4 Distribuciones Muestrales de los Estadísticos más usuales

4.1 El estadístico media muestral

Consideremos una v.a. poblacional X que sigue una distribución de probabilidad con $E(X) = \mu$ y $Var(X) = \sigma^2$.

Si queremos estimar la media poblacional μ, parece razonable escoger una muestra y calcular la media de esta muestra.

Sea $(X_1, X_2, ..., X_n)$ una m.a.s. de la v.a. $X \Rightarrow$

Las variables muestrales $X_1, X_2, ..., X_n$ son independientes, $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$, $\forall i \;\; \Rightarrow$

Definimos el estadístico MEDIA MUESTRAL como sigue:

$$\boxed{\overline{X} = \frac{\sum\limits_{i=1}^{n} X_i}{n}}$$

Se tiene que:

$$\begin{aligned} E(\overline{X}) &= \mu \\ Var(\overline{X}) &= \frac{\sigma^2}{n} \end{aligned}$$

sin más que aplicar las propiedades vistas de la esperanza y la varianza para variables independientes.
CUANDO LA VARIANZA σ^2 ES CONOCIDA, su distribución muestral viene dada por los dos resultados siguientes:

Teorema de la aditividad de la distribución normal:

Sea $(X_1, X_2, ..., X_n)$ una m.a.s. de una v.a. $X \sim N(\mu, \sigma)$, entonces

$$\boxed{\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \implies Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)}$$

Teorema central del límite:

Sea $(X_1, X_2, ..., X_n)$ una m.a.s. de una v.a. X tal que $E(X) = \mu$ y $Var(X) = \sigma^2$, entonces la variable aleatoria

$$\overline{X} \text{ se aproxima a la distribución } N(\mu, \frac{\sigma}{\sqrt{n}}) \text{ cuando } n \longrightarrow \infty.$$

Esto es,

$$\boxed{\overline{X} \underset{n\geq 30}{\approx} N(\mu, \frac{\sigma}{\sqrt{n}}) \implies Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \underset{n\geq 30}{\approx} N(0, 1)}$$

Nota: En general, para cualquier distribución, la variable media muestral se puede aproximar por la distribución $N(\mu, \frac{\sigma}{\sqrt{n}})$ cuando $n \geq 30$, entonces $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \underset{n \geq 30}{\approx} N(0,1)$.

CUANDO LA VARIANZA σ^2 ES DESCONOCIDA, la estimamos a partir de los datos mediante la

VARIANZA MUESTRAL $\boxed{S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} = \frac{n}{n-1}\left(\overline{X^2} - \overline{X}^2\right)}$

A la raíz cuadrada $S = \sqrt{S^2}$ se le llama **desviación típica o estándar muestral**.

Entonces para conocer la distribución de probabilidad de la variable aleatoria $\frac{\overline{X} - \mu}{S/\sqrt{n}}$ tenemos el siguiente resultado:

Teorema de Fisher:

Sea $(X_1, X_2, ..., X_n)$ una m.a.s. de una v.a. $X \sim N(\mu, \sigma)$, entonces:

- Los estadísticos $\overline{X}$ y S^2 son independientes.
- La variable aleatoria

$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$ sigue una **distribución t de Student con n-1 grados de libertad**

Esto es, $T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

Nota: En general, para cualquier distribución, la variable media muestral se puede aproximar por la distribución de una $N(\mu, \frac{S}{\sqrt{n}})$ cuando $n \geq 30$, entonces $\frac{\overline{X} - \mu}{S/\sqrt{n}} \underset{n \geq 30}{\approx} N(0,1)$.

4.2 El estadístico proporción muestral

Supongamos que los individuos de una población determinada pueden presentar o no una cierta característica y queremos estimar la **proporción de unidades que en la población presentan dicha característica**, p.
La variable aleatoria poblacional la definimos como sigue:

$$X = \begin{cases} 1, & \text{si el individuo presenta la característica} \\ 0, & \text{en otro caso} \end{cases} \quad \Longrightarrow \quad X \sim b(p)$$

Sea $(X_1, X_2, ..., X_n)$ una m.a.s. de la v.a. $X \Rightarrow$

Las variables muestrales $X_1, X_2, ..., X_n$ son independientes, $E(X_i) = p$ y $Var(X_i) = p(1-p)$, $\forall i \ \Rightarrow$

Definimos el estadístico **PROPORCIÓN MUESTRAL** como sigue:

$$\boxed{\widehat{p} = \frac{\sum_{i=1}^{n} X_i}{n}}$$

Se tiene que:

$$E(\widehat{p}) = p$$

$$Var(\widehat{p}) = \frac{p(1-p)}{n}$$

por las propiedades de la esperanza y la varianza de variables aleatorias independientes ya citadas.

Y, aplicando el teorema central del límite, tenemos que su distribución muestral es:

$$\boxed{\widehat{p} \underset{n\geq 30}{\approx} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) \Rightarrow Z = \frac{\widehat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \underset{n\geq 30}{\approx} N(0,1)}$$