

Inferencia estadística: Introducción

Recuento estadística descriptiva

Msc. Carlos Andres Lopez

Universidad Tecnológica de Pereira
Maestria en Ingenieria de sistemas y computación
asrael@utp.edu.co

8 de febrero de 2020

Índice

1 Estadística

- Definición
- Conceptos
- Escalas

2 Descriptiva

- Tipos de datos
- Medidas de tendencia central
- Anotaciones
- Variabilidad de los datos

3 Aplicación

4 Bibliografía

Conceptos basicos

Estadística (def): disciplina que se ocupa de los metodos para recolectar, organizar y analizar datos con el fin de establecer conclusiones validas.

Existen tres grandes ramas:

- Analisis de datoa: Recolecta, organiza y muestra de froma resumida los datos.
- Probabilidad: las leyes del azar mas allá de los juegos de azar.
- Inferencia estadística: la ciencia de obtener conclusiones sobre un conjunto de datos especificos, usando conocimiento sobre las probabilidades.

Conceptos Básicos

- Datos: hechos (numeros, cantidades) que se recolectan para interpretar la realidad
- Conjunto de datos: datos recolectados asociados a un fenómeno
- Elementos: entidades sobre las que se recolectan los datos
- Variables: características de un elemento
- Observación: conjunto de mediciones obtenidas de un elemento
- Población: conjunto de elementos de interés para un estudio en particular
- Muestra: subconjunto de la población

Escalas de Medida

- Escala nominal: escala de medición cuando los datos son etiquetas o nombres, pueden ser o no numericos. No poseen orden, no es posible decir si uno es mejor que otro. Ejemplo: nacionalidad, genero, ciudad.
- Escala ordinal: variables con las propiedades de los datos nominales pero pueden ser ordenados o jerarquizado, pueden ser o no numericos. Ejemplo: pequeño, mediano, grande.
- Escala de intervalo: variables con las propiedades de los datos ordinales, se expresan en términos de alguna unidad de medición.

Tipos de datos

- Datos cualitativos: nombran o etiquetan categorías de elementos.
- Datos cuantitativos: cuantifican los elementos.
- Distribución de frecuencia: resumen plasmado en una tabla que muestra el número de elementos en cada una de las clases (que no se superponen)

Medidas de distribución central central

- Media: suma de todos los valores dividido por el numero de observaciones

$$\bar{x} = \sum \frac{x}{n} \quad (1)$$

La media muestral tiene como denominador n y la media poblacional N.

- Mediana: valor en medio de las observaciones cuando estas se encuentran ordenadas de menor a mayor. Divide la distribucion en dos partes.
- Moda: valor que se encuentra con mayor frecuencia, valor mas comun.

Anotaciones

- En **variables nominales** no se puede hallar **media o mediana**.
- La **mediana** ayuda a establecer la existencia de datos **atipicos**.
- **Dato atipico** es un valor que se encuentra a distancia desproporcionada del valor **medio**.
- Para describir correctamente la distribución se necesita mas que las medidas de tendencia central.

Rango

Diferencia entre el valor mas alto y el valor mas bajo.

- Facil de entender
- Facil de calcular
- No nos da buena información sobre la variabilidad
- Solo considera los valores extremos

Rango intercuartil

- No contiene valores extremos
- Distribución de datos en cuatro partes iguales
- El segundo cuartil Q_2 divide la distribución en dos partes iguales
- Q_2 es la mediana

El rango intercuartil (IQR) es la distancia entre el primer y tercer cuartil

$$IQR = Q_3 - Q_1 \quad (2)$$

La principal ventaja del IQR es que no es afectado por los datos atípicos.

Caja de bigotes

Es una grafica para describir el centro, la variabilidad y los datos atipicos. los bigotes contienen valores sin incluir los datos atipicos.
Calculo de valor atipico:

$$valor < Q_1 - 1,5(IQR) \quad (3)$$

$$valor > Q_3 + 1,5(IQR) \quad (4)$$

Varianza

Varianza: muestra la cantidad de valores esparcidos al rededor de la media. A mayor varianza mayor variabilidad de los datos.

La varianza esta basada en la distancia entre cada valos (x) y la media.

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \quad (5)$$

Desviación estandar

Es el promedio de la distancia de una observación desde la media.

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad (6)$$

La desviación estandar es la medida de la dispersión mas empleada. La desviación estandar es la raiz cuadrada de la varianza, de esta forma los datos son expresados en las mismas unidades de la variable analizada.

Coefficiente de variación

Permite describir que tan grande es la desviación estandar en relación con la media.

$$cv = \left(\frac{s}{media} * 100 \right) \quad (7)$$

Es útil para comparar la variabilidad de variables que tienen desviaciones estándar distintas y medias distintas.

Ejercicio demostrativo 1

Las notas de un grupo de 11 en quimica fueron:

x	nota
1	7.4
2	7.9
3	4.1
4	8.1
5	6.2
6	7.1
7	7.4
8	6.7

- En una grafica de puntos, como se vería?
- Cual es el centro de la distribución?
- Cual es la variabilidad de la distribución?
- Construya la caja de bigotes

Ejercicio demostrativo 2

Las tarifas de renta de automóviles por día en siete ciudades de Estados Unidos son las siguientes :

Ciudad	Tarifa por día
Boston	\$43
Atlanta	35
Miami	34
New York	58
Orlando	30
Pittsburgh	30
Washington, D.C.	36

- Encuentre la media, la mediana, la moda, la varianza y la desviación estándar de estas tarifas.

Ejercicio demostrativo, continuación

- En una muestra similar de siete ciudades de la costa opuesta de Estado Unidos la media muestral de las tarifas fue de \$38 por día. La varianza y la desviación estándar fueron 12.3 y 3.5 cada una. Analice la diferencia entre las tarifas de las ciudades del este y del oeste.

Bibliografía

- [1] Anderson, David R., Dennis J. Sweeney y Thomas A. Williams. Estadística para administración y economía, 10a. ed. Cengage Learning Editores, S.A.
- [2] Murray R. Spiegel, Larry J. Stephens. Estadística, 4a. ed. McGRAW-HILL/INTERAMERICANA EDITORES, S.A.