

M2 - Análisis de Datos: descriptivo e inferencial

Dora Suárez, Juan F. Pérez

Departamento MACC
Matemáticas Aplicadas y Ciencias de la Computación
Universidad del Rosario

juanferna.perez@urosario.edu.co

Primer Semestre de 2019

Contenidos

- 1 Datos: una característica de una población
- 2 Variable aleatoria
- 3 Datos: una característica de una muestra

Datos: una característica de una población

Análisis descriptivo de datos

- Tomamos datos respecto a una característica de una población

Análisis descriptivo de datos

- Tomamos datos respecto a una característica de una población
- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso

Análisis descriptivo de datos

- Tomamos datos respecto a una característica de una población
- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- Población: los estudiantes de este curso

Análisis descriptivo de datos

- Tomamos datos respecto a una característica de una población
- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- Población: los estudiantes de este curso
- Característica: número de tazas de café que tomaron ayer

Análisis descriptivo de datos

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso

Análisis descriptivo de datos

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- $2, 0, 1, \dots$

Análisis descriptivo de datos

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- 2, 0, 1, ...
- Medidas descriptivas:
 - Mínimo, máximo

Análisis descriptivo de datos

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- 2, 0, 1, ...
- Medidas descriptivas:
 - Mínimo, máximo
 - Promedio

Análisis descriptivo de datos

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- 2, 0, 1, ...
- Medidas descriptivas:
 - Mínimo, máximo
 - Promedio
 - Variabilidad alrededor de la media

Análisis descriptivo de datos

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- 2, 0, 1, ...
- Medidas descriptivas:
 - Mínimo, máximo
 - Promedio
 - Variabilidad alrededor de la media
 - Frecuencia (absoluta, relativa)

Análisis descriptivo de datos (ejemplo)

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso

Análisis descriptivo de datos (ejemplo)

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- $[0, 1, 0, 2, 3, 0, 1, 2, 0, 2, 2, 2, 1, 0, 3, 4]$

Análisis descriptivo de datos (ejemplo)

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- $[0, 1, 0, 2, 3, 0, 1, 2, 0, 2, 2, 2, 1, 0, 3, 4]$
 - Número de observaciones: 16

Análisis descriptivo de datos (ejemplo)

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- $[0, 1, 0, 2, 3, 0, 1, 2, 0, 2, 2, 2, 1, 0, 3, 4]$
 - Número de observaciones: 16
 - Mínimo: 0

Análisis descriptivo de datos (ejemplo)

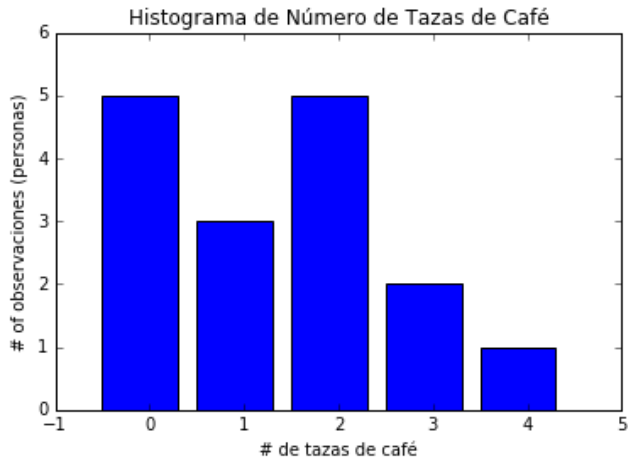
- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- $[0, 1, 0, 2, 3, 0, 1, 2, 0, 2, 2, 2, 1, 0, 3, 4]$
 - Número de observaciones: 16
 - Mínimo: 0
 - Máximo: 4

Análisis descriptivo de datos (ejemplo)

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso
- $[0, 1, 0, 2, 3, 0, 1, 2, 0, 2, 2, 2, 1, 0, 3, 4]$
 - Número de observaciones: 16
 - Mínimo: 0
 - Máximo: 4
 - Promedio: 1.4375

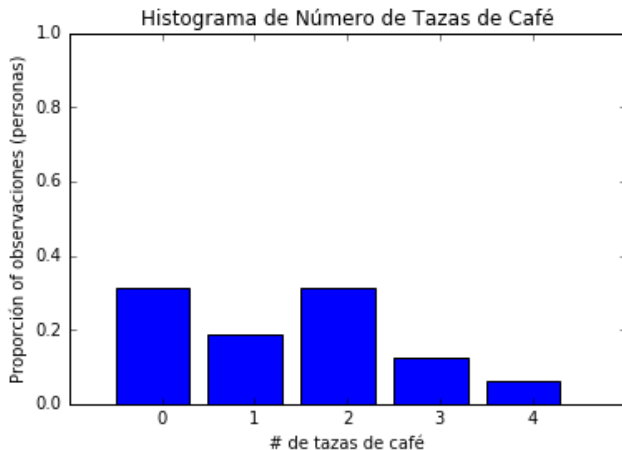
Análisis descriptivo de datos (ejemplo)

Frecuencia absoluta:



Análisis descriptivo de datos (ejemplo)

Frecuencia relativa:



Generalizando a cualquier conjunto de datos

- Datos unidimensionales: una sola característica

Generalizando a cualquier conjunto de datos

- Datos unidimensionales: una sola característica
- Número de datos: N

Generalizando a cualquier conjunto de datos

- Datos unidimensionales: una sola característica
- Número de datos: N
- **Datos:** $[x_1, x_2, \dots, x_N]$

Generalizando a cualquier conjunto de datos

- Datos unidimensionales: una sola característica
- Número de datos: N
- **Datos:** $[x_1, x_2, \dots, x_N]$
- **Mínimo:** $\min_i \{x_i\}$

Generalizando a cualquier conjunto de datos

- Datos unidimensionales: una sola característica
- Número de datos: N
- **Datos:** $[x_1, x_2, \dots, x_N]$
- **Mínimo:** $\min_i \{x_i\}$
- **Máximo:** $\max_i \{x_i\}$

Generalizando a cualquier conjunto de datos

- Datos unidimensionales: una sola característica
- Número de datos: N
- **Datos:** $[x_1, x_2, \dots, x_N]$
- **Mínimo:** $\min_i \{x_i\}$
- **Máximo:** $\max_i \{x_i\}$
- **Promedio:**

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

Generalizando a cualquier conjunto de datos

- **Datos:** $[x_1, x_2, \dots, x_N]$

Generalizando a cualquier conjunto de datos

- **Datos:** $[x_1, x_2, \dots, x_N]$
- **Frecuencia absoluta:**
número de veces que aparece j en las observaciones

$$f_j = \#\{x_i = j\}$$

Generalizando a cualquier conjunto de datos

- **Datos:** $[x_1, x_2, \dots, x_N]$
- **Frecuencia absoluta:**
número de veces que aparece j en las observaciones

$$f_j = \#\{x_i = j\}$$

- **Frecuencia relativa:**
proporción del veces que aparece j en las observaciones

$$p_j = \frac{f_j}{N}$$

Variable aleatoria

Variable aleatoria

- X : número de tazas de café que tomó ayer un estudiante de este curso

Variable aleatoria

- X : número de tazas de café que tomó ayer un estudiante de este curso
- Seleccionamos un estudiante al azar, ¿cuál es la probabilidad de que el estudiante haya tomado 3 tazas de café?

Variable aleatoria

- X : número de tazas de café que tomó ayer un estudiante de este curso
- Seleccionamos un estudiante al azar, ¿cuál es la probabilidad de que el estudiante haya tomado 3 tazas de café?

■

$$P(X = 3) = p_3 = \frac{2}{16} = 0,125$$

Variable aleatoria

- Se realiza un **experimento aleatorio**

Variable aleatoria

- Se realiza un **experimento aleatorio**
- **Espacio muestral**: conjunto de todos los posibles resultados

Variable aleatoria

- Se realiza un **experimento aleatorio**
- **Espacio muestral**: conjunto de todos los posibles resultados
- **Variable aleatoria**: función del espacio muestral en los números reales (asignamos valores numéricos al resultado del experimento)

Variable aleatoria (ejemplo)

- **Experimento aleatorio:** los estudiantes de este curso toman un número de tazas de café

Variable aleatoria (ejemplo)

- **Experimento aleatorio:** los estudiantes de este curso toman un número de tazas de café
- **Espacio muestral:** todas las posibles combinaciones de número de tazas de café consumidas por cada estudiante

Variable aleatoria (ejemplo)

- **Experimento aleatorio:** los estudiantes de este curso toman un número de tazas de café
- **Espacio muestral:** todas las posibles combinaciones de número de tazas de café consumidas por cada estudiante
- **Variable aleatoria:** el número de tazas que consumió un estudiante seleccionado al azar

Variable aleatoria discreta

- **Variable aleatoria** X : número de tazas de café que tomó ayer un estudiante de este curso

Variable aleatoria discreta

- **Variable aleatoria** X : número de tazas de café que tomó ayer un estudiante de este curso
- **Discreta**: toma valores en un conjunto finito (o contable)

Variable aleatoria discreta

- **Variable aleatoria** X : número de tazas de café que tomó ayer un estudiante de este curso
- **Discreta**: toma valores en un conjunto finito (o contable)
- $X \in \{0, 1, 2, 3, \dots\}$

Variable aleatoria discreta

- **Variable aleatoria** X : número de tazas de café que tomó ayer un estudiante de este curso
- **Discreta**: toma valores en un conjunto finito (o contable)
- $X \in \{0, 1, 2, 3, \dots\}$
- Función de masa de probabilidad:

$$p_j = P(X = j)$$

Variable aleatoria discreta

- **Variable aleatoria** X : número de tazas de café que tomó ayer un estudiante de este curso
- **Discreta**: toma valores en un conjunto finito (o contable)
- $X \in \{0, 1, 2, 3, \dots\}$
- Función de masa de probabilidad:

$$p_j = P(X = j)$$

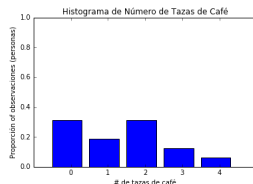
- p_j : probabilidad de que X sea igual a j

Variable aleatoria discreta (ejemplo)

- **Variable aleatoria** X : número de tazas de café que tomó ayer un estudiante de este curso

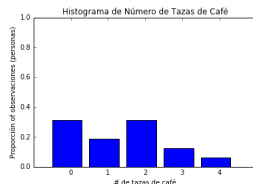
Variable aleatoria discreta (ejemplo)

- **Variable aleatoria X :** número de tazas de café que tomó ayer un estudiante de este curso



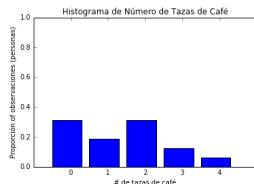
Variable aleatoria discreta (ejemplo)

- **Variable aleatoria X :** número de tazas de café que tomó ayer un estudiante de este curso



Variable aleatoria discreta (ejemplo)

- **Variable aleatoria** X : número de tazas de café que tomó ayer un estudiante de este curso



- Función de masa de probabilidad:

$$P(X = j) = \begin{cases} 5/16, & j = 0, \\ 3/16, & j = 1, \\ 5/16, & j = 2, \\ 2/16, & j = 3, \\ 1/16, & j = 4, \end{cases}$$

Datos: una característica de una muestra

Muestras

- Tomamos datos respecto a una característica de una población

Muestras

- Tomamos datos respecto a una característica de una población
- Ejemplo: número de tazas de café que tomaron ayer los habitantes del país

Muestras

- Tomamos datos respecto a una característica de una población
- Ejemplo: número de tazas de café que tomaron ayer los habitantes del país
- Población: los habitantes del país

Muestras

- Tomamos datos respecto a una característica de una población
- Ejemplo: número de tazas de café que tomaron ayer los habitantes del país
- Población: los habitantes del país
- Característica: número de tazas de café que tomaron ayer

Muestras

- Es infactible, muy costoso o incluso imposible medir la característica en todos los elementos de la población

Muestras

- Es infactible, muy costoso o incluso imposible medir la característica en todos los elementos de la población
- Ejemplos:

Muestras

- Es infactible, muy costoso o incluso imposible medir la característica en todos los elementos de la población
- Ejemplos:
 - Grandes poblaciones

Muestras

- Es infactible, muy costoso o incluso imposible medir la característica en todos los elementos de la población
- Ejemplos:
 - Grandes poblaciones
 - Características cambiantes

Muestras

- Es infactible, muy costoso o incluso imposible medir la característica en todos los elementos de la población
- Ejemplos:
 - Grandes poblaciones
 - Características cambiantes
 - Pruebas destructivas

Muestras

- Es infactible, muy costoso o incluso imposible medir la característica en todos los elementos de la población
- Ejemplos:
 - Grandes poblaciones
 - Características cambiantes
 - Pruebas destructivas
- **Muestra:** selección de un número limitado de observaciones de la población (no todas)

Muestra aleatoria

- Muestra debe ser representativa de la población

Muestra aleatoria

- Muestra debe ser representativa de la población
- Ejemplo: los estudiantes de este curso no son una muestra representativa

Muestra aleatoria

- Muestra debe ser representativa de la población
- Ejemplo: los estudiantes de este curso no son una muestra representativa
- Selección al azar: todos los miembros de la población tienen la misma probabilidad de ser seleccionados

Muestra aleatoria

- Muestra debe ser representativa de la población
- Ejemplo: los estudiantes de este curso no son una muestra representativa
- Selección al azar: todos los miembros de la población tienen la misma probabilidad de ser seleccionados
- Si la característica poblacional es una variable aleatoria X ...

Muestra aleatoria

- Muestra debe ser representativa de la población
- Ejemplo: los estudiantes de este curso no son una muestra representativa
- Selección al azar: todos los miembros de la población tienen la misma probabilidad de ser seleccionados
- Si la característica poblacional es una variable aleatoria X ...
- ... obtenemos una muestra aleatoria de tamaño n : X_1, \dots, X_n

Muestra aleatoria (ejemplo)

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso (población de tamaño 16)

Muestra aleatoria (ejemplo)

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso (población de tamaño 16)
- X : número de tazas de café que tomó ayer un estudiante de este curso

$$P(X = j) = \begin{cases} 5/16, & j = 0, \\ 3/16, & j = 1, \\ 5/16, & j = 2, \\ 2/16, & j = 3, \\ 1/16, & j = 4, \end{cases}$$

Muestra aleatoria (ejemplo)

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso (población de tamaño 16)
- X : número de tazas de café que tomó ayer un estudiante de este curso

$$P(X = j) = \begin{cases} 5/16, & j = 0, \\ 3/16, & j = 1, \\ 5/16, & j = 2, \\ 2/16, & j = 3, \\ 1/16, & j = 4, \end{cases}$$

- Tomamos una muestra aleatoria de tamaño 3: X_1, X_2, X_3

Muestra aleatoria (ejemplo)

- Ejemplo: número de tazas de café que tomaron ayer los estudiantes de este curso (población de tamaño 16)
- X : número de tazas de café que tomó ayer un estudiante de este curso

$$P(X = j) = \begin{cases} 5/16, & j = 0, \\ 3/16, & j = 1, \\ 5/16, & j = 2, \\ 2/16, & j = 3, \\ 1/16, & j = 4, \end{cases}$$

- Tomamos una muestra aleatoria de tamaño 3: X_1, X_2, X_3
- Cada una de las tres observaciones tiene la misma *distribución* de X

Muestra aleatoria: inferencias

- A partir de la **muestra** buscamos **inferir** características de la **población**

Muestra aleatoria: inferencias

- A partir de la **muestra** buscamos **inferir** características de la **población**
- Muestra es una visión parcial de la población: incertidumbre

Muestra aleatoria: inferencias

- A partir de la **muestra** buscamos **inferir** características de la **población**
- Muestra es una visión parcial de la población: incertidumbre
- Al aumentar el tamaño de la muestra (aleatoria) la incertidumbre debe reducirse

Inferencias a partir de muestras aleatorias

- Característica de la **Población**: variable aleatoria X

Objetivo: a partir de estas cantidades medidas sobre la muestra inferir el comportamiento de la población

Inferencias a partir de muestras aleatorias

- Característica de la **Población**: variable aleatoria X
- **Muestra aleatoria** de tamaño n : X_1, \dots, X_n

Objetivo: a partir de estas cantidades medidas sobre la muestra inferir el comportamiento de la población

Inferencias a partir de muestras aleatorias

- Característica de la **Población**: variable aleatoria X
- **Muestra aleatoria** de tamaño n : X_1, \dots, X_n
- **Datos**: $[x_1, x_2, \dots, x_n]$

Objetivo: a partir de estas cantidades medidas sobre la muestra inferir el comportamiento de la población

Inferencias a partir de muestras aleatorias

- Característica de la **Población**: variable aleatoria X
- **Muestra aleatoria** de tamaño n : X_1, \dots, X_n
- **Datos**: $[x_1, x_2, \dots, x_n]$
- Mínimo, Máximo

Objetivo: a partir de estas cantidades medidas sobre la muestra inferir el comportamiento de la población

Inferencias a partir de muestras aleatorias

- Característica de la **Población**: variable aleatoria X
- **Muestra aleatoria** de tamaño n : X_1, \dots, X_n
- **Datos**: $[x_1, x_2, \dots, x_n]$
- Mínimo, Máximo
- Promedio

Objetivo: a partir de estas cantidades medidas sobre la muestra inferir el comportamiento de la población

Inferencias a partir de muestras aleatorias

- Característica de la **Población**: variable aleatoria X
- **Muestra aleatoria** de tamaño n : X_1, \dots, X_n
- **Datos**: $[x_1, x_2, \dots, x_n]$
- Mínimo, Máximo
- Promedio
- Frecuencia absoluta

Objetivo: a partir de estas cantidades medidas sobre la muestra inferir el comportamiento de la población

Inferencias a partir de muestras aleatorias

- Característica de la **Población**: variable aleatoria X
- **Muestra aleatoria** de tamaño n : X_1, \dots, X_n
- **Datos**: $[x_1, x_2, \dots, x_n]$
- Mínimo, Máximo
- Promedio
- Frecuencia absoluta
- Frecuencia relativa

Objetivo: a partir de estas cantidades medidas sobre la muestra inferir el comportamiento de la población