| DATOS PERSONALES | FIRMA |
| --- | --- |
| |  |  | | --- | --- | | Nombre: María Alejandra | DNI: 23681944-2  PP: AAB412563 | |  |
| Apellidos: Alvarez |

| 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.

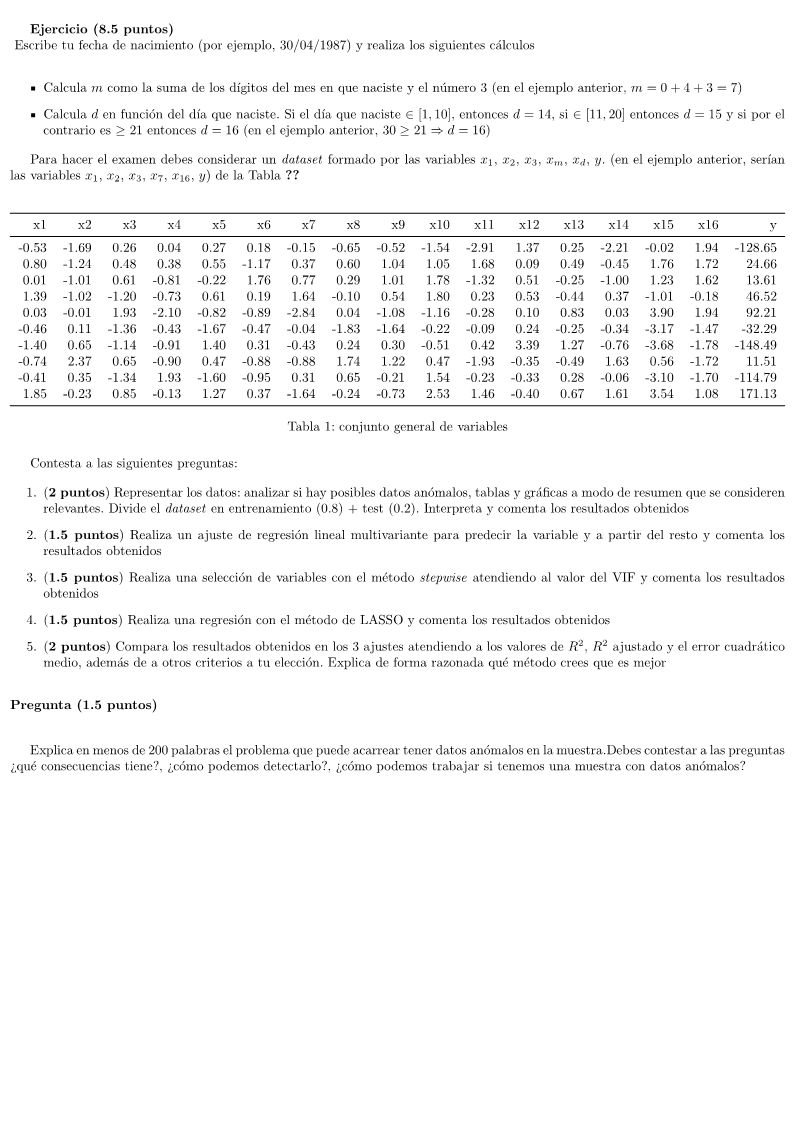
**Puntuación**

**Preguntas**

* Puntuación máxima 10.00 puntos

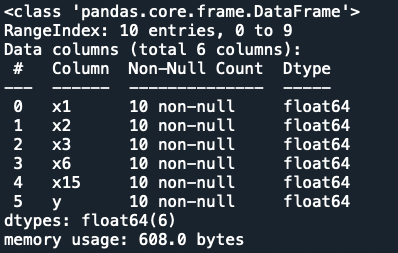
El examen constará de un ejercicio práctico (8,5 puntos) y una pregunta teórica (1,5 puntos). Los enunciados están en la página 14 y el espacio para responder el examen está entre las práginas 4 y 13.  
  
**1.** Pregunta

 (Responder en 10 caras)



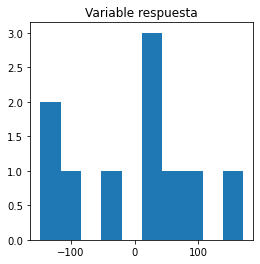
1. En mi caso m=6 y d=15 con lo cual considero x1, x2, x3, x6, x15, y como dataset.

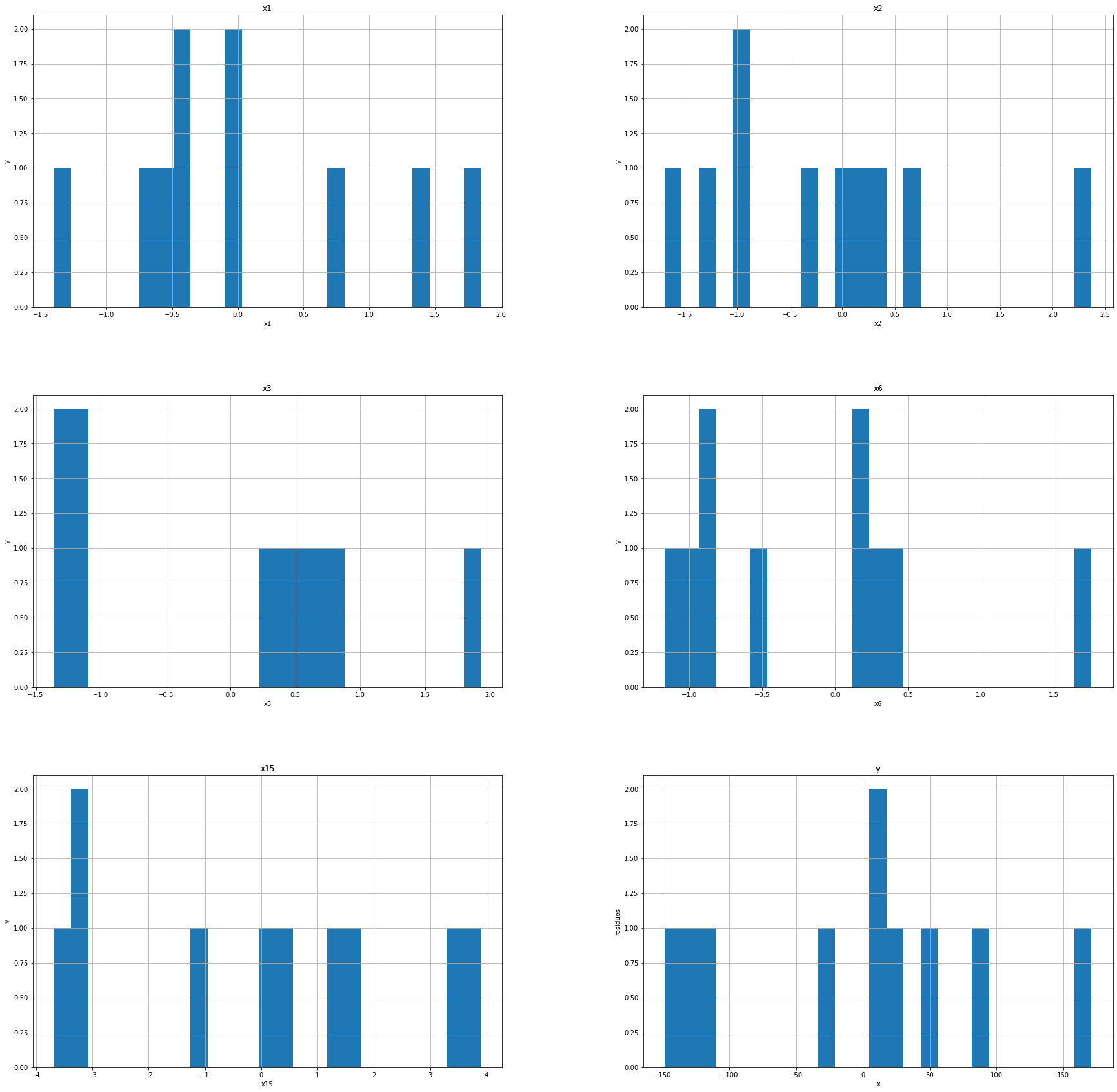
El resultado que se obtiene con info() y describe() es el siguiente



Como era de esperar tenemos 10 datos numéricos (float64) en cada variable y no hay faltantes.

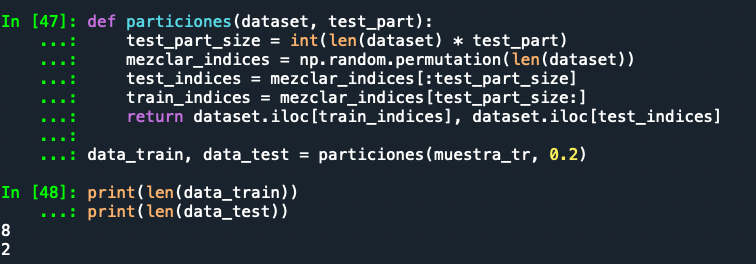
Podemos estandarizar las variables (pero en realidad no hay mucha diferencia) y dibujar los histogramas de cada variable independiente y de la respuesta:





Podemos ver que en algunos casos podría parecer una distribución normal, pero no en todos, quizá porque son pocos datos.

Dividimos el dataset en 80-20, obteniendo, como era de esperar 8 datos para entrenamiento y 2 para test:



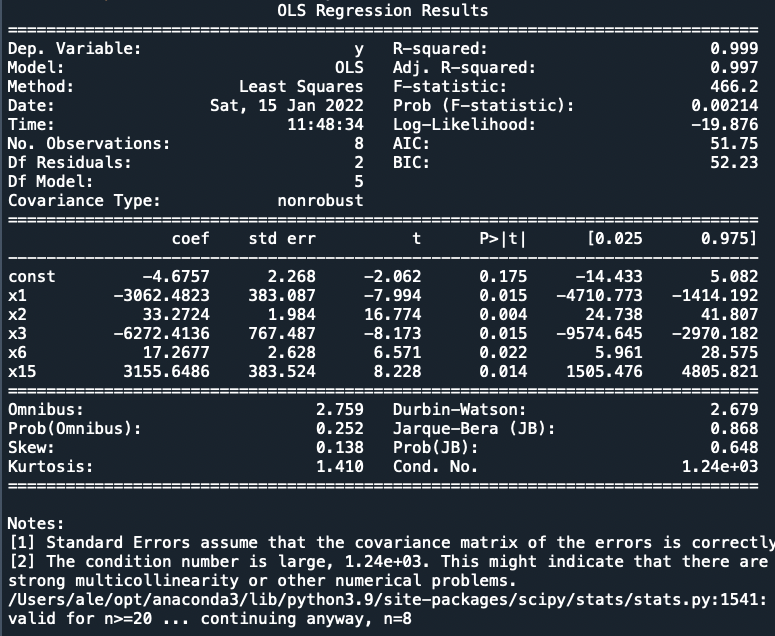
1. Ajustamos por regresión multivariante y obtenemos

Intercepto: ﻿ ﻿-4.67570023181861

Coeficientes: ﻿[-3062.48230712 33.27238159 -6272.41357299 17.2676895

3155.64862492]

Y el resultado



Aquí podemos ver que el coeficiente de determinación R^2 es 0.999 y el ajustado 0.997 muy similares.

También podemos ver que todas las variables son significativas debido a su p.valor < 0.05.

Finalmente, el error cuadrático medio es de ﻿ ﻿977.6409042896014.

1. Con la regresión Lasso dando varios valores de alpha, obtenemos:

para alpha=1 que el intercepto es ﻿ ﻿-8.897217136929523 y los coeficientes son ﻿[66.58407257 33.0020399 -0. 9.87787831 21.1819691 ] . Esto implica que la variable x3 no aporta en este modelo.

para alpha=1e4 que el intercepto es ﻿ ﻿10.3125 y los coeficientes son ﻿[ 0. 0. 0. -0. 0.]. Esto implica que y no depende de las variables.

Buscando el alpha óptimo, obtenemos un error cuadrático medio de ﻿55.194029473912714.

5) De la comparación de los ecm podemos concluir que conviene la regresión Lasso, pero va a depender de la elección de alpha.

Pregunta 2:

Tener datos anómalos puede traer como consecuencia la obtención de resultados erróneos y conclusiones equivocadas si no se analiza el porqué de esos datos. Los datos anómalos pueden tener su origen en errores en la obtención de datos (error grosero) o pueden ser correctos y parte de la muestra considerada. Una forma de detectarlos, cuando se puede, es gráficamente. Si podemos representar los datos (dimensión 2 o 3), podemos apreciar fácilmente la presencia de datos anómalos. Por ejemplo en un scatter se pueden ver alejados de la nube de puntos, o en un boxplot como puntos fuera de las cajas y bigotes.

En el caso de que sean incorrectos, se deben eliminar. Si los datos son correctos, existen varias formas de trabajar con ellos: Una primera es simplemente eliminarlos, pero esto puede ser un error que implique modelos o decisiones incorrectas. Otra segunda opción es trabajar con ellos y aplicar aplicar alguna estrategia apropiada al modelo de regresión (transformación de la variable respuesta, cambio de método, inclusión de nuevas variables, etc.). Por último, también se puede aplicar algún método robusto (por ejemplo regresión por mínimos cuadrados recortados).