Capítulo 6. Inferencia

6.1 Resumen Teórico de Inferencia

6.1.1 Introducción

La inferencia estadística trata de obtener conclusiones sobre la población a partir de la información proporcionada por una muestra aleatoria; es decir, obtener de las propiedades de la muestra una aproximación fiable a las del colectivo o población en estudio.

Las inferencias sobre el valor de un parámetro poblacional, como es la media μ o la proporción p, se pueden hacer mediante **estimaciones** (puntuales o por intervalos de confianza) y mediante **contrastes de hipótesis.**

6.1.2 Conceptos

Parámetro: Es un valor numérico que describe una característica de la población (μ, p, σ^2, etc.)

Estadístico: Es toda función de los datos muestrales. Dicha función asigna a cada muestra de tamaño n elegida de la población (por m.a.s.), un valor numérico; se tiene, por tanto, una variable aleatoria que tendrá una distribución de probabilidad llamada Distribución en el muestreo del estadístico.

Estimador para un parámetro poblacional desconocido es un estadístico que nos da un valor que pertenece al conjunto de valores que puede tomar el parámetro que se estima. Para la mayoría de las muestras ese valor está próximo al parámetro poblacional desconocido.

- Se utiliza para la media poblacional μ el **estimador MEDIA MUESTRAL**

$$\overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Y para la proporción muestral p el **estimador PROPORCIÓN MUESTRAL**

$$\hat{P} \approx N\left(p, \sqrt{\frac{pq}{n}}\right), \qquad q = 1 - p.$$

(Se considerarán las muestras de tamaño $n \geq 30$ para poder aplicar el Teorema Central del Límite y asegurar la distribución anterior).

6.1.3 Estimación Puntual

Consiste en tomar como valor del parámetro poblacional desconocido (**μ, p...**), el de un estadístico ($\bar{x}, \hat{p}, ...$), obtenido en una muestra aleatoria elegida de la población objeto de estudio; es decir, al ofrecido por el estimador sobre una muestra.

Para que un estimador proporcione estimaciones fiables del parámetro poblacional debe ser:

- Centrado o insesgado: la media de la distribución muestral del estadístico coincide con el parámetro poblacional desconocido.
- Consistente: al aumentar el tamaño de la muestra, el valor medio de la distribución muestral del estadístico tiende al parámetro estimado.
- Eficiente: que sea el de menor varianza de entre todos los insesgados.

Se utilizarán los estimadores definidos en el apartado anterior para hacer estimaciones de la media y la proporción poblacional.

6.1.4 Estimación por intervalos de confianza

Consiste en encontrar un intervalo (a, b) de manera que tengamos una cierta confianza de que el parámetro poblacional desconocido se encuentre en dicho intervalo.

Se considera que la población de partida sigue una distribución Normal con desviación típica conocida (σ) para la estimación de μ, o una distribución Binomial para la estimación de p.

Para construir el intervalo:

a) Se elige un **estimador** del parámetro que se desea estimar ($\overline{X}$ para μ, y $\hat{P}$ para p).
b) Se elige el **nivel de confianza 1-α** con el que se desea construir el intervalo, eso quiere decir que, antes de elegir la muestra, se tendrá una probabilidad 1-α de que el intervalo construido a partir de esa muestra contenga al parámetro de la población.
c) **Se toma una muestra aleatoria de la población de tamaño n** y en ella se obtiene el valor del estadístico correspondiente. (Con cada muestra elegida, el valor del estadístico sería distinto y se obtendría un intervalo diferente, de ahí que el 100(1-α) % de los posibles intervalos construidos contendría el verdadero valor del parámetro desconocido de la población).
d) Se construye el intervalo centrado en el estadístico ($\bar{x}$, o $\hat{p}$) obtenido en la muestra; que sería:

$$I_{100(1-\alpha)\%}(\mu) = \left(\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \ , \ \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \right) \text{ para estimar } \mu$$

$$\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} \ , \ \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} \right) \text{ para estimar } p$$

Donde $z_{1-\alpha/2}$ **es el punto crítico** de la variable aleatoria Normal tipificada $Z \approx N(0,1)$ tal que $P(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}) = 1-\alpha$, siendo éste el nivel de confianza elegido. De esa igualdad se deduce la $P(Z \le z_{1-\alpha/2}) = 1-\frac{\alpha}{2}$ en la tabla de la distribución Normal, que nos dará el correspondiente valor crítico $z_{1-\alpha/2}$.

- Se llama **amplitud del intervalo** a la diferencia= Extremo superior – Extremo inferior.

- El **error de la estimación** es la diferencia, en valor absoluto, entre el parámetro poblacional y el estadístico muestral, por lo tanto el error máximo de estimación será el radio del intervalo:

Error máximo = $z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ para el intervalo de la media y

Error máximo = $z_{1-\alpha/2}\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}$ para el intervalo de la proporción.

6.1.5 Tamaño de la Muestra

-Si aumentamos el tamaño de la muestra elegida, manteniendo la misma confianza, o sea, el mismo valor crítico $z_{1-\alpha/2}$, menor error cometemos al inferir el valor del parámetro.

Esto quiere decir que podemos calcular el tamaño de muestra necesario para tener un error máximo concreto con un nivel de confianza previamente fijado.

-Si se mantiene fijo el tamaño de la muestra y se desea aumentar el nivel de confianza $1-\alpha$ (con lo que aumentaría el valor crítico $z_{1-\alpha/2}$), aumentaría también el error de la estimación.