

TAREA # 1

Por ningún motivo se recibirán entregas posteriores a la hora límite. En caso de no realizar la entrega la nota asignada será 0.

Notas: La entrega del taller deberá hacerse en un informe autocontenido incluyendo las gráficas, interpretación y conclusiones. Adicionalmente debe entregarse el código utilizado en un archivo ipynb.

Para todos los literales que lo requieran, se recomienda utilizar una significancia del 5% (confianza del 95%). En caso de utilizar otro nivel de significancia, hacerlo explícito en el informe.

1. (40%) La base de datos contenida en resultadosicfes.csv tiene información de los resultados obtenidos en 2019, estos resultados muestran todos los departamentos de Colombia, por favor tomar como muestra los resultados de los siguientes 5 departamentos; La Guajira, Valle, Atlántico, Bogotá (la cual en las estadísticas se tiene como departamento) y Huila. El objetivo es realizar un análisis exploratorio de los datos.

- a. (20%) Identifique las variables continuas y para CADA UNA de éstas:
 - Determine cuántos valores perdidos tiene cada una.
 - Calcule la media, desviación estándar, máximo y mínimo.
 - Identifique los departamentos donde los estudiantes obtuvieron un puntaje en matemáticas mayor a 55, un puntaje en lectura mayor a 60 y un puntaje en ingles mayor a 70 reemplace estos valores por valores perdidos. En lo que sigue trabaje con estas variables corregidas.
 - Calcule nuevamente la media, desviación estándar, máximo, mínimo. Compare los resultados con los hallados los 2 literales anteriores y comente.
- b. (5%) Compare en una tabla los puntajes de matemáticas, lectura y ciencias naturales, e ilustre gráficamente con gráficas de caja para cada una de las materias. Comente.
- c. (5%) Identifique el departamento que tiene los puntajes globales más altos. Identifique el que tiene los mejores puntajes en matemáticas.
- d. (5%) Muestre en una tabla el número de estudiantes que pertenece a colegios oficiales y no oficiales. Ilustre mediante una gráfica. ¿Cree Ud. que los resultados del puntaje global se ven influenciado por el tipo de colegio?
- e. (5%) ¿Cree usted que hay alguna relación entre el puntaje global y el estrato del estudiante?

2. (30%) Teniendo en cuenta los resultados obtenidos por los estudiantes en sus pruebas ICFES una academia dedicada a cursos pre-icfes decidió ofertar una variedad de cursos con el fin de lograr que los estudiantes obtengan puntajes mayores a 400 puntos. La academia optó por desarrollar 4 cursos distintos, estos están pensados variando la intensidad de horas de la siguiente manera; intensivo (2 horas diarias), semi-intensivo (4 horas semanales), semanal (5 horas en un solo día) y super intensivo (3 horas diarias). En la base de datos cursos.csv, encontrara los resultados obtenidos por los estudiantes de acuerdo con el curso tomado. Con base en la información, conteste las siguientes preguntas:

- a. (10%) Realice estadísticas descriptivas (promedio, frecuencia, desviación, entre otros) que permitan observar diferencias de los puntajes obtenidos por los estudiantes según el curso tomado y concluya.
- b. (10%) Un estudiante que se encuentra indeciso por el curso a tomar considera que el curso intensivo y el curso super intensivo no tienen ninguna diferencia. ¿Está de acuerdo con esta afirmación?
- c. (10%) Uno de los profesores de la academia desea conocer los puntajes promedio de los estudiantes con el fin de establecer estrategias para que algún estudiante tenga un puntaje mayor a 483 que es el mejor de la academia, ¿cree usted que el profesor puede ayudar a sus estudiantes sugiriéndoles que tomen el curso semanal? ¿Qué curso cree usted que debe sugerir el profesor a sus estudiantes teniendo en cuenta los resultados que se han obtenido?

3. (30%) Battlife se dedica a la producción de baterías, se sabe la vida útil promedio de sus baterías siguen una distribución normal con una media de 800 horas y una varianza σ^2 . Un grupo de ingenieros quieren verificar que esta afirmación sea veraz, por lo tanto deciden tomar una muestra aleatoria de 30 baterías con las cuales obtienen que la vida útil promedio es de 790 horas, con una desviación estándar muestral de 50 horas.

Teniendo en cuenta lo anterior, por favor resuelva las siguientes preguntas:

- a. (10%) ¿Cuál es la probabilidad de que la vida útil promedio real de las baterías sea menor a 780 horas?
- b. (10%) Un comprador está interesado en adquirir un lote de baterías, sin embargo, solicitó que, en una muestra de 40 baterías, la variabilidad de las vidas útiles no debe exceder 2500 horas. Calcule la probabilidad de que la varianza de las vidas útiles de una muestra de 40 baterías supere este valor.
- c. (10%) El área de desarrollo esta haciendo un estudio de mercado dado que desean que las baterías producidas sean las que tengan la mejor vida útil del mercado. Por lo cual, se han obtenido los resultados de una empresa competidora "BattteriesInc". De "BattteriesInc". se conoce que la vida útil promedio de las baterías es una variable aleatoria que sigue una distribución Normal con media μ y varianza 0.005. Se tomó una muestra aleatoria de 20 baterías de Battlife y 15 baterías de BattteriesInc. Asumiendo que las variables aleatorias son independientes, ¿cuál es la probabilidad de que la varianza muestral de la vida útil producida por Battlife menor a la varianza muestral del espesor de la vida útil de BattteriesInc?