

## **Guía de Ejercicios N°2**

En esta guía se dará comienzo al **Análisis de Datos Univariado**. Para ello se requiere que el estudiante aprenda a:

1. Reconocer escala de medición de una variable.
2. Armar tablas de frecuencia según el tipo de variable.
3. Interpretar la información que proporcionan las diferentes frecuencias.
4. Saber aplicar correctamente medidas de resumen según el tipo de variable.
5. Interpretar la información que brindan las medidas de resumen.
6. Determinar características de la distribución de la variable, con enfoque en comparación con la distribución normal.
7. Definir qué medida (si la media, la mediana, o moda) resulta mejor para representar la variable.
8. Reconocer outliers, y decidir su tratamiento.

- 1) Dado el dataset limpio del departamento asignado a su grupo, trabajado en la primera clase de práctica, descargue el mismo con la librería writexl.
- 2) En un archivo de jupyter nuevo, cargué el nuevo dataset y aplicando a las subdivisiones que usted realizó (primaria-secundaria) utilice la función frq para ver el estado en general en que quedo su dataset nuevo. Conteste:
  - a) ¿Quedan variables que deberían eliminarse? Si la respuesta es afirmativa indique el motivo.
  - b) De las asignaturas de interés en el estudio en los cual se normalizaron los nombres, ¿quedaron todas bien cargadas?
  - c) ¿La proporción de datos mal cargados en las notas es significativo?
- 3) A partir del dataset del departamento asignado a su grupo, tomando como punto de partida variables cualitativas realice una tabla de frecuencias completa, y conteste:
  - a) ¿Cómo es la distribución de estudiantes por nivel en el departamento en general?, ¿es similar la distribución en primaria y secundaria?
  - b) ¿Cómo es la distribución por turno en el departamento en general? ¿Cuál es la distribución por turno si se segmenta en primaria y secundaria? Si, se mantiene la tendencia actual de estudiantes en primaria, el sistema secundario ¿deberá ampliar o reducir sus bancos por curso?
  - c) ¿Cómo es la distribución por sector público y privado si se segmenta en primaria y secundaria?
  - d) ¿Cómo es distribuye en su departamento la cantidad de estudiantes por año de cursado? ¿cuál es el año con mayor matrícula? ¿cuál es el año con menor matrícula?
  - e) ¿Hay otras variables cualitativas que puedan resultan de interés analizar?

- f) ¿La información de a) b) c) d) se puede contrastar con la base de secundaria y primaria? De ser así indique las que se pudieron contrastar y si las distribuciones son similares.
- g) ¿Hay alguna otra información que se le ocurra se pueda obtener a partir de este análisis?
- 4) A partir del dataset del departamento asignado a su grupo, segmentando por nivel, año, y materia (centraremos nuestra mirada en los promedios de Matemática y Lengua para comenzar), tomando como punto de partida la variable cuantitativa promedio realice una tabla de frecuencias completa, y conteste:
- a) ¿En qué rango se encuentran la mayoría de las calificaciones?
  - b) ¿Qué calificaciones son las más inusuales?
  - c) El 70% de los estudiantes obtuvo menos de...?
  - d) ¿Qué porcentaje de estudiantes obtuvo notas menores a 4?
  - e) ¿Qué porcentaje de estudiantes obtuvo notas menores a 6?
  - f) ¿Qué porcentaje de estudiantes aprobó la asignatura, si se considera aprobado si el porcentaje es mayor a 6? ¿Esto es así en el sistema educativo primario o secundario de Entre Ríos?
  - g) ¿Qué porcentaje de estudiantes obtuvo notas mayores a 8?
  - h) Observando los resultados, ¿considera que el rendimiento en la asignatura fue óptimo?
- 5) A partir del dataset del departamento asignado a su grupo, segmentando por nivel, año, y materia (centraremos nuestra mirada en los promedios de Matemática y Lengua para comenzar), tomando como punto de partida la variable cuantitativa promedio realice un resumen de estadísticos, y conteste:
- a) ¿Se corresponden los valores de Media y Mediana? ¿Y la Moda?
  - b) ¿Cuál es el promedio medio de la muestra seleccionada? ¿La media es representativa de la muestra o es conveniente elegir otra medida?
  - c) ¿Qué podría decir de la distribución de esta variable? ¿Está centrada?
  - d) ¿Qué puede decir de la dispersión de los datos?
  - e) ¿Hay outliers? ¿Son valores atípicos? ¿Son valores extremos?
  - f) ¿A partir de qué valor se considera un promedio outliers?
  - g) Si se descarta el 25% de los sueldos menores y el 25% de los ingresos mayores, ¿entre qué valores se concentran el 50% de los promedios?
  - h) ¿El 75% de las notas son menores que...?
  - i) ¿El 50% de las notas son mayores que...?
  - j) Si se comparan los datos de Lengua y Matemática, ¿cuál tiene mayor variabilidad?
  - k) ¿Hay alguna otra información que pueda brindar a partir de las medidas de resumen estudiadas?
  - l) ¿Existen otras variables cuantitativas en las que se deba realizar este mismo análisis para extraer información útil a la departamental?
- 6) Escriba un texto donde resuma la información obtenida, a partir del análisis realizado en esta guía.