Algunas distribuciones muestrales

ESCUELA POLITÉCNICA SUPERIOR DE CÓRDOBA
Universidad de Córdoba
Departamento de Estadística

Distribución de la media muestral

Sea $\vec{\mathcal{X}} = (\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n) \in D\left(\mu, \sigma^2\right)$. El estadístico $\bar{X}$ (media muestral)

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i \in D\left(\mu, \frac{\sigma^2}{n}\right)$$

Casos a considerar:

- Si el tamaño de la muestra es grande, aplicando el Teorema Central del Límite:

$$\bar{X} \in D\left(\mu, \frac{\sigma^2}{n}\right) \to N\left(\mu, \frac{\sigma^2}{n}\right)$$ O lo que es lo mismo $$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \to N(0,1)$$

- σ^2 conocida:

  Si $n \geq 30$ la aproximación es buena.

- σ^2 desconocida:

  Se puede calcular (σ^2) mediante la varianza muestral S^2 y la aproximación es buena si $n \geq 100$, es decir:

$$Z = \frac{\bar{X} - \mu}{S / \sqrt{n}} \to N(0,1)$$

Si la muestra proviene de una población Normal:

- σ^2 conocida:

$$\mathcal{X}_i \in D\left(\mu,\sigma^2\right) \Rightarrow \bar{X} \in N\left(\mu,\frac{\sigma^2}{n}\right) \Leftrightarrow Z = \frac{\bar{X}-\mu}{\sigma / \sqrt{n}} \to N(0,1)$$

- σ^2 desconocida:

  Se calcula (σ^2) mediante S^2, la aproximación anterior es buena si $n \geq 20$.

$$Z = \frac{\bar{X}-\mu}{S / \sqrt{n}} \to N(0,1)$$

  En caso contrario (n < 20) se utiliza la siguiente aproximación, que por tanto, es válida para todo n:

$$T = \frac{\bar{X}-\mu}{S / \sqrt{n-1}} \to t(n-1)$$

Apéndice 1

$$N(0,1) \quad t_{(n-1)} \quad \chi^2_{(n-1)}$$

$$\frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{\dfrac{nS^2}{\sigma^2}}{n-1}}} = \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\dfrac{\dfrac{S\sqrt{n}}{\sigma}}{\sqrt{n-1}}} = \frac{(\bar{X}-\mu)\sqrt{n-1}}{\dfrac{\sigma}{\sqrt{n}}\,\dfrac{\sqrt{n}S}{\sigma}} = \frac{(\bar{X}-\mu)\sqrt{n-1}}{S} = \frac{\bar{X}-\mu}{S/\sqrt{n-1}} \rightarrow t_{(n-1)}$$

Distribución de la varianza muestral

Sea $\vec{\mathcal{X}} = (\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n) \in D(\mu, \sigma^2)$

- Si μ es conocida: $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$
- Si μ es desconocida: $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$E(S^2) = \sigma^2 - \frac{\sigma^2}{n}$$

$$\mathrm{V}(S^2) = \frac{\mu_4 - \sigma^4}{n}$$

- Si el muestreo se realiza en una población normal:

$$S^2 \to N\left(\sigma^2 - \frac{\sigma^2}{n}, \frac{\mu_4 - \sigma^4}{n}\right)$$

- El estadístico usado en el estudio de la varianza es (Teorema de Craig):

$$Q = \frac{nS^2}{\sigma^2} \to \chi^2_{(n-1)}$$

Distribución del parámetro p en una población Binaria

Sea $\vec{X} = (X_1, X_2, \ldots, X_n) \in B(p)$ $\begin{cases} E[X_i] = p \\ V(X_i) = pq \end{cases}$

Definimos el estadístico $\hat{p}$ (proporción) $\hat{p} = \dfrac{\dot{X}}{n}$

siendo: $\dot{X} = \sum_{i=1}^{n} X_i \in b(n;p) \xrightarrow{T.C.L.} N(np;npq)$

$$\Rightarrow Z = \frac{\dot{X} - np}{\sqrt{npq}} \to N(0;1)$$

por tanto: $\hat{p} \to N\left(p; {}^{pq}\!/_{n}\right)$ O equivalentemente $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}} \to N(0\,;1)$

Distribución de la diferencia de medias en poblaciones normales

Sean:

$$\left.\begin{array}{l}\vec{X} \in N\left(\mu_X, \sigma_X^2\right) \\ \vec{Y} \in N\left(\mu_Y, \sigma_Y^2\right)\end{array}\right\} \Rightarrow \begin{array}{l}\overline{X} \in N\left(\mu_X, \dfrac{\sigma_X^2}{n_X}\right) \\ \overline{Y} \in N\left(\mu_Y, \dfrac{\sigma_Y^2}{n_Y}\right)\end{array}$$

Consideremos la nueva v.a. $\overline{Z} = \overline{X} - \overline{Y}$

$$E\left[\overline{Z}\right] = E\left[\overline{X}\right] - E\left[\overline{Y}\right] = \mu_X - \mu_Y$$

$$V\left[\overline{Z}\right] = V\left(\overline{X} - \overline{Y}\right) = 1^2 V(\overline{X}) + (-1)^2 V(\overline{Y}) - 2COV(\overline{X}, \overline{Y}) \quad \Rightarrow$$

$$\Rightarrow \quad \overline{Z} \in N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y} - 2COV(\overline{X}, \overline{Y})\right)$$

donde $COV(\overline{X}, \overline{Y}) = \dfrac{\sigma_{XY}}{n_X n_Y}$

- Si las muestras están relacionadas $\sigma_{XY} \neq 0$, generalmente se realiza un estudio de la v.a. $Z = X - Y$, donde por regla general $n_X = n_Y$

- Si las muestras son independientes, $\sigma_{XY} = 0$

$$\overline{Z} \in N(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y})$$

Casos posibles para σ_X^2 y σ_Y^2 :

- σ_X^2 y σ_Y^2 son conocidas:

$$\overline{Z} = \overline{X} - \overline{Y} \in N(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}) \Rightarrow Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \to N(0,1)$$

- $\sigma_X^2 = \sigma_Y^2$ y desconocidas:

$$\overline{Z} \in N\left(\mu_X - \mu_Y, \sigma^2\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)\right) \Rightarrow U = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \to N(0,1)$$

Si calculamos σ^2 mediante $S^2 = \dfrac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}$ entonces:

$$U = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \quad \to \quad t_{(n_X + n_Y - 2)}$$

- $\sigma_X^2 \neq \sigma_Y^2$ y desconocidas:

$$U = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \quad \rightarrow \quad t_{(n)}$$

donde $n \cong n_X + n_Y - 2$, si $n_X \cong n_Y$ y relativamente grande, o bien:

$$n = \frac{\left({S_X^2}/{n_X} + {S_Y^2}/{n_Y} \right)^2}{\frac{\left({S_X^2}/{n_X} \right)^2}{n_X - 1} + \frac{\left({S_Y^2}/{n_Y} \right)^2}{n_Y - 1}}$$

ESCUELA POLITÉCNICA SUPERIOR DE CÓRDOBA
Universidad de Córdoba
DEPARTAMENTO DE ESTADÍSTICA

Distribución del cociente de varianzas en poblaciones normales

Sean: $\vec{X} \in \mathcal{N}(\mu_X, \sigma_X^2)$ y $\vec{Y} \in \mathcal{N}(\mu_Y, \sigma_Y^2)$. Considerando que: $\dfrac{n_X S_X^2}{\sigma_X^2} \in \chi^2_{(n_X-1)}$

y que $\dfrac{n_Y S_Y^2}{\sigma_Y^2} \in \chi^2_{(n_Y-1)}$, entonces:

$$F = \frac{\dfrac{\underbrace{\dfrac{n_X S_X^2}{\sigma_X^2}}_{\chi^2_{(n_x-1)}}}{n_X - 1}}{\dfrac{\underbrace{\dfrac{n_Y S_Y^2}{\sigma_Y^2}}_{\chi^2_{(n_y-1)}}}{n_Y - 1}} = \frac{(n_Y - 1) n_X S_X^2 \sigma_Y^2}{(n_X - 1) n_Y S_Y^2 \sigma_X^2} \in \mathcal{F}_{(n_X-1, n_Y-1)}$$

Si $n_X = n_Y = n$, la expresión anterior se reduce a:

$$F = \frac{S_X^2 \sigma_Y^2}{S_Y^2 \sigma_X^2} \in \mathcal{F}_{(n-1,n-1)}$$

Diferencia de proporciones en dos poblaciones Binaria Independientes

Sean $\vec{X} = \left(X_1, X_2, \ldots, X_{n_x}\right) \in B\left(p_X\right)$ $\begin{cases} E\left[X_i\right] = p_X \\ V\left(X_i\right) = p_X\, q_X \end{cases}$ $\hat{p}_X \to N\left(p_X ; {p_X q_X}/{n_X} \right)$

$\vec{Y} = \left(Y_1, Y_2, \ldots, Y_{n_y}\right) \in B\left(p_Y\right)$ $\begin{cases} E\left[Y_i\right] = p_Y \\ V\left(Y_i\right) = p_Y\, q_Y \end{cases}$ $\hat{p}_Y \to N\left(p_Y ; {p_Y q_Y}/{n_Y} \right)$

Definimos el estadístico Diferencia de proporciones: $\hat{p}_X - \hat{p}_Y$

$$\hat{p}_X - \hat{p}_Y \to N\left(p_X - p_y \; ; \frac{p_X q_X}{n_X} + \frac{p_Y q_Y}{n_Y} \right)$$

$$Z = \frac{\hat{p}_X - \hat{p}_Y - \left(p_X - p_y\right)}{\sqrt{\frac{p_X q_X}{n_X} + \frac{p_Y q_Y}{n_Y}}} \to N(0 \,; 1)$$

Corrección para poblaciones finitas

Los estadísticos que hemos estudiado para una población:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \in N\left(\mu; \sigma^2/n\right)$$

- σ^2 conocida: $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \in N(0;1)$
- σ^2 desconocida: $T = \dfrac{\bar{X}-\mu}{S/\sqrt{n-1}} = \dfrac{\bar{X}-\mu}{\bar{S}/\sqrt{n}} \in t_{(n-1)}$

corresponden a la situación en la que el tamaño de la población donde se obtienen las muestras es o se supone infinito, pero si la población de donde se va obtener la muestra es de tamaño FINITO, se ha de aplicar un coeficiente de corrección *C.*

$$C = \sqrt{\frac{N-n}{N-1}}$$

donde: *N* ≡ “Tamaño de la población”,
n ≡ “Tamaño muestral”

Si la relación $\dfrac{n}{N} > 0.05$ la población se considera de tamaño <u>finito</u>.

Los estadísticos anteriores quedan transformados:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \in N\left(\mu; \frac{\sigma^2}{n}\frac{N-n}{N-1}\right)$$

$$Z = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \in N(0;1)$$

$$T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n-1}}\sqrt{\frac{N-n}{N-1}}} = \frac{\bar{X}-\mu}{\frac{\bar{S}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \in t_{(n-1)}$$

Algunas distribuciones muestrales

ESCUELA POLITÉCNICA SUPERIOR DE CÓRDOBA
Universidad de Córdoba

DEPARTAMENTO DE ESTADÍSTICA