Universidad Autónoma de Querétaro

Curso Propedéutico Actuaría

Estadística y probabilidad

Medidas de tendencia central y dispersión

En general se dice que existen dos medidas de interés que son aplicadas a cualquier conjunto de datos: la localización del centro de estos, y su respectiva variabilidad. La tendencia central de un conjunto de datos se define como la disposición de dichos datos para agruparse, ya sea alrededor del centro o de ciertos valores numéricos. Por otro lado, la variabilidad de un conjunto de datos se dice que es la dispersión de las observaciones en el conjunto.

Se puede decir que principalmente existen tres medidas de tendencia central: la media, la mediana y la moda.

Definición 1.1. La media de las observaciones $x_1, x_2, \dots, x_n$ es el promedio aritmético de éstas y se denota como: $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$

La media es una medida de tendencia central que se usa en diversos conjuntos de datos para obtener información de estos, sin embargo, es necesario destacar que al ser una medida que usa todos los datos del conjunto inicial puede verse afectada por algún dato atípico que exista.

A continuación, daremos dos ejemplos, suponga que tenemos los salarios de los trabajadores de una sucursal de una empresa transnacional que contiene 15 empleados (sin contar al CEO de la empresa):

Salarios={12907,7209,14093,6698,8153,12549,8217,15036,15476,15772,14247,

6685,6556,6670,15534}

Entonces, computando en la definición 1.1 obtenemos un promedio de 11053.47, pero ¿Qué pasaría si agregamos el salario del CEO de la empresa?, recordemos que dijimos que los salarios eran solamente de una sucursal, y el CEO es el director general, es decir, no de la sucursal sino de toda la empresa.

Salarios_nuevos={12907,7209,14093,6698,8153,12549,8217,15036,15476,15772, 14247, 6685,6556,6670,15534,160000}

Ahora nuestro promedio se dispara de los 11053.47 a los 20362.62, es decir, estamos diciendo que en general los trabajadores del conjunto “Salarios_nuevos” están ganando aproximadamente 20362.62, pero esto claramente es absurdo porque nadie gana más de 15600 en el conjunto “Salarios”. Como el lector puede notar, el ejemplo es bastante simple, y el método lo es aún más ya que solo se trata de sumar y dividir, pero al igual que es importante saber realizar los cálculos, es igual de importante saber que significan, ya que de nada sirve tener un numero y no tener idea de que dice.

Definición 1.2. La mediana de un conjunto de observaciones es el valor para el cual, cuando todas las observaciones se ordenan de manera creciente, la mitad de éstas es menor que este valor y la otra mitad mayor.

Es importante decir que, si el número de observaciones en el conjunto es impar, la mediana es el valor de la observación que se encuentra a la mitad del conjunto ordenado. Si el número es par se considera la mediana como el promedio aritmético de los valores de las dos observaciones que se encuentren a la mitad del conjunto ordenado. Por otro lado, se puede determinar a partir de la distribución acumulativa, es decir, la mediana es el percentil cincuenta.

Debido a que la mediana es un valor que se basa en la secuencia ordenada de las observaciones en un conjunto de datos, es necesario saber que la existencia de algunos valores extremos no afectará su valor. Por lo tanto, si un conjunto contiene unos cuantos valores extremos atípicos no afectaran al valor de la mediana, de esta manera la mediana puede ser una medida de tendencia central más deseable que la media.

Definición 1.3. La moda de un conjunto de observaciones es el valor de la observación que ocurre con mayor frecuencia en el conjunto.

La moda muestra hacia qué valor tienden los datos a agruparse. En conjuntos relativamente pequeños, puede que no exista un par de observaciones cuyo valor sea el mismo. En esta situación no es clara la definición de moda. También puede suceder que la frecuencia más alta se encuentre compartida por dos o más observaciones. En estos casos, la moda tiene una utilidad limitada como medida de tendencia central.

Definición 1.4. La varianza de las observaciones $x_1, x_2, \dots, x_n$ es en general la media del cuadrado de las distancias entre cada observación y la media del conjunto de observaciones. La varianza se denota por:

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

La varianza es una medida de dispersión bastante buena de la variabilidad debido a que, si muchas de las diferencias son grandes o pequeñas en un conjunto de datos, entonces el valor de la varianza será grande o pequeño. Cabe destacar que la varianza no se salva de los efectos de la existencia de datos atípicos.

Definición 1.5. La desviación estándar es la raíz cuadrada de la varianza y se denota por:

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

Naturalmente, la varianza y la desviación estándar no son medidas de variabilidad ajenas, ya que la ultima no puede determinarse a menos que se conozca la primera. Usualmente existe una preferencia por la desviación estándar porque ésta se expresa en las mismas unidades físicas de las observaciones.

Una vez que introdujimos las nociones básicas de la estadística vamos a proceder a dar una breve introducción de su hermana, la probabilidad.

Introducción a la probabilidad

Comenzaremos dando unas definiciones básicas:

Definición 1.6. Experimento determinístico. Resultados conocidos con exactitud.

Un ejemplo seria conocer la distancia de un objeto si tenemos previamente los datos de la velocidad y del tiempo, es determinístico porque si tengo las dos variables anteriores entonces tengo exactamente el resultado.

Definición 1.7. Experimento aleatorio. Aquel experimento donde sabemos que puede ocurrir, pero no lo que ocurrirá con exactitud.

El ejemplo clásico seria el del dado, sabemos que si arrojamos un dado vamos a obtener un número del 1 al 6, pero no sabemos que número vamos a obtener, es decir, sabemos que será un número en el rango ya definido, pero no podemos decir con exactitud qué número será.

Definición 1.8. Espacio muestral. Conjunto Ω de todos los posibles resultados de un experimento aleatorio. Puede ser:

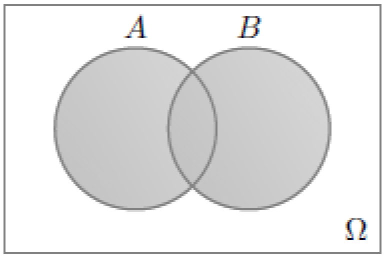
- Discreto: finito o infinito numerable.
- Continuo: infinito no numerable.

Definición 1.9. Suceso. Es un subconjunto del espacio muestral.

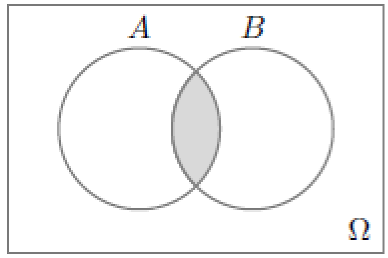
En el ejemplo del dado, el espacio muestral seria {1,2,3,4,5,6} y un suceso sería, por ejemplo, el de sacar un numero par {2,4,6}.

Definición 1.10. Sucesos disjuntos. A y B son disjuntos si $A \cap B = \emptyset$

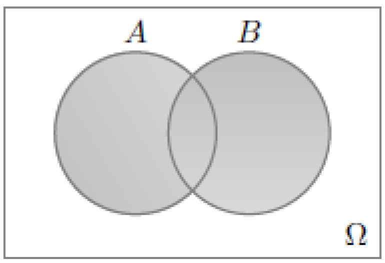
Sean A y B dos subconjuntos cualesquiera de Ω. Recordamos a continuación las operaciones básicas de unión, intersección, diferencia y complemento.



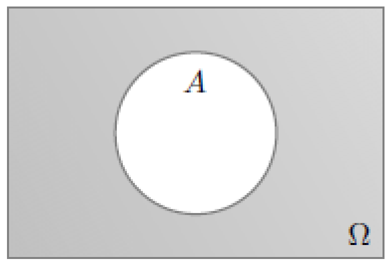
$A \cup B$



$A \cap B$



$A - B$



A^c

Por otro lado, es fácil verificar las siguientes propiedades, cabe destacar que estas propiedades servirán en un futuro para demostrar algunos teoremas básicos de probabilidad.

$A \cup \varnothing = A.$ $\quad A \cap \varnothing = \varnothing.$ $\quad A \cup \Omega = \Omega.$

$A \cap \Omega = A.$ $\quad A \cup A^c = \Omega.$ $\quad A \cap A^c = \varnothing.$

Cabe destacar que las operaciones de conjuntos de unión e intersección se caracterizan por ser asociativas:

$$\begin{aligned} A \cup (B \cup C) &= (A \cup B) \cup C \\ A \cap (B \cap C) &= (A \cap B) \cap C \end{aligned}$$

Y así mismo son distributivas:

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \end{aligned}$$

Para finalizar daremos a conocer las leyes de De Morgan, recordemos que A^c se lee como A complemento, o complemento de A.

$$
\begin{aligned}
(A \cup B)^c &= A^c \cap B^c \\
(A \cap B)^c &= A^c \cup B^c
\end{aligned}
$$