Probabilidad y Estadística

Agustín G. Bonifacio

UNSL

Estadística: Descripción de Datos con Medidas Numéricas

Pretendemos describir o predecir la conducta de la población con base en la información obtenida a partir de una muestra representativa de esa población.

Definición

La estadística descriptiva consta de los procedimientos usados para resumir y describir las características importantes de un conjunto de medidas.

Definición

La estadística inferencial consiste en los procedimientos que se aplican para plantear inferencias con respecto a las características de la población a partir de la información contenida en una muestra tomada de esa población.

El objetivo de la estadística inferencial es hacer inferencias (conclusiones, predicciones, toma de decisiones) con respecto a las características de una población a partir de la información contenida en una muestra.

Pasos necesarios para lograr el objetivo de la Estadística Inferencial

1. Especificar preguntas que se deben plantear e identificar la población de interés.
2. Decidir cómo seleccionar la muestra.
3. Seleccionar la muestra y analizar la información de la muestra.
4. Usar la información del paso anterior para hacer inferencias con respecto a la población.
5. Determinar la confiabilidad de la inferencia.

Definición

Las medidas numéricas descriptivas relacionadas con una población de mediciones se llaman parámetros, aquellas calculadas desde una muestra de mediciones se llaman estadísticos.

Definición

La media aritmética o promedio de una lista $x_1, \ldots, x_n$ de n mediciones viene dada por

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Observaciones

- Un uso importante de $\bar{x}$ es como estimador de μ, la media poblacional.
- $\bar{x}$ varía con cada muestra.

Definición

La mediana de una lista $x_1, \ldots, x_n$ de n mediciones es el valor que se encuentra en la posición central cuando las mediciones están ordenadas de menor a mayor.

Observaciones

- Cuando n es par, la *mediana* se toma como el promedio de los dos valores centrales.
- La *mediana* es menos sensible a los valores extremos o atípicos que la *media*.

Definición

La moda de una lista $x_1, \ldots, x_n$ de n mediciones es el valor que se presenta con más frecuencia.

Definición

El rango de una lista de n mediciones se define como la diferencia entre la medición mayor y la menor.

Definición

La varianza (muestral) de una lista de n mediciones $x_1, \ldots, x_n$ se define mediante la fórmula

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Definición

La desviación estándar de una lista de n mediciones $x_1, \ldots, x_n$ es igual a la raíz cuadrada (positiva) de la varianza.

Definición

Dadas n mediciones ordenadas de menor a mayor, el p-ésimo percentil es el valor que excede $p\,\%$ de las mediciones y es menor que el restante $(100-p)\,\%$.

Definición

- El 25-percentil se denomina cuartil inferior, Q_1.
- El 50-percentil es la mediana.
- El 75-percentil se denomina cuartil superior, Q_3.

Ejemplo

Determine los cuartiles inferior y superior para el siguiente grupo de mediciones:

$$16, 25, 4, 18, 11, 13, 20, 8, 11, 9.$$

Definición

El rango intercuartil, IQR para un grupo de mediciones es $IQR = Q_3 - Q_1$.

Definición

El resumen de 5 números de un grupo de n mediciones $x_1, \ldots, x_n$ viene dado por

$$\min x_i, Q_1, \bar{x}, Q_3, \max x_i.$$

Definición

La puntuación z de una observación x mide la distancia de una observación al promedio, en unidades de desviación estándar, esto es,

$$z = \frac{x - \bar{x}}{s}.$$

Definición

Se dice que un estimador $\hat{\theta}$ de un parámetro θ es insesgado si $E(\hat{\theta}) = \theta$.

Definición

La distancia entre una estimación y el parámetro estimado se llama error de estimación.

De ahora en más suponemos que los tamaños de las muestras son *grandes*, por lo que los estimadores insesgados que estudiaremos tienden a distribuirse de forma normal, como consecuencia del Teorema Central del Límite.

Definición

El margen de error (ME) del estimador viene dado por:

$$ME = 1,96\ SE$$

donde SE denota el error estándar del estimador.

Ejemplo

Para estimar la media poblacional μ, el estimador $\bar{x}$ es insesgado y tiene error estándar dado por

$$SE = \frac{\sigma}{\sqrt{n}}.$$

Entonces

$$ME = \pm 1,96 \frac{\sigma}{\sqrt{n}}.$$

Si no se conoce σ y $n \geq 30$, se puede aproximar σ con s.

Ejemplo

Para estimar la proporción poblacional p en una población binomial, el estimador $\hat{p} = x/n$ es insesgado con

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

por lo cual

$$ME = \pm 1,96\sqrt{\frac{p(1-p)}{n}},$$

que se estima como

$$\pm 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Definición

Un estimador por intervalo es una regla para calcular dos números a y b que crean un intervalo $[a, b]$ que contiene con una probabilidad de $(1-\alpha)$ al parámetro de interés. El número $(1-\alpha)$ se denomina coeficiente de confianza.

¿Cómo se construye un intervalo de confianza?

Intervalo de confianza de $(1-\alpha)100\,\%$ para una muestra grande

$$\text{Estimador Puntual} \pm z_{\alpha/2} SE$$

Ejemplo

Intervalo de confianza de $(1-\alpha)100\,\%$ para la media poblacional μ de una muestra grande

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

donde n es el tamaño de la muestra y σ es la desviación estándar de la población. Cuando no se conoce σ, se puede aproximar mediante s.

Observación

Notemos que lo anterior es equivalente a escribir

$$P\left(\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]\right) = 1-\alpha.$$

Ejemplo

Intervalo de confianza de $(1-\alpha)100\,\%$ para una proporción poblacional p en una muestra grande

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

donde n es el tamaño de la muestra.

Observación

Un buen intervalo de confianza debe cumplir:

1. Ser tan estrecho como sea posible.
2. Tener un coeficiente de confianza grande, cercano a 1.

Límite de confianza inferior de $(1-\alpha)100\,\%$:

$$\text{Estimador Puntual} - z_\alpha SE$$

Límite de confianza superior de $(1-\alpha)100\,\%$:

$$\text{Estimador Puntual} + z_\alpha SE$$