| DATOS PERSONALES | FIRMA |
| --- | --- |
| |  |  | | --- | --- | | Nombre: Darío Esteban | DNI:1717630964 | |  |
| Apellidos: Arroyo Huerta |

| 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 | Quito, Ecuador |

| 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)

Fecha de nacimiento: 31/01/1991

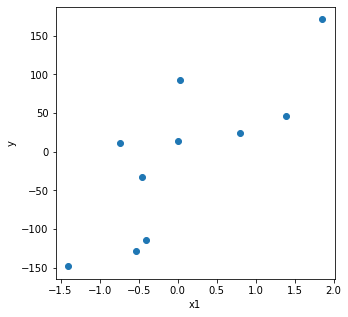
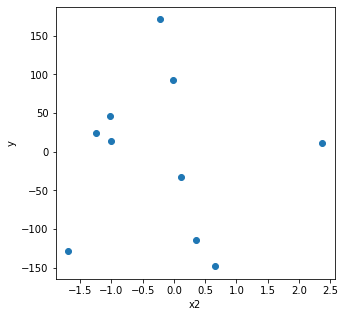
m=0+1+3=4

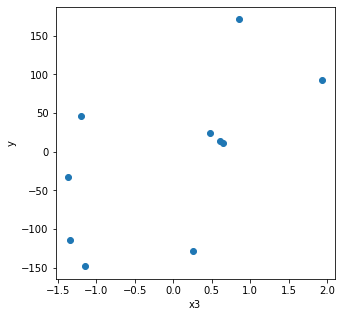
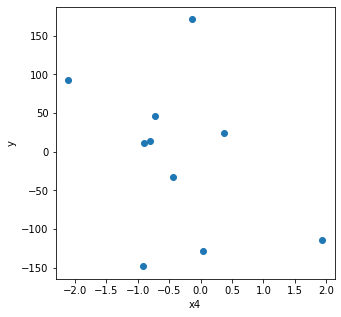
d=16

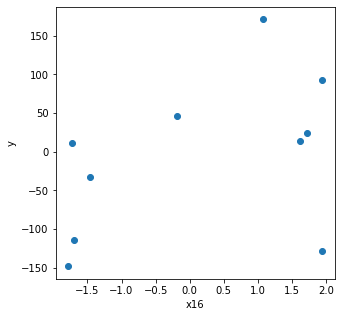
mis datos:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **x1** | **x2** | **x3** | **x4** | **x16** | **y** |
| -0,53 | -1,69 | 0,26 | 0,04 | 1,94 | -128,65 |
| 0,8 | -1,24 | 0,48 | 0,38 | 1,72 | 24,66 |
| 0,01 | -1,01 | 0,61 | -0,81 | 1,62 | 13,61 |
| 1,39 | -1,02 | -1,2 | -0,73 | -0,18 | 46,52 |
| 0,03 | -0,01 | 1,93 | -2,1 | 1,94 | 92,21 |
| -0,46 | 0,11 | -1,36 | -0,43 | -1,47 | -32,29 |
| -1,4 | 0,65 | -1,14 | -0,91 | -1,78 | -148,49 |
| -0,74 | 2,37 | 0,65 | -0,9 | -1,72 | 11,51 |
| -0,41 | 0,35 | -1,34 | 1,93 | -1,7 | -114,79 |
| 1,85 | -0,23 | 0,85 | -0,13 | 1,08 | 171,13 |

A continuación se presentan las imágenes obtenidas al graficar cada una de las variables “xi” contra la variable “y”







Al realizar el análisis de regresión lineal se obtienen los coeficientes de regresión:

[86.26808608, -57.88394201, 115.33955295, -23.52902514, -129.13242558]

Sin embargo cuando se continua el análisis se obtiene esta tabla:

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

const -6.4580 10.121 -0.638 0.558 -34.558 21.642

x1 86.2681 12.030 7.171 0.002 52.866 119.670

x2 -57.8839 3068.950 -0.019 0.986 -8578.654 8462.886

x3 115.3396 3007.868 0.038 0.971 -8235.841 8466.520

x4 -23.5290 16.554 -1.421 0.228 -69.491 22.433

x5 -129.1324 4389.428 -0.029 0.978 -1.23e+04 1.21e+04

==============================================================================

Se observa que los p-valores son mayores a 0.005 excepto para x1, por lo que no parecería existir una correlación estadísitica significativa entre los datos. Sin embargo, esto nos puede indicar la presencia de datos anómalos dentro de los datos, o simplemente que estos datos no se ajustan muy bien con regresión lineal.

Por otro lado, al realizar la regresión con LASSO (alpha=1) obtenemos estos coeficientes:

[ 88.36826922 27.71170853 24.42067704 -22.45400905 0. ]

Esto nos dice por otro lado que se ha dado selección de variables al utilizar un coeficiente de penalización 1, y que los datos, según este análisis no dependería de x16.

**2.** Pregunta

Si se tienen datos anómalos, esto conlleva a que se tenga problemas en el momento de analizar los resultados de todos los demás datos, pues usualmente se utilizan variables como la media y la varianza para analizar el comportamiento de una distribución, los cuales pueden ser muy influenciados por la presencia de datos anómalos. Si no se conoce la distribución que siguen nuestros datos, entonces usualmente se utiliza “muestreo” para tener una idea de cómo puede ser la distribución de los datos bajo estudio.

Sin embargo, los datos anómalos pueden influenciar mucho la aproximación que se tiene respecto a la distribución aproximada con la real. Para reducir este efecto, se tienen varias estrategias:

* 1. Se puede eliminar el dato anómalo de la muestra bajo estudio, lo cual requiere ignorar por completo su influencia.
  2. Se puede utilizar otros estimadores que no son tan influenciados por los datos anómalos para analizar la muestra de los datos, por ejemplo la mediana, la media recortada, la media winsorizada, la trimedia, el MAD, etc. Estos también son estimadores robustos.
  3. Debido a que los datos anómalos tienen mayor influencia en conjuntos de datos pequeños, podríamos reducir su influencia incrementando el número de muestras.

Para detectar si una muestra tiene un dato “anómalo”, se lo puede realizar comparando la media y la varianza de la muestra total, y alguna de las cantidades mencionadas anteriormente, y si hay demasiada diferencia, significa que tenemos datos anómalos.

