Conceptos Esenciales

- Uno de los objetivos de la inferencia estadística es sacar una conclusión sobre una población sobre la base de una muestra aleatoria de la población. Obviamente, las muestras aleatorias varían, por lo que necesitamos entender cuánto varían y cómo se relacionan con la población. Nuestro objetivo final es crear un modelo de probabilidad que describa el comportamiento a largo plazo de las mediciones de muestra. Utilizamos este modelo para hacer inferencias sobre la población.

- En las estadísticas, cuando queremos describir las características de una muestra, llamamos a las estadísticas de valores. Sin embargo, cuando queremos describir las características de una población, llamamos a esos valores parámetros.

- Podemos usar la teoría matemática para derivar expresiones para la media y la desviación estándar de la distribución de muestreo de la proporción de la muestra. Al tomar muchas muestras aleatorias de tamaño n de una distribución de la población con proporción de población p:
    - La media de la distribución de las proporciones de la muestra es p.
    - La desviación estándar de la distribución de las proporciones de la muestra es $\sqrt{\frac{p(1-p)}{n}}$

- Para tener una idea del patrón de variación en las proporciones de la muestra, necesitamos generar más de cinco muestras. La distribución que muestra cómo las proporciones de la muestra varían de una muestra a otra se denomina distribución de muestreo de la proporción de la muestra.

- Al tomar muchas muestras aleatorias de tamaño n de una distribución de la población con proporción de población p:
    - La media de la distribución de las proporciones de la muestra es p
    - La desviación estándar de la distribución de las proporciones de la muestra es $\sqrt{\frac{p(1-p)}{n}}$
    - Si $np \geq 10$ y $n(1-p) \geq 10$ entonces el Teorema Del Límite Central establece que la distribución de las proporciones de la muestra sigue una distribución normal aproximada con la media p y la desviación estándar

- En la práctica, no conocemos la proporción de la población, ni podemos permitirnos tomar miles de muestras aleatorias. En su lugar, observamos una sola muestra aleatoria. En este caso, necesitamos estimar la media y la desviación estándar de la proporción de la muestra:
    - La media estimada de la distribución de las proporciones de la muestra es $\widehat{p}$.
    - Para distinguirlo de la verdadera desviación estándar de las proporciones de muestra, llamamos a la desviación estándar estimada de las proporciones de muestra el error estándar de $\widehat{p}$.

Ecuaciones clave

error estándar:

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

desviación estándar

$$\sqrt{\frac{p(1-p)}{n}}$$

Glossario

teorema del límite central

si $np \geq 10$ y $n(1-p) \geq 10$ entonces el Teorema Del Límite Central establece que la distribución de las proporciones de la muestra sigue una distribución normal aproximada con la media p y la desviación estándar $\sqrt{\frac{p(1-p)}{n}}$

parámetros

números que describen una población

población

la población es toda la colección de individuos o objetos sobre los que quieres aprender

muestra

una muestra es una parte de la población que se selecciona para el estudio

error estándar

un percentil de una distribución es el valor en el que un cierto porcentaje cae por debajo de ese valor

estadística

números que son calculados de una muestra