

Taller Estadística Descriptiva

Universidad Externado de Colombia

Departamento de Matemáticas

Estadística 2

Ejercicios:

Considere la base de datos dada en **saber11_2015_1.csv** correspondiente a la información oficial de las pruebas saber 11 correspondientes al primer semestre del año 2015. Las variables del conjunto de datos son: código del municipio; nombre del municipio; departamento; calendario; naturaleza; jornada; número de estudiantes evaluados; calificación promedio de la institución en: lectura crítica, matemáticas, sociales, ciencias naturales, inglés, razonamiento cuantitativo, competencias ciudadanas.

1. ¿Cuántos registros tienen datos faltantes? Remover todos los registros con datos faltantes.
2. ¿Cuántas instituciones educativas? ¿Cuántas variables?
3. Clasificar las variables según su naturaleza y escala de medición.
4. ¿Cuántos municipios se consideran en esta base de datos?
5. ¿Cuál es la cantidad total de estudiantes evaluados?
6. Identificar los cinco municipios con mayor cantidad de instituciones educativas.
7. Identificar la(s) institución(es) con mayor cantidad de estudiantes. ¿Cuáles son la jornada, el calendario, la naturaleza y el municipio correspondientes?
8. Hacer la distribución de frecuencias de la jornada de las instituciones educativas. Hacer el gráfico de barras correspondiente. Comentar brevemente los resultados obtenidos.
9. Hacer la distribución de frecuencias de la naturaleza de las instituciones educativas. Hacer el gráfico de barras correspondiente. Comentar brevemente los resultados obtenidos.

10. Hacer la distribución de frecuencias (relativas) conjunta de la jornada y la naturaleza de las instituciones educativas. Hacer el gráfico de barras correspondiente. Comentar brevemente los resultados obtenidos.
11. Obtener los perfiles fila y columna de la tabla anterior. Comentar brevemente los resultados obtenidos.
12. Calcular e interpretar el mínimo, el máximo, la media, la mediana y los cuartiles del promedio de lectura crítica de las instituciones educativas.
13. Calcular e interpretar el rango, el rango intercuartílico y el coeficiente de variación del promedio de lectura crítica de las instituciones educativas.
14. Comparar gráfica y numéricamente la distribución del promedio de lectura crítica y del promedio de matemáticas de todas las instituciones educativas.
15. Comparar gráfica y numéricamente las instituciones oficiales y no oficiales en términos del promedio de lectura crítica.
16. Hacer los dispersogramas (en un solo gráfico) y la matriz de correlaciones entre el promedio de lectura, matemáticas y sociales de las instituciones educativas. Comentar brevemente los resultados obtenidos.