

## Conceptos Esenciales

- Una frecuencia es el número de veces que un valor de datos ocurre. Una tabla de frecuencia lista los diferentes resultados y sus frecuencias. Una frecuencia relativa es la relación (fractura o proporción) de la frecuencia de un resultado de los datos con el total del número de observaciones.
- Representaciones visuales de variables categorical
  - Gráficas de barra son mostradores visuales de datos en cual la frecuencia de cada categoría listada al otro lado del axis horizontal es indicada por la altura de su barra rectangular correspondiente (o la longitud de la gráfica está desplegada horizontalmente).
  - Gráficas circulares despliegan datos en una gráfica redonda, dividido en “piezas de tarta”, cada una representando la portion de la cantidad o frecuencia que representa. En otras palabras, gráficas circulares despliegan la frecuencia relativa de tu conjunto de datos.
- Una tabla contingencia es una tabla que despliega los resultados de dos frecuencias relativas de tu conjunto de datos.
- Gráficas circulares y gráficas de barra son buenas representaciones visuales de variables categóricas de una sola población o grupo. Pero, ¿qué podemos hacer si queremos comparar una variable categórica en varios grupos?
  - Gráficas de barras apiladas despliegan valores para diferentes categorías pero sino más bien de demostrar una barra diferente para cada categoría, demuestran subcategorías como segmentos con cada barra.
  - Gráficas de barra de lado a lado presentan datos para dos variables categóricas para más de un grupo por creando dos barras en la gráfica para cada grupo - una barra para cada variable.
- Cuando es presentado con grandes conjuntos de datos, gráficas de puntos a veces es incómodo para juntar. En adicción, no es la mejor forma para presentar los datos. Para grandes conjuntos de datos, un histograma puede presentar numerosos puntos de datos más fácilmente que barras en cambio de inmensa cantidad de puntos de datos en una gráfica de puntos .
  - Histogramas son buenas opciones para mostrar conjuntos de datos que tienen un gran número de observaciones ya que agrupan observaciones en “contenedores” de tamaño iguales.
  - Gráficas de puntos muestran cuántas observaciones individuales hay de cada variable observada. Cada observación en el conjunto de datos aparece como su propio punto en la gráfica.
- Las características para describir la distribución de un variable cuantitativo son forma, centro, el extiende, y la presencia de los valores atípicos.
  - Para describir la forma de una distribución, imagina dibujando el esbozo de los datos para enfatizar la tendencia general.
    - Sesgada izquierda: un racimo de datos en el lado derecho con una cola de datos estrechándose a la izquierda.

- Simétrico con un pico central (también llamado forma campana): el lado izquierdo y derecho de la distribución (cercanamente) se reflejan uno a otro.
- Sesgada derecha: un racimo de datos en el lado izquierdo con una cola de datos estrechándose a la derecha.
- La forma también se puede describir por el número de picos.
  - Unimodal: hay un pico prominente
  - Bimodal: hay dos picos prominente
  - Multimodal: hay tres o mas picos prominentes
  - Uniforme: no hay picos prominentes. Una forma rectangular, la misma cantidad de datos para cada valor variable
- El centro describe la locación del medio de distribución.
- El extiende es la medida de cuanto los valores en un conjunto de datos tienden a diferir o variar de uno a otro.
  - Una forma como podemos medir el extiende es por encontrar el variable minimo y maximo de los datos y calculando la diferencia entre los dos. Esta diferencia se llama el rango.
- Las partes aisladas son observaciones en los datos que son inusuales y fuera del patrón general del resto de observaciones en la distribución.

## Ecuaciones clave

$$\text{frecuencia relativa} = \frac{\text{grupo conjunto}}{\text{total}}$$

## Glossario

### grafica de barra

una gráfica en cual las categorías son representadas por barras cuales están separadas de unas de otras

### forma campana

el lado izquierdo y derecho de la distribución cercanamente se reflejan uno al otro

### bimodal

dos picos prominente en los datos

### centro

la locación del medio de la distribución

**tabla de contingencia (tabla de dos lados)**

tabla que despliega el resultado de dos variables categóricas simultáneamente

**grafica de puntos**

despliega cuantas observaciones individuales hay para cada observación variable

**tabla de frecuencia**

lista el número de observaciones (la frecuencia) de cada valor único de un variable

**histograma**

despliega los conjuntos de datos cuales tienen un gran número de observaciones por agrupando “contenedores” de tamaño iguales

**sesgada izquierda**

un racimo de datos en el lado derecho con una cola de datos estrechándose a la izquierda

**multimodal**

tres o más picos prominentes en los datos

**unidades observacionales**

contiene información de un grupo de individuales/cosas

**valores atipicos**

observaciones en los datos cuales son inusual y fuera de el patrón general del resto de observaciones en la distribución

**grafico circular**

gráfica donde las categorías son representadas por porciones en un círculo y son proporcionales en tamaño al porcentaje de individuos/articulo en cada categoría

**poblacion**

gente o entidades como animales o objetos

### **rango**

la diferencia entre el el valor mínimo y máximo de los datos

### **frecuencia relativa**

representa la proporción de observaciones que están en una categoría particular y puede ser expresada como un decimal o porcentaje

### **sesgada derecha**

un racimo de datos en el lado izquierdo con una cola de datos estrechándose a la derecha

### **sample**

un grupo de individuales o entidades donde recolectamos datos

### **forma**

patrones en general y el número de picos en los datos

### **gráfica de barra de lado a lado, gráfica de barra apilada**

una extensión de gráfica de barra que nos deja conducir comparaciones entre múltiples de conjuntos de datos

### **el extender (diferencia)**

la medida de cuanto los valores en un conjunto de datos tienden a diferir o variar de uno a otro

### **uniforme**

no hay picos prominentes en los datos

### **unimodal**

hay un pico prominente en los datos

### **variables**

unidades observacionales son grabadas

