| DATOS PERSONALES | FIRMA |
| --- | --- |
| |  |  | | --- | --- | | Nombre: Néstor | DNI: 0603412883 | |  |
| Apellidos: Estrada Brito |

| ESTUDIO | ASIGNATURA | CONVOCATORIA |
| --- | --- | --- |
| MÁSTER UNIVERSITARIO EN INGENIERÍA MATEMÁTICA Y COMPUTACIÓN (PLAN 2016) | 4391020006.- TÉCNICAS MULTIVARIANTES | Ordinaria Número periodo 1823 |

| FECHA | MODELO | CIUDAD DEL EXAMEN |
| --- | --- | --- |
| 14-16/01/2022 | Modelo - D |  |

| Etiqueta identificativa |
| --- |
|  |

**INSTRUCCIONES GENERALES**

1. Ten disponible tu documentación oficial para identificarte, en el caso de que se te solicite.
2. Si tu examen consta de una parte tipo test, indica las respuestas en la plantilla según las características de este.
3. Debes contestar en el documento adjunto, respetando en todo momento el espaciado indicado para cada pregunta. Si este es en formato digital, los márgenes, el interlineado, fuente y tamaño de letra vienen dados por defecto y no deben modificarse. En cualquier caso, asegúrate de que la presentación es suficientemente clara y legible. Entrega toda la documentación relativa al examen, revisando con detenimiento que los archivos o documentos son los correctos. El envío de archivos erróneos o un envío incompleto supondrá una calificación de “no presentado”.
4. Durante el examen y en la corrección por parte del docente, se aplicará el Reglamento de Evaluación Académica de UNIR que regula las consecuencias derivadas de las posibles irregularidades y prácticas académicas incorrectas con relación al plagio y uso inadecuado de materiales y recursos.
5. No está permitido el uso de Internet ni ningún tipo de comunicación con otra persona.Durante todo el examen tu teléfono móvil debe estar en modo avión.
6. La parte principal de cada pregunta consiste en interpretar y comentar los resultados obtenidos. Si te limitas a hacer los cálculos no vas a poder superar el examen.
7. Es fundamental que las respuestas estén debidamente redactadas, de forma clara y precisa y sin faltas de ortografía.
8. Para hacer el examen puedes utilizar los apuntes del curso y los scripts que hayas preparado y Python para hacer los cálculos.

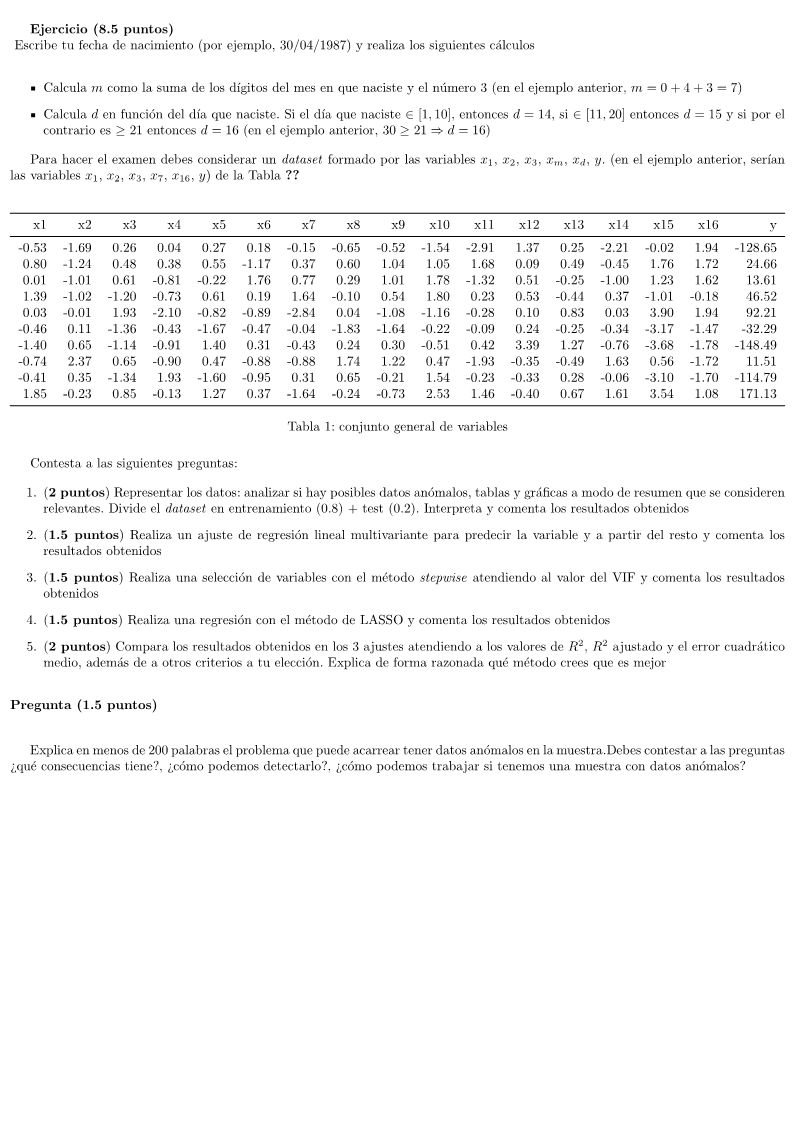
**Puntuación**

**Preguntas**

* Puntuación máxima 10.00 puntos

El examen constará de un ejercicio práctico (8,5 puntos) y una pregunta teórica (1,5 puntos). Los enunciados están en la página 14 y el espacio para responder el examen está entre las práginas 4 y 13.  
  
**1.** Pregunta

 (Responder en 10 caras)



PREGUNTA 1

* 1. **DATABASE**

Para la realización del dataframe me salen los siguientes resultados

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza bajaTexto

Descripción generada automáticamente

En la siguiente gráfica podemor ver la describción y la información de los datos, donde nos podemos dar cuenta que cada variable tiene 10 datos no nulos, en las variables predictoras la media tiende a cero menos en la variable x1, la cual puede ser un valor atípico, a parte la media y la desviación tipida de las variables predictoras difieren mucho de la media y la desviación típica de la variable respuesta

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza mediaInterfaz de usuario gráfica, Texto

Descripción generada automáticamente

Entonces como m=5 y d=15, debo coger los valores x1, x2, x3, x5, x15 para realizar el nuevo database

Y me quedaría como en la gráfica de abajo. En la gráfica podemor ver la descripción y la información de los datos, donde nos podemos dar cuenta que cada variable tiene 10 datos no nulos, en las variables predictoras la media tiende a cero menos en la variable x1, x15, la cual puede ser un valor atípico, a parte la media y la desviación tipida de las variables predictoras difieren mucho de la media y la desviación típica de la variable respuesta. Una vez visto los datos descritos de las variables predictoras y de la variable respuesta, se puede observar que las variables predictoras tienen una media que se encuentra próximo al valor 0 y una desviación estándar próximo a 1. Mientras que para la variable respuesta se tiene una media de -16.8 y una desviación estándar de 119.11.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Se procede a realizar la representación de las variables predictoras, en las gráficas ponemos observar que la variable x2, x5, x15 tienen valores constantes centrados en el valor de 1, en las demas variables presentan valores atípicos. Esto se debe a q hay pocos datos en cada variable, esto cambiaría si se representa mas de 200 valores para cada variable.

Imagen que contiene Gráfico

Descripción generada automáticamente

Representación de la variable respuesta, se puede observar que algunos datos se centran en el valor de 2.0 y otros en el valor de 1.0, si hubiese más valores se tendría una distribución diferente.

Gráfico, Histograma

Descripción generada automáticamente

Se procede a dividir el dato de entrenamiento con (0.8 y test de 0.2) y queda como resultado lo siguiente:

Para el entrenamiento se muestra en la siguiente gráfica, en la que se puede observar que son asignados 8 valores para el entrenamiento.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

En la siguiente gráfica se muestra que son asignados 2 valores para la validación.

Imagen que contiene tabla, computadora, grande, computer

Descripción generada automáticamente

Las lineas de programación se adjuntan en otro archivo en la plataforma de unir.

* 1. **REGRESIÓN LINEAL MULTIVARIANTE**

**Realizando la regresión linela muiltivariante obtenemos el intercepto con un valor de 103.6754128 y los coeficientes de regresión de: -108.9199 19.7777 13.44832 -19.101700 -76.2777**

En la Figura de abajo se puede observar que, el p\_valor de las variables predictora (x3, x5) se acercan a 1, con lo cual se podría a eliminar dichas variables predictoras. Se puede observar en la Figura que el coeficiente de determinación del modelo es igual a 0.97, y que el coeficiente de determinación ajustado es 0.896. La diferencia no es tan grande por las observaciones dadas de la Figura. Los coeficientes salen altos por la variable respuesta que tenía valores igual altos. Se podría eliminar las variables con el p\_ valor elevado mediante el stepwise, pero no pide el ejercicio.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

* 1. **VIF- STEPWISE**

En la siguiente figura podemos observar que la variable x2 posee un valor VIF más elevado que las otras variables, por lo que podría presentar problemas de colinealidad. Mediante stepwise se procede a eliminar esa variable.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Realixando el stepwise para la variable x2 muestra los siguientes resultados vistos en la gráfica siguiente, en la cual nos podemos dar cuenta que el r cuadrado ajustado mejoró su valor con respecto al anterior, pero el pvalor de la variable x3 es cercano a uno. Con lo cual procedemos a eliminar la variable x2 y realizar la regresión.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente con confianza media

Una vez eliminado la variable x2 se observa los siguientes resultados de la siguiente gráfica, con valores del intercepto de -6.935 y coeficientes de 0.8166 61.839 -24.573 -23.6340

Texto

Descripción generada automáticamente

* 1. **LASSO**

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

**En la siguiente gráfica muestra los resultado de la regresión de LASSO, CON UN INTERCEPTO DE 87.3175, y unos coeficientes de -78.80 29.1 25.838 -0 -67.607**



* 1. **COMPARACIÓN**

**Se encuentra el error cuadrático medio y se compara**

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

estos valores salen muy grandes porque las variables estan normalizadas, pero la varibe y, x lo tanto

es un error coherente con la variable y

el error màs pequeño lo tiene la red STEPWISE, POR LO QUE SERÍA EL MEJOR MÉTODO PARA ÉSTE PROBLEMA

EJERCICIO 2 TEORIA

Un árbol de decisión es en sí una secuencia de condiciones para distintas variables predictoras y su relación con la variable de respuesta. Estas condiciones pasan en varios caminos del árbol, generando divisiones con distintas posibilidades en ramas. Estas ramas, así pues, representan divisiones del espacio de los predictores en diferentes regiones. Para realizar las predicciones se emplea la media o la moda de la región a la cual pertenece dicha observación. Las decisiones que clasifican las observaciones están organizadas en nodos. Estos nodos se pueden identificar por la profundidad, teniendo varios niveles de profundidad. Un árbol de decisión empieza con un único nodo y luego se ramifica en resultados posibles, cada resultado crea nodos adicionales, que se ramifican en otras posibilidades. Esta representación es similar a la de un árbol. Sirven para categorizar y representar varias condiciones que ocurren de forma sucesiva, para la resolución analítica de cierto problema categórico.Los árboles de decisión son simples y son fáciles de interpretar, aunque suelen tener unas capacidades de ajuste y de predicción peores que los metodos de regresion lineal. Las ventajas que presentan son: sencillos de implementar, fáciles de entender e interpretar, versátiles y permiten resolver tanto problemas de clasificación como de regresión, las desventajas serían que son muy sensibles ante rotaciones de los datos de la muestra de entrenamiento, y cuando el árbol es muycomplejo tambien se hace complejo la interpretación.