| DATOS PERSONALES | FIRMA |
| --- | --- |
| |  |  | | --- | --- | | Nombre: Juan Carlos | DNI: 1203546161 | |  |
| Apellidos: Yturralde Villagómez |

| 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 | Guayaquil |

| 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.

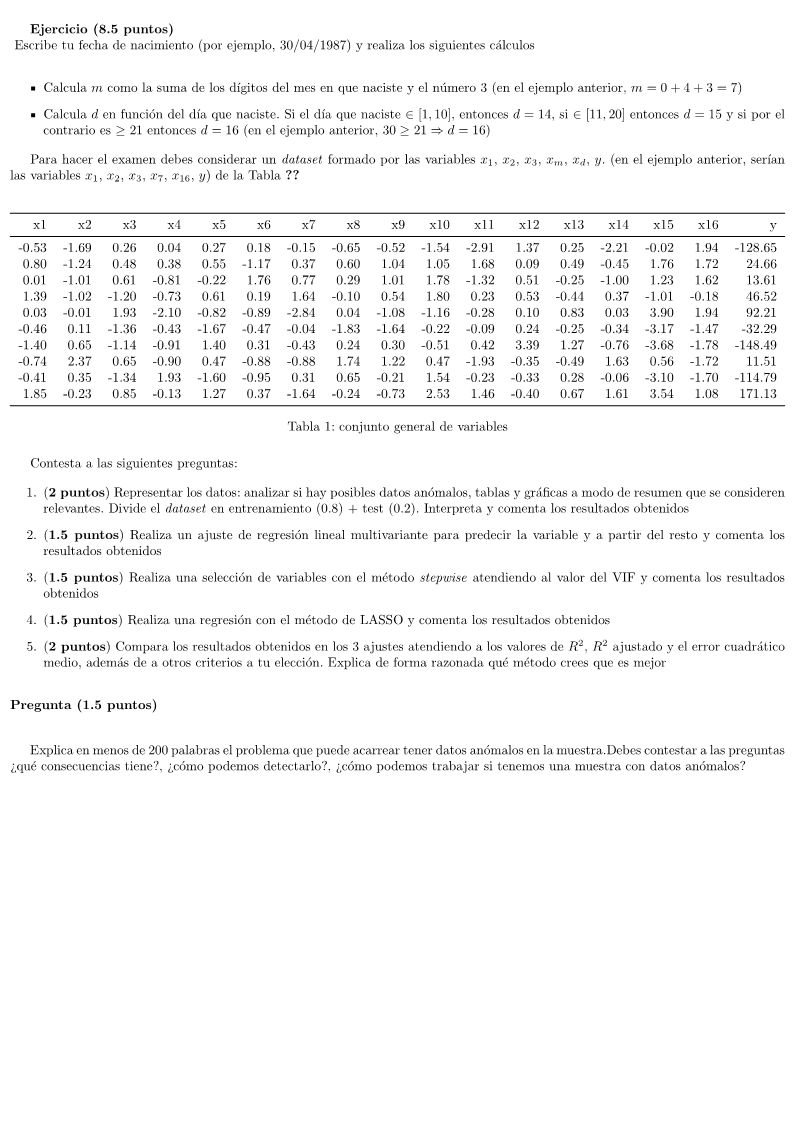
**Puntuación**

**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 = 13-07-1981**

**Calculo de m**

* m= 0+7+3
* **m=10**

**Cálculo de d**

* **d=15**

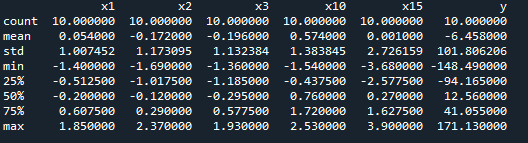
**Variable del Dataset**

* **x1,x2,x3,x10,x15**

**Pregunta 1**

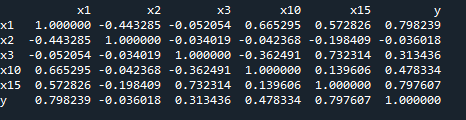
Se realizo el método describe de la variable del dataset donde se pudo visualizar que el conjunto de observaciones son 10 datos para cada variable. La media de la variable x1 y x15 son aproximadamente cero, mientras que la variable x2 y x3 su media es aproximadamente -0.2, y la variebla x10 tiene una media de 0,57. La variables y tiene un media de -6.45. La desviación standar de la variable x1, x2,x3,x10 estan en el rango de 1 a 1.38. La variable x15 tiene la desviación standar de 2.72 y la Varible Respuesta y tiene un desviación de 101.80. Los datos de la variables predictores tienes valores pequeño, mientras que los datos de la variables respuesta tienen valores más altos.

**print(df.describe())**

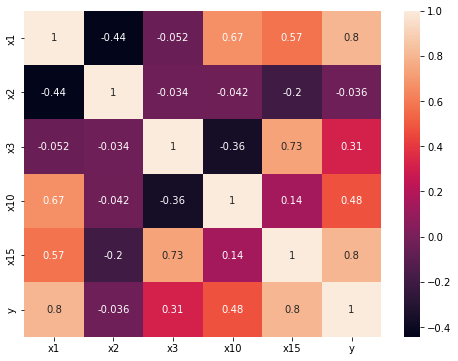


**Se puede visualizar que la correlación de algunas variable son menores a 0.5, es posible que exista colinealidad entre varias variables. Pero la variable que es posible que tenga mayor colinealidad es la variable x2 como se puede visualizar en la grafica**

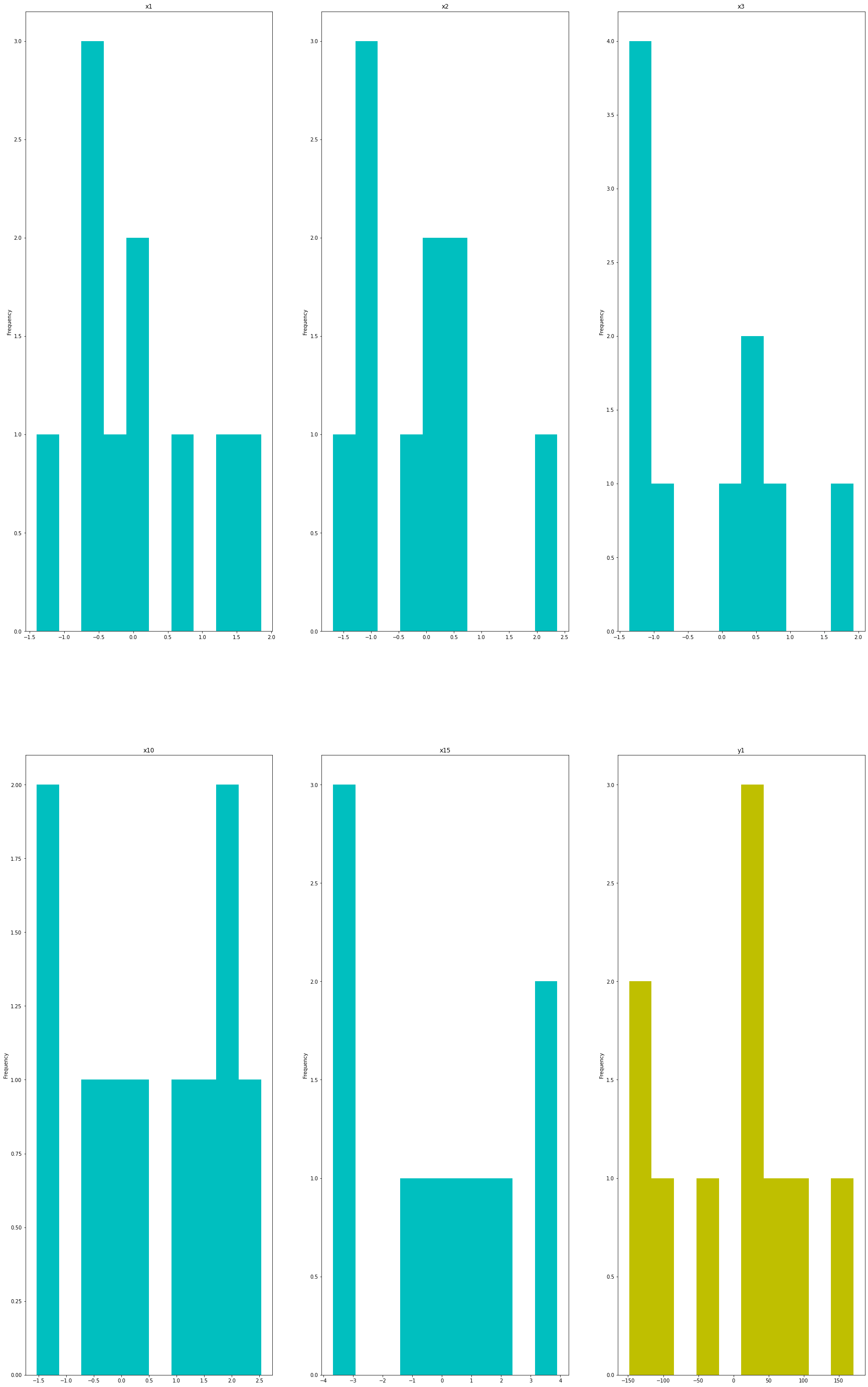
**Matriz de Correlacion del dataframe:**



**Grafico de la Correlación**



**Grafico de la Variable Predictora y la Variables Respuesta**



**De la grafica se puede visualizar claramente que no es una distribución normal con una campana de gauss, sino fueran por dos grupo que difieren se podria decir que es una distribucion uniforme.**

**Pregunta 2**

**Regresion Lineal Multivariante**

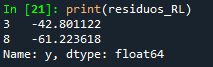
**Intercepto:**

**10.345485800323601**

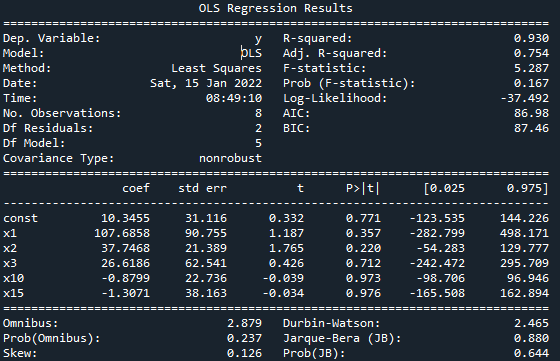
**Coeficientes**

**[107.68583302 37.74676887 26.61859756 -0.87987872 -1.30706074]**

**Residuo:**



**Regresion Lineal Multivariante OLS**



**Se puede visualizar claramente que el coeficiente de determinación r2 es alto y el R2 adjutado es 0.754. ademas se visualizar que todas la variables tiene un p.valor < 0.05 lo que nos indican que son variables menos significativas.**

**Pregunta 3**

**Metodo de StepWise**

**Me salio un error al correr el step wise, realice igual que la tarea pero lamentablemente no me corrio el programa que utilizando VIF queria elimar las variable que tenia mayor VIF, y ir analizando caso por caso pero lamentablemente el programa no me corrio.**

**Pregunta 4**

**Método Lasso**

**Alfha=1**

**Intercepto**

-6.704488674195167

**Coeficientes**

[53.08524115 17.68857738 -0. -0.44092166 23.29962616]

Con Alpha = 1 en relación con la regresión Lineal multiple, el intercepto disminuyo, y los coeficientes tambien

**Alfha =10000**

**Intercepto**

12.044999999999996

**Coeficientes**

[ 0. -0. 0. 0. 0.]

Con Alpha = 10000 todos coeficientes de las variables desaparecen,. Esto debido a que cuando dos variables están correlacionadas tiende a descartar una y ajustar la otra.

**Pregunta 5**

Error Cuadratico Medio Regresion Lineal Multivariante

4285.935862961027

Error Cuadratico Método Lasso

4552.072466246831

Podemos visualizar que el error cuadratico medio de la lasso es menor, podemos concluir que ajustar este conjunto de datos a una regresion lineal multivariangte no es lo más optimo ya que se detectaron que todas la variables predictoras eran menos significativas.

Pregunta Teorica

Los datoas anómalos en la muestra tienen la consecuensia de realizar una mala predicción en la regresion lineales, por lo que es importante separarlo y asi ajustar un mejor modelo.

Los datos anómalos se puedes con graficos tipos cajon, y ademas verificando la importante de la variables, como tambien encontrandos la variables menos significativos con los diferentes modelos estudiados en clase.

Si tenemos datos anomalos podria utilizar un selector de variable como stepwise para eliminar la varias menos significativas y así ajustar mejor el modelo