

**Proyecto grupal
Curso 2023/2024
Parte de Inferencia Estadística**

Instrucciones

- Se trabajará con el conjunto de datos proporcionado por las profesoras.
- Dicho conjunto de datos versa sobre el consumo de vehículos. De cada modelo de vehículo se han recogido varias variables como cilindrada, consumo, emisiones de CO₂, tipo de vehículo, fabricante.
- A cada grupo se le asignará un conjunto de datos correspondiente a un año determinado.
- Los datos se pueden encontrar en:
<https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64>
(Archivo que hay que utilizar: “**2023 Fuel Consumption Ratings (2023-08-18)**”)
- Cada alumno debe elegir al menos una variable categórica para realizar el análisis estadístico por categorías.
- Se debe crear una variable categórica llamada “país” indicando a qué país pertenece el fabricante del vehículo.
- Para realizar ciertos análisis, el alumno puede tomar todos o una parte de los datos.

Estadística Descriptiva

Instrucciones

Se procederá a realizar un análisis descriptivo básico de los datos, empleando los estadísticos más comunes, de dispersión y de centralidad (media, mediana, desviación típica, ...). El análisis se realizará de modo global y de modo específico según niveles o variables categóricas escogidas. Además, en esta etapa del estudio se deberá representar visualmente los datos, usando los distintos tipos de gráficos vistos.

Procedimiento y requisitos mínimos

La longitud de esta parte del informe (en PDF) debe ser de entre 4 y 6 páginas y deberá contener como mínimo:

- **Estadísticos descriptivos básicos totales:** media, mediana, desviación típica, y cuartiles; y por niveles según la variable categóricas escogida.
- **Gráficas que representen e ilustren aspectos significativos del conjunto de datos.** Pueden emplearse: histogramas, gráficos Q- Q y/o gráficos de probabilidad para comparar muestras, box-plots, diagramas de dispersión ... del conjunto total de los datos y de algunos subgrupos o niveles concretos de la población.

La discusión y la visualización de los datos obtenidos es donde recaerá la mayor parte del peso de la evaluación.

Inferencia

Instrucciones

La última parte se basa en realizar diversos análisis inferenciales a partir los datos analizados, complementando el anterior análisis exploratorio. Esta parte del informe (en PDF) ocupará una longitud entre 5 y 7 páginas.

En la discusión y análisis de los datos obtenidos es dónde recaerá la mayor parte del peso de la evaluación. Se valorarán todos aquellos análisis extras realizados al margen de los requisitos mínimos.

Procedimiento y requisitos mínimos

- **Distribuciones y contrastes de bondad de ajuste:** a la vista de los histogramas representados en la parte descriptiva, se deberá proponer un modelo de probabilidad a la que se ajuste una de las variables dadas (normal, gamma, exponencial ...), estimando los parámetros necesarios (media y Varianza en caso de distribución normal, λ en caso de exponencial, etc). Y, para comprobarlo, se debe realizar un contraste sobre la bondad del ajuste mediante la distribución chi-cuadrado de forma teórica según lo visto en la asignatura de Inferencia estadística.
- **Intervalos de confianza:** se incluirán intervalos de confianza para la media y la varianza de una de las variables numéricas escogidas. Se calcularán ambos intervalos de forma teórica, según lo visto en la asignatura de Inferencia estadística.
- **Contraste de hipótesis:** deberá realizarse algún contraste de hipótesis, en función de las de las variables escogidas, y de las observaciones recogidas en la parte descriptiva. Por ejemplo, se puede contrastar la diferencia de medias en el precio entre distintos tipos de vehículos, o diferencia entre el consumo de coches de distinto país.

Observaciones

Para la realización de los cálculos puede utilizarse alguna herramienta informática, tipo Excel o Python, a elección de cada persona. Tiene que quedar explicado y plasmado, de manera clara, qué cálculo se ha hecho en cada paso.

La entrega constará de:

- un único **informe PDF** relatando todo el desarrollo de los cálculos debidamente ordenados y explicados (portada, índice, introducción, etc.)
 - Una **presentación tipo PowerPoint**, o similar, resumiendo el trabajo y las conclusiones más relevantes, con la explicación oral grabada (mediante notas de voz en las diapositivas, por ejemplo)
-