

Muestreo

Población:

El conjunto de intereses de estudio.

Muestra:

Es un subconjunto de una población. Al hecho de obtener una muestra se le llama **muestreo**.

Muestra aleatoria:

Es una muestra elegida de manera aleatoria. Para ello deben cumplirse las siguientes condiciones.

- Cada individuo debe tener una probabilidad conocida de ser seleccionado.
- La muestra debe ser de elementos independientes, es decir, la elección de un elemento es independiente de los demás.

Existen dos tipos de muestreo

- Remplazo: Cada elemento escogido en la muestra es reintegrado una vez hecha la observación es reintegrado a la población antes de elegirse al siguiente individuo. Todo individuo puede aparecer más de una vez en una muestra.
- Sin remplazo: Cuando el individuo escogido no es reintegrado a la población y solo puede aparecer una vez en la muestra.

Una población se considera conocida bajo un modelo X cuando se sabe la distribución de probabilidad $f_X(x)$. A los parámetros de la distribución que describen a una población se les llama **parámetros poblacionales**.

La utilidad de las muestras es poder hacer predicciones o estimaciones sobre una población sin la necesidad de tener que estudiar toda la población, pueden haber muchos factores, como;

- el costo de hacer un estudio,
- la imposibilidad física de hacer el estudio,
- observaciones que aún no estén disponibles.

Una muestra puede ser significativa para lograr describir una población, de la misma muestra depende nuestro estudio.

Estadísticos muestrales

Supongamos el caso de una población a la que se desea conocer la altura de sus ciudadanos. Sea X la variable aleatoria que mide los valores de las alturas. Supongamos además que se tienen 100 personas al azar y que las observaciones son las siguientes,

- x_1 corresponde a una primera observación X_1
- x_2 corresponde a una segunda observación X_2 .

y en general x_n es la observación X_n .

Una muestra de tamaño n es descrita por observaciones x_1, \dots, x_n de las variables X_1, \dots, X_n estas variables conforman la parte aleatoria de la muestra y nos interesa ahora poder calcular

$$P(X_1 = x_1, \dots, X_n = x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

Se le llama *estadístico muestral* a una función g tal que, al obtener observaciones del vector $\hat{X} = (X_1, \dots, X_n)$, el valor $g(\hat{X})$ estima un parámetro poblacional. También a la distribución de \hat{X} se le llama distribución muestral.

Por inferir nos referimos a las conclusiones que se pueden obtener mediante una muestra y que tan confiables son las conclusiones.

Media muestral

Se define al media muestral del vector aleatorio de tamaño n como

$$g(\hat{X}) = \frac{1}{n} \sum_{i=1}^n X_i$$

Es usual denotar a $g(\hat{X}) = \bar{X}$. El teorema limite central es importante en este aspecto, pues;

$$\bar{X} \sim Norm(\mu, \sigma)$$

entonces cuando $n \rightarrow \infty$.