

Limpieza y Análisis de Datos

Lucas Gómez Torres y Joan Amengual Mesquida

13 de marzo, 2023

Índice General

1	1. Descripción del dataset. ¿Por qué es importante y qué pregunta/problema pretende responder?	2
2	Integración y selección de los datos de interés a analizar.	2
3	Limpieza de los datos.	2
3.1	¿Los datos contienen ceros o elementos vacíos? Gestiona cada uno de estos casos.	2
3.2	Identifica y gestiona los valores extremos.	2
4	Análisis de los datos.	2
4.1	Selección de los grupos de datos que se quieren analizar/comparar (p.ej., si se van a comparar grupos de datos, ¿cuáles son estos grupos y qué tipo de análisis se van a aplicar?).	2
4.2	Comprobación de la normalidad y homogeneidad de la varianza.	2
4.3	Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc. Aplicar al menos tres métodos de análisis diferentes.	2
5	Representación de los resultados a partir de tablas y gráficas.	2
6	Resolución del problema. A partir de los resultados obtenidos, ¿cuáles son las conclusiones? ¿Los resultados permiten responder al problema?	2
7	Código.	2
8	Vídeo.	2

1 1. Descripción del dataset. ¿Por qué es importante y qué pregunta/problema pretende responder?

2 Integración y selección de los datos de interés a analizar.

Puede ser el resultado de adicionar diferentes datasets o una subselección útil de los datos originales, en base al objetivo que se quiera conseguir.

3 Limpieza de los datos.

3.1 ¿Los datos contienen ceros o elementos vacíos? Gestiona cada uno de estos casos.

3.2 Identifica y gestiona los valores extremos.

4 Análisis de los datos.

4.1 Selección de los grupos de datos que se quieren analizar/comparar (p.ej., si se van a comparar grupos de datos, ¿cuáles son estos grupos y qué tipo de análisis se van a aplicar?).

4.2 Comprobación de la normalidad y homogeneidad de la varianza.

4.3 Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc. Aplicar al menos tres métodos de análisis diferentes.

5 Representación de los resultados a partir de tablas y gráficas.

Este apartado se puede responder a lo largo de la práctica, sin necesidad de concentrar todas las representaciones en este punto de la práctica.

6 Resolución del problema. A partir de los resultados obtenidos, ¿cuáles son las conclusiones? ¿Los resultados permiten responder al problema?

7 Código.

Hay que adjuntar el código, preferiblemente en R, con el que se ha realizado la limpieza, análisis y representación de los datos. Si lo preferís, también podéis trabajar en Python.

8 Vídeo.

Realizar un breve vídeo explicativo de la práctica (máximo 10 minutos), donde ambos integrantes del equipo expliquen con sus propias palabras el desarrollo de la práctica, basándose en las preguntas del enunciado

para justificar y explicar el código desarrollado. Este vídeo se deberá entregar a través de un enlace al Google Drive de la UOC, junto con enlace al repositorio Git entregado.