Distribuciones fundamentales de muestreo

Gustavo Ahumada

Distribuciones muestrales

El objetivo del análisis estadístico es obtener conocimiento con respecto a ciertas propiedades en un población que son de interés para un investigador. El **muestreo** es una de las formas más frecuentes de obtener información acerca de una población de interés. Por ejemplo, podemos afirmar, con base en las opiniones de varias personas entrevistadas en las calles Santiago de Chile, que aproximadamente el 70% prefiere un determinada marca automovíl. En este caso tratamos con una muestra aleatoria de opiniones de una población finita muy grande. Finalmente, consideremos una máquina despachadora de refrescos en la que la cantidad promedio de bebida servida se mantiene en 240 mililitros. Un inspector de calidad de la compañia calcula la media de 40 bebidas y obtiene $\bar{x} = 236$ mililitros, y con base en este valor decide que la maquina aún sirve bebidas con un contenido premedio de $\mu = 240$ mililitros. Las 40 bebidas representan una muestra de la población infinita de posibles bebidas que esta máquina servirá.

Definición. La distribución de probabilidad de una estadística se llama **distribución muestral.** La distribución muestral de probabilidad de $\bar{X}$ se llamada **distribución muestral de la media.**

Distribuciones muestrales de medias

La primera distribución muestral importante a considerar es la media $\bar{X}$. Suponga una muestra aleatoria de tamaño n se toma de una población normal con media μ y varianza $\sigma 2$. Cada observación $\bar{X}_i, i = 1, 2..., n$, de la muestra aleatoria tendrá entonces la misma distribución normal que la población que se muestrea. Tenemos:

$$\bar{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

tiene una distribución normal con media

$$\mu_{\bar{X}} = \frac{\mu + \mu + ... + \mu}{n} = \mu$$

y varianza

$$\sigma^2_{\bar{x}} = \frac{\sigma^2 + \sigma^2 + ... + \sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Si tomamos muestras de una población con distribución desconocida, finita o infinita, la distribución muestral de $\bar{X}$ aún será aproximadamente normal con media μ y varianza σ^2/n siempre que el tamaño sea grande. El resultado anterior se debe al teorema del límite central.

Definición. Teorema del límite central: si $\bar{X}$ es la media de una muestra aleatoria de tamaño n tomada de una población con media μ y varianza σ^2/n, entonces la forma límite de la distribución de

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

conforme $n \longrightarrow \infty$, es la distribución normal estándar $n(z; 0, 1)$.

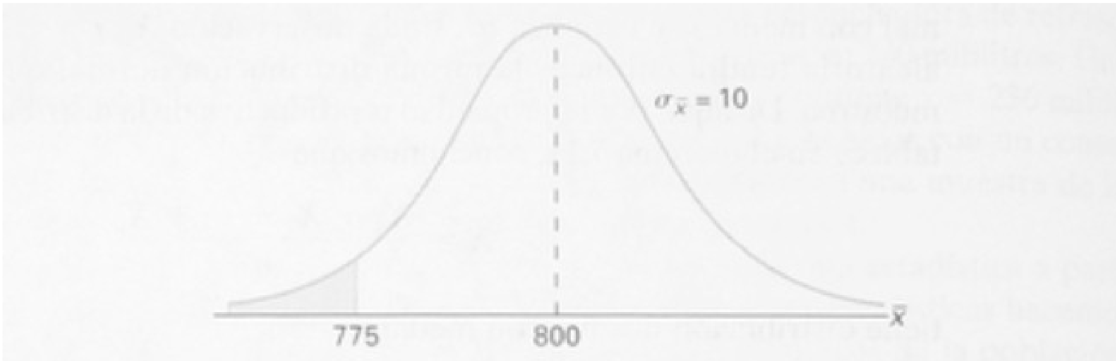

Figure 1: Área de focos con una vida menor a 775 horas

Ejemplo: Una empresa eléctrica fabrica focos que tienen una duración que se distribuye aproximadamente en forma normal, con media 800 horas y desviación estándar de 40 horas. Encuentre la probabilidad de que una muestra aleatoria de 16 focos tenga una vida promedio de menos de 775 horas.

Solución: La distribución muestral de $\bar{X}$ será aproximadamente normal, con $\mu_{\bar{X}} = 800$ y $\sigma_{\bar{X}} = 40/\sqrt{16} = 10$. La probabilidad que se desea está dada por el área de la región sombreada de la figura de arriba.

En correspondencia con $\bar{x} = 775$, encontramos que

$$z = \frac{775 - 800}{10} = -2.5,$$

y por lo tanto

$$P(\bar{X} < 775) = P(Z < -2.5) = 0.0062.$$

Inferencia sobre la media de la población

Una aplicación muy importante del teorema central del límite es determinar los valores razonables de la media de la población μ. Temas como la prueba de hipótesis, estimación, control de calidad emplean este teorema.

Ejempllo

pagina 222 Walpole