

Tarea evaluable de análisis de datos

1. En ciertas empresas se realizan pruebas de selección para conocer la aptitud para el trabajo de los aspirantes a obtener un empleo. Resulta por tanto clave saber qué pruebas pueden predecir la aptitud para el trabajo de una persona. En el archivo `trabajo.dat` se encuentran las puntuaciones de un test psicotécnico (`psico`) y test de adecuación (`adec`) que obtuvieron 20 individuos que fueron contratados por la compañía. También aparece el resultado de la evaluación a que fueron sometidos estos 20 empleados tras un periodo de dos años (`eval`).
 - a) Plantea un modelo teórico para estudiar la posible relación de la evaluación con el resultado de los dos test. Justifica la elección de dicho modelo.
 - b) Indica las conclusiones que se obtienen de la aplicación del modelo. ¿Es conveniente cambiar el modelo para mejorar el proceso de selección? Justifica las respuestas.
 - c) Realiza un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.
2. Un ingeniero sospecha que el tiempo (en minutos) para la elaboración de cierta pieza metal está influenciado por la procedencia del metal empleado y por la profundidad del corte de la máquina. Para comprobar su hipótesis, toma metal de cada uno de los tres proveedores de la empresa y elige distintas profundidades de corte (en pulgadas) para cada proveedor. Los datos obtenidos una vez llevado a cabo el experimento se encuentran en el archivo `pieza2.dat`.
 - a) Plantea un modelo teórico para estudiar si la procedencia del metal y la profundidad del corte influyen en el tiempo de elaboración de dicha pieza de metal. Justifica la elección de dicho modelo.
 - b) Indica las conclusiones que se obtienen de la aplicación del modelo. ¿Es necesario cambiar de modelo para analizar los datos? Justifica las respuestas.
 - c) Realiza un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.
3. La hembra de salmón entierra los huevos tras el proceso de apareamiento y fertilización en unos nidos en el lecho del río. Para un estudio comparativo en los ríos Navia, Negro y Esva se desenterraron 6 de estos nidos en cada río y se registró el porcentaje de huevos que había sido fecundado y la temperatura del agua en torno al nido. Los datos se encuentran en el archivo `salmones.dat`.
 - a) Plantea un modelo teórico para estudiar la posible relación entre el porcentaje de huevos fecundados y la temperatura, determinando al mismo tiempo en qué medida puede venir dicha relación condicionada por el río en que se encuentran los nidos. Justifica la elección del modelo.
 - b) Indica las conclusiones que se obtienen de la aplicación del modelo. ¿Ha sido necesario cambiar el modelo inicial en este estudio? Justifica las respuestas.
 - c) Realiza un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.
4. Se estudia la posible influencia que determinados índices económicos pueden tener en el precio de la acción de grandes y medianas empresas. En una muestra de 50 empresas, se investiga el precio de la acción en bolsa y además se registra información sobre: facturación (en millones de euros), número de trabajadores, deuda, valor de las propiedades e inversiones realizadas (esta tres últimas en miles de euros). Los datos se encuentran en el archivo `accion.dat`. Analiza estos datos de la forma más detallada posible. Se pide:
 - a) Plantea un modelo teórico para estudiar la posible relación entre el precio de la acción y las otras variables registradas. Justifica la elección de dicho modelo.
 - b) Indica las conclusiones que se obtienen de la aplicación del modelo. ¿Es conveniente cambiar el modelo para un mejor análisis de las oscilaciones del precio de la acción? Justifica las respuestas.
 - c) Realiza un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.

5. Se piensa que el voltaje máximo que puede proporcionar una batería puede estar influenciado por el material con el que está fabricada y por la temperatura a que trabaja. Se diseñó una experiencia en laboratorio con tres tipos de material y a distintas temperaturas. Los datos obtenidos se encuentran en `voltaje2.dat`.
- Plantea un modelo teórico para estudiar si el material y la temperatura influyen en el voltaje de la batería. Justifica la elección del modelo.
 - Indica las conclusiones que se obtienen de la aplicación del modelo. ¿Es conveniente cambiar el modelo para analizar los datos? Justifica las respuestas.
 - Realiza un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.
6. En el archivo `ericksen.dat`, se encuentran los datos correspondientes a un estudio llevado a cabo en 1980 para determinar la influencia que determinadas variables demográficas tenían en la tasa de criminalidad (`crime`). En los 50 estados de los Estados Unidos se registraron el porcentaje de hispanos o negros (`minority`), el porcentaje de población en situación de pobreza (`poverty`), el porcentaje de población con dificultades para hablar o entender el idioma (`language`), el porcentaje de mayores de 25 años sin finalizar la educación secundaria (`highschool`) y el porcentaje de la población con problemas de vivienda (`housing`).
- Plantea un modelo teórico para estudiar la posible relación de la tasa de criminalidad con el resto de variables. Justifica la elección del modelo.
 - Indica las conclusiones que se obtienen de la aplicación del modelo. ¿Es conveniente cambiar de modelo para dar una mejor interpretación de la tasa de criminalidad en términos de las variables demográficas? Justifica las respuestas.
 - Realiza un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.
7. En los datos `highway.dat` aparecen los registros de la tasa de accidentes por millón de vehículos en 39 secciones de diversas carreteras en el estado de Minnesota en 1973. El objetivo del análisis era entender la influencia que ciertas variables relativas al tráfico y al diseño de las carreteras podían tener sobre dicha tasa de accidentes. Las variables que aparecen en el archivo son:
- `rate`: tasa de accidentes por millón de vehículos
 - `len`: longitud del segmento de carretera en millas
 - `ADT`: tráfico diario promedio (en miles)
 - `trks`: porcentaje de camiones sobre el tráfico total
 - `sigs1`: número de señales por milla de carretera
 - `slim`: límite de velocidad del tramo de carretera
 - `shld`: anchura del arcén en pies
 - `acpt`: número de puntos de acceso por milla de carretera
 - `itg`: número de cruces por milla de carretera
 - `lwid`: anchura del carril (en pies)
- Plantea un modelo teórico para estudiar la posible relación de la tasa de accidentes con el resto de variables. Justifica la elección del modelo.
 - Indica las conclusiones que se obtienen de la aplicación del modelo. ¿Es posible eliminar del modelo alguna de las variables relativas al tráfico y al diseño de las carreteras? Justifica la respuesta.
 - Realiza un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.

8. En el archivo `leinhardt.dat` aparecen los datos de mortalidad infantil en 1970 de 105 países, junto con la renta per capita y el continente donde se encuadran. Se pretende determinar si la renta per capita influye en la mortalidad infantil y si esta influencia viene condicionada por el continente donde se ubica el país. Las variables que aparecen en el archivo son:

`income`: renta per cápita (en dólares)

`infant`: tasa de mortalidad infantil por cada 1000 nacimientos

`region`: continente donde se ubica el país

- a) Plantea un modelo teórico para estudiar la posible relación de la mortalidad infantil de un país con su renta per capita y si dicha relación es igual para los 4 continentes considerados o si por el contrario existen diferencias entre continentes.. Justifica la elección del modelo.
 - b) Indica las conclusiones que se obtienen de la aplicación del modelo. ¿Es conveniente un cambio del modelo inicialmente elegido? Justifica las respuestas.
 - c) Realiza un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.
9. Se toman datos de 209 ordenadores con el objeto de predecir el rendimiento de la cpu (y) en función de una serie de variables. El rendimiento se mide tomando como punto de referencia la cpu de un determinado modelo de la firma IBM. Las variables que pueden ser influyentes son el tiempo de cada ciclo en nanosegundos (x_1), la memoria principal máxima (x_2), la memoria principal mínima (x_3), ambas en kilobytes, el tamaño de la memoria caché (x_4), también en kilobytes, el número mínimo (x_5) y máximo (x_6) de canales. Los datos se encuentran en el archivo `cpus.dat`.

Realizar el siguiente estudio estadístico de estos datos:

- a) Plantear un modelo teórico para estudiar la posible relación del rendimiento de la cpu con el resto de variables.
 - b) ¿Es posible eliminar alguna de las variables influyentes del modelo? Razona la respuesta.
 - c) Realizar un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.
10. Un distribuidor de bebidas está estudiando la efectividad de los métodos de reparto. La variable de interés es el tiempo de reparto en minutos (y); se sospecha que existe cierta relación con el volumen de la caja que se reparte. Por otra parte, los empleados utilizan tres tipos de carretilla. Se diseña el experimento de forma que se utiliza 4 veces cada uno de los tipos de carretilla, obteniéndose los datos que aparecen en el archivo `reparto.dat`.

Realizar el siguiente estudio estadístico de estos datos:

- a) Plantear un modelo teórico para estudiar la posible relación del tiempo de reparto con el volumen de la caja y el tipo de carretilla.
- b) ¿Es posible eliminar alguna de las variables influyentes del modelo? Razona la respuesta.
- c) Realizar un diagnóstico de las hipótesis teóricas del modelo finalmente elegido y en función de dicho diagnóstico determinar si las conclusiones de los apartados anteriores son válidas.

Criterios de valoración

A la hora de calificar estos trabajos se valorará:

- Que se haya realizado el planteamiento teórico adecuado de forma justificada.
- Que se detallen los pasos seguidos en el análisis y el motivo por el que se realizan.
- Que se relacionen los resultados con los contenidos teóricos.
- Que se contextualicen los resultados en el marco del problema concreto planteado.
- La presentación del trabajo.