# Estadística

Javier Alberto Pérez Garza

11 de septiembre de 2019

# Estadística

La estadística es la disciplina encargada de estudiar la recolección, organización, análisis, visualización e interpretación de datos con el objetivo de generar conocimiento.

# Variables aleatorias

Son variables que toman valores de forma aleatoria como resultado de un experimento.

- Una **variable aleatoria** es **discreta** si el conjunto de valores que puede tomar se encuentra en un espacio discreto, el conjunto de valores se puede numerar.
  Ejemplo, el lanzamiento de un dado.
- Una **variable aleatoria** es **continua** si los valores que puede tomar se encuentra en un espacio continuo, el conjunto de valores no se puede numerar.
  Ejemplo, mediciones en un rango de números reales.

# Distribución de probabilidad

Una función de probabilidad de una variable aleatoria asigna a cada valor que la variable puede tomar un probabilidad de suceder en un experimento.

# Función de distribución acumulada

La función de distribución acumulada de una variable aleatoria $X$, es una función que asigna a cada uno de sus valores $x$ la probabilidad de que la variable $X$ asuma un valor igual o menor a $x$. Matemáticamente:

$$F(x) = P(X \leq x)$$

# Función de probabilidad

En muchos casos es más importante conocer la probabilidad de una variable aleatoria tome un valor particular. Para esto usamos las funciones de masa o densidad.

- Para una variable **discreta**. La función de masa de probabilidad para una variable discreta es la función que mapea la probabilidad de la variable $X$ tome el valor $x$ en algún experimento. Matemáticamente:

$$F(x) = P(X = x)$$

- Para una variable **continua**: La función de densidad de probabilidad para una variable continua es la función que mapea la probabilidad de que la variable $X$ se encuentre en un intervalo [a, b].
  Matemáticamente:

$$P(a \leq X \leq b)$$

# Distribución binomial

Una variable aleatoria binomial cuenta el número de éxitos al repetir $n$ experimentos con dos resultados posibles (éxito o fracaso), cada experimento con una probabilidad $p$ de ser exitoso (y $1 - p$ de fracasar).

Función de probabilidad de una variable aleatoria binomial:

$$B(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

$n$ - número de experimentos.
$p$ - probabilidad de experimento exitoso.

# Distribución uniforme

Los valores de una variable aleatoria uniforme tienen la misma probabilidad de presentarse en cualquier experimento.

Función de densidad de una variable aleatoria uniforme continua en un intervalo $[A, B]$:

$$U(x; A, B) = \begin{cases} \dfrac{1}{B - A} & A \leq x \leq B \\ 0 & \text{cualquier otro caso.} \end{cases}$$

# Distribución normal

Los valores de una variable aleatoria normal tienen mayor probabilidad de estar cerca de la media y se dispersan disminuyendo su probabilidad al alejarse de ésta de acuerdo con su desviación estándar.

Función de densidad de una variable aleatoria normal con media $\mu$ y varianza $\sigma^2$:

$$N(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}), \quad -\infty < x < \infty$$

# Teorema del límite central

El teorema del límite central indica que la suma o la media de $n$ variables aleatorias independientes se aproxima a una distribución normal cuando el tamaño de $n$ aumenta. Matemáticamente:
Sean $X_1, X_2 \dots X_n$ variables aleatorias independientes con media $u$ y varianza finita $\sigma^2$ y $\bar{X} = \sum_{i=1}^{n} X_i/n$, entonces:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

cuando $n$ es lo suficientemente grande.

Este teorema es importante porque en la práctica muchos de los variables observadas pueden interpretarse como la suma de múltiples variables aleatorias independientes. Esto justifica que estimadores de una distribución normal puedan ser utilizados en muestras de gran tamaño.

# Estimadores tendencia central

- Media aritmética:

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

  Para la población ($\mu$) o muestra ($\bar{x}$)

- Mediana:

$$x(\frac{n+1}{2})\text{si } n \text{ impar} \qquad \frac{x(\frac{n+1}{2}) + x(\frac{n}{2})}{2}\text{si } n \text{ par}$$

- Moda:
  Es el valor con más ocurrencias en la muestra o población.

# Estimadores dispersión

- Varianza:

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

Para la población $\sigma^2$ o una muestra $s^2$.

- Desviación estándar:

$$\sigma = \sqrt{\sigma^2} \qquad \text{o} \qquad s = \sqrt{s^2}$$

Para la población $\sigma$ o una muestra $s$.