



Unidad III Inferencia estadística (Estimación)

"La mayoría de nosotros encontramos pronósticos, proyecciones, estimaciones y aproximaciones en alguna forma u otra casi a diario. El pronóstico del Servicio Meteorológico acerca de la probabilidad de que llueva; la proyección del economista del último giro de la economía, el informe mensual del índice de precios al mayoreo el índice de precios al consumidor son todos ejemplos comunes de estimaciones." (Estadística paso a paso, Christensen, 1983, p. 309)

Estimación

La **estimación** es el procedimiento utilizado cuando se desea conocer el valor de un **parámetro**. Por ejemplo, ¿Cuál es la media de la distancia que deben viajar los estudiantes de su casa al CCH? Para contestar a esa pregunta, es necesario tomar una muestra y calcular la media muestral \bar{x} . Supóngase que se selecciona una muestra aleatoria de 100 distancias y resulta una media de 10.22 kilómetros, por lo que este valor será una **estimación puntual** de la media de la distancia que deben viajar los estudiantes de su casa al CCH.



Otro ejemplo, se selecciona una muestra aleatoria de 120 bolsas chicas de papas Zavrithas para estimar la media de contenido neto de papas fritas. Y después de efectuado el muestreo, se obtiene una media de 39.5 gramos. Es decir, la media muestral es $\bar{x} = 39.5$, con ello se ha estimado que la media de contenido de papas fritas de bolsas chicas Zavrithas en 39.5 gramos. Esto no quiere decir que la media poblacional, es decir μ , sea exactamente igual a 39.5, sino que se interpreta esta estimación como " μ está próxima a 39.5". El término de proximidad es relativo, pero quizás en este caso "próximo" debe ser definido arbitrariamente como estar "dentro de 1 gramo de μ ". Si 1 gramo satisface la idea intuitiva de proximidad, entonces decir " μ está próxima a 39.5" es comparable a decir " μ está entre 38.5 ($39.5 - 1$) y 40.5 ($39.5 + 1$)". Este tipo de estimación utiliza el concepto de intervalo y asigna una medida a la confiabilidad de éste en la estimación del parámetro en cuestión.



Cuando realizamos un estudio en una muestra, nuestra **muestra solo es una de las posibles muestras que podríamos haber seleccionado partir de la población**. Por lo tanto, nuestra muestra nos proporciona una estimación puntual (por ejemplo, media, proporción, diferencia de medias, diferencia de proporciones, etc.), que será solo una de las múltiples estimaciones puntuales teóricas que hubiéramos obtenido con otras muestras. Si estas estimaciones teóricas siguen alguna distribución de probabilidad conocida, podremos estimar su precisión (intervalo de confianza) y hacer pruebas de hipótesis. La distribución de probabilidad de estas estimaciones puntuales va a depender del tipo de variable, de tamaño muestral, de su magnitud y, para variables continuas, de otros parámetros descritos en la muestra (por ejemplo, desviación estándar). Con estos parámetros calcularemos los estadísticos de referencia, equivalentes al valor z que vimos para la distribución normal, a los que les corresponderá una probabilidad, que podemos consultar en tablas o con algún programa computacional. También podremos hacer el paso inverso, encontrar el valor o rango de valores que corresponde una probabilidad; esto nos servirá, por ejemplo, para calcular entre qué valores se encontrarán el 95% de la estimación teórica, lo que corresponde al **intervalo de confianza** del 95%.

Nivel de confianza ($1-\alpha$)

Probabilidad de que la muestra por seleccionarse produzca valores límite que se localicen en lados opuestos del parámetro que se estima. Algunas veces el nivel de confianza se llama **coeficiente de confianza**.

Sabemos que parte de la idea de la Estadística es utilizar los resultados obtenidos de la muestra como valores aproximados a los que se supone operan para la población; es decir, **usamos los estadísticos para “tantear” o estimar los parámetros**, y es eso precisamente a lo que se refiere el término estimación. Existen dos formas de realizar **estimación: puntual y por intervalos**. La **estimación puntual consiste en calcular un estadístico y asumir que su valor se aproxima al valor del parámetro**. La **estimación por intervalos consiste en generar precisamente un intervalo a partir de valores obtenidos de una muestra, y dentro del que se espera esté contenido algún parámetro en particular**. Ahora, se sabe que una muestra es tomada de manera aleatoria, por lo que deberá de incluirse en la estimación un elemento de probabilidad, y este elemento nos permitirá dar cierto grado de certidumbre a nuestras estimaciones; esto es, si sabemos cómo se comporta la población de interés, o al menos el estimador con el que se trabaje, en términos de distribución de probabilidad, entonces podemos decidir el porcentaje de seguridad para nuestra estimación; esto es, podemos establecer de antemano la probabilidad de que al construir el intervalo de confianza, efectivamente consigamos que dentro de éste se encuentre el parámetro a estimar.

La interpretación de estos intervalos de confianza se basará en que la inferencia estadística estudia los métodos para obtener conclusiones de toda una población a partir del análisis de una muestra.

En nuestra vida cotidiana a menudo oímos o realizamos este tipo de interpretaciones, por ejemplo:

- De 4 novios que ha tenido mi amiga, 3 la han engañado, dice **“Casi todos los hombres son iguales” (engaños)**.



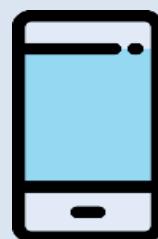
- Probar una sopa cuando se está cocinando, para decidir: **está lista**. Es decir, se realiza un muestreo (probadita) y ese resultado se usa para caracterizar a toda la población (Toda la sopa).



- Cuando consideras que **requieres de 30 minutos para trasladarte de tu casa a la escuela**, debido a que en otras ocasiones (registros de tiempo) has requerido aproximadamente ese tiempo.



- Al concluir que un **98% de los del CCH tienen celular**, realizas una interpretación basada en los estudiantes que conoces o has observado y sabes tienen celular (muestreo), por lo que, basado en esto, extrapolas a la población de estudiantes.



- Ver los cortos de una película y concluir: "**parece ser que está buena esa película**", en donde debes notar que al decir "parece ser" asumes que no tienes la seguridad (probabilidad de 1) que la película será de tu agrado, ya que, en este caso, el muestreo no es aleatorio y seguramente esté conformado por cortos (muestreo) que en conjunto NO representen a la película (población).

