma_X^2}{\frac{\frac{(m-1)S_Y^2}{\sigma_Y^2}}{m-1} \cdot \sigma_Y^2} = \frac{\left(\frac{V}{n-1}\right)}{\left(\frac{W}{m-1}\right)} \cdot \left(\frac{\sigma_X}{\sigma_Y}\right)^2$$

χ^2_{n-1} $\quad \chi^2_{m-1}$

$$\Rightarrow \frac{S_X^2}{S_Y^2} = \frac{V/(n-1)}{W/(m-1)} \cdot \frac{\sigma_X^2}{\sigma_Y^2}$$

$$\Rightarrow \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2} = \frac{V/(n-1)}{W/(m-1)} \cdot \cancel{\frac{\sigma_X^2}{\sigma_Y^2}} \cdot \cancel{\frac{\sigma_Y^2}{\sigma_X^2}}$$

$$\therefore \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2} \sim F_{n-1,m-1}$$

Paréntesis cultural

Si $X \sim F_{n,m}$ entonces

$$f_X(x; n, m) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \cdot \left(\frac{n}{m}\right)^{\frac{n}{2}} \cdot x^{\frac{n}{2}-1} \cdot \left(1 - \frac{n}{m}x\right)^{-\frac{n+m}{2}} \cdot I_{\mathbb{R}^+}(x)$$

Ejemplo

La estatura de los mexicanos sigue una distribución normal. La varianza poblacional es la misma en cada estado del país. Por otro lado, se cuenta con dos muestras: una de 16 personas de Oaxaca y otra de 22 personas de Baja California, cada una con desviación estándar de 4.7 cm y 8.8 cm, respectivamente.

(a) Calcular la probabilidad de que la varianza muestral de los oaxaqueños sea del doble o más que la varianza muestral de los bajacalifornianos.

(b) Bajo la evidencia muestral, ¿es razonable pensar que la varianza de la estatura de los oaxaqueños y los bajacalifornianos es la misma?

a) Sea X la estatura de los oaxaqueños

y sea Y " " " BC $\quad \boxed{\sigma_X^2 = \sigma_Y^2}$

$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $\quad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

$$P(S_X^2 > 2S_Y^2) = P\left(\frac{S_X^2}{S_Y^2} > 2\right) = P\left(\frac{S_X^2}{S_Y^2} \cdot \cancel{\frac{\sigma_Y^2}{\sigma_X^2}}^{1} > 2 \cdot \cancel{\frac{\sigma_Y^2}{\sigma_X^2}}^{1}\right)$$

$$= P(F > 2) \quad \text{donde } F \sim F_{15,21}$$

$= 0.07$ (con calculadora)

$= 0.075$ (con tablas)

$1.83 \sim 2 \sim 2.18$

$90\% \sim 92.5\% \sim 95\% \}$ $P(F \leq 2)$

Caso 6 Queremos estimar $\mu_X - \mu_Y$ con $\bar{X} - \bar{Y}$ sin conocer σ_X^2 y σ_Y^2.

Trabajaremos con el supuesto adicional de que

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2$$

$$X \sim \mathcal{N}(\mu_X, \sigma^2) \qquad Y \sim \mathcal{N}(\mu_Y, \sigma^2)$$

es decir que la varianza poblacional es "común"!

En este caso, podemos aproximar σ^2 mediante la varianza muestral combinada ("pooled variance"), s_p^2.

Def $X \sim \mathcal{N}(\mu_X, \sigma^2) \quad Y \sim \mathcal{N}(\mu_Y, \sigma^2)$

$$s_X^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2 \qquad s_Y^2 = \frac{1}{m-1}\sum (Y_i - \bar{Y})^2$$

$$\Rightarrow \quad S_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}$$

Teo 1: $$W = \frac{(n+m-2)s_p^2}{\sigma^2} \sim \chi^2_{n+m-2}$$

$$\left[\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}\right]$$

Teo 2. $$T = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_p^2}{n} + \frac{S_p^2}{m}}} = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

$$\left[\frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim \mathcal{N}(0,1)\right]$$