| DATOS PERSONALES | FIRMA |
| --- | --- |
| |  |  | | --- | --- | | Nombre: ALEX DARÍO | DNI: 1804393583 | |  |
| Apellidos: SÁNCHEZ PAREDES |

| 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.

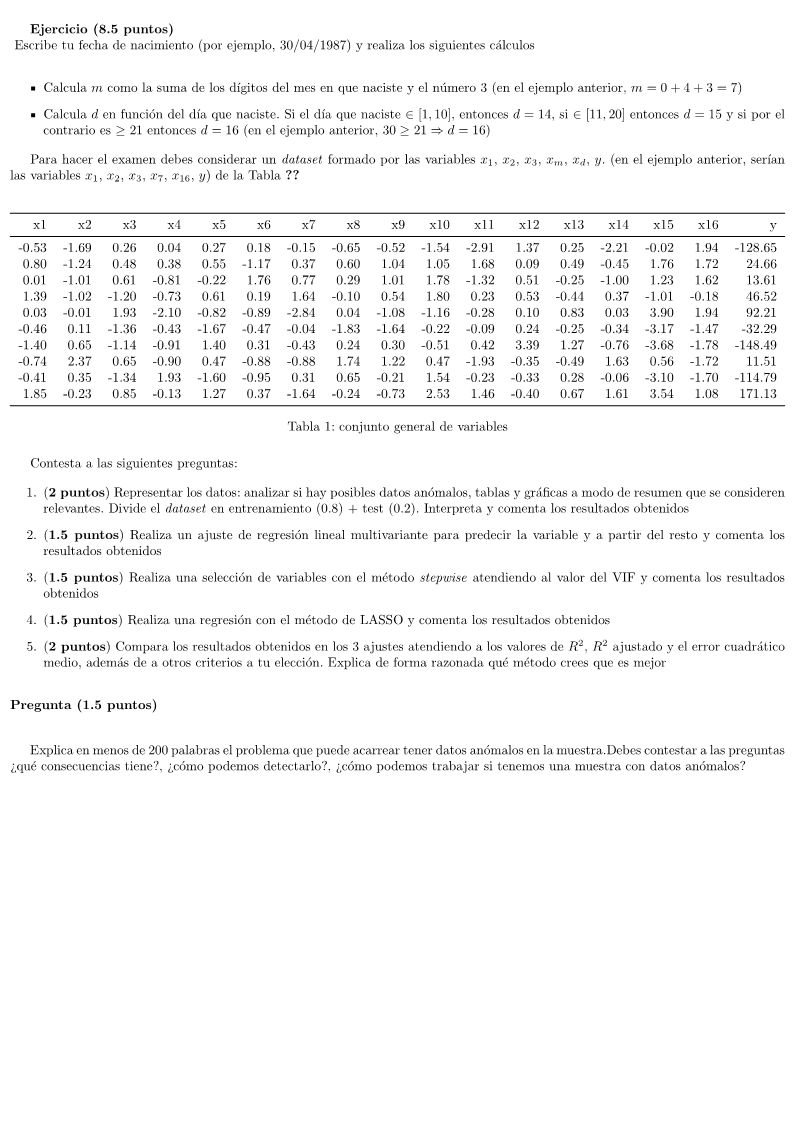
**Puntuación**

**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

#Cargas Librerias

import pandas as pd

import numpy as np

from pandas.core.common import flatten

from plotnine import \*

from array import \*

import scipy.stats as stats

import math

import matplotlib as mpl

import matplotlib.pyplot as plt

import statsmodels.api as sm

from sklearn import linear\_model

from sklearn.metrics import mean\_squared\_error, r2\_score

import statsmodels.formula.api

import statsmodels.stats.api as sms

se ajusta la fecha y el dia

#fecha de nacimiento 17/06/1993

m=0+6+3

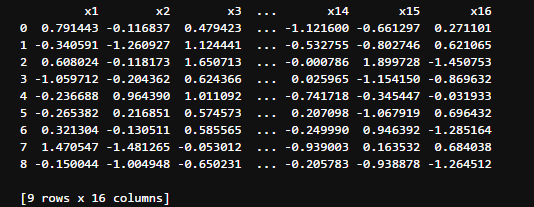
d=16

#se crea el dataframe acorde a los datos creados tomando en cuenta las filas y columnas respectivas

nom\_var = ["x%d" % i for i in range(1, 16 + 1)]

df = pd.DataFrame(np.random.randn(9, 16),columns=nom\_var )

print(df)



p\_train = 0.80 # Porcentaje de train. Acorde a lo solicitado

df['is\_train'] = np.random.uniform(0, 1, len(df)) <= p\_train

train, test = df[df['is\_train']==True], df[df['is\_train']==False]

df = df.drop('is\_train', 1)

se reazliza la impresion del entrnamiento y de las pruebas

print("Ejemplos usados para entrenar: ", len(train))

print("Ejemplos usados para test: ", len(test))

df.describe()



Realizacion de la graficacion de los datos en busca de los datos anómalos

fig, axes = plt.subplots(nrows=4, ncols=4, figsize=(30,30))

#creacion de histograma

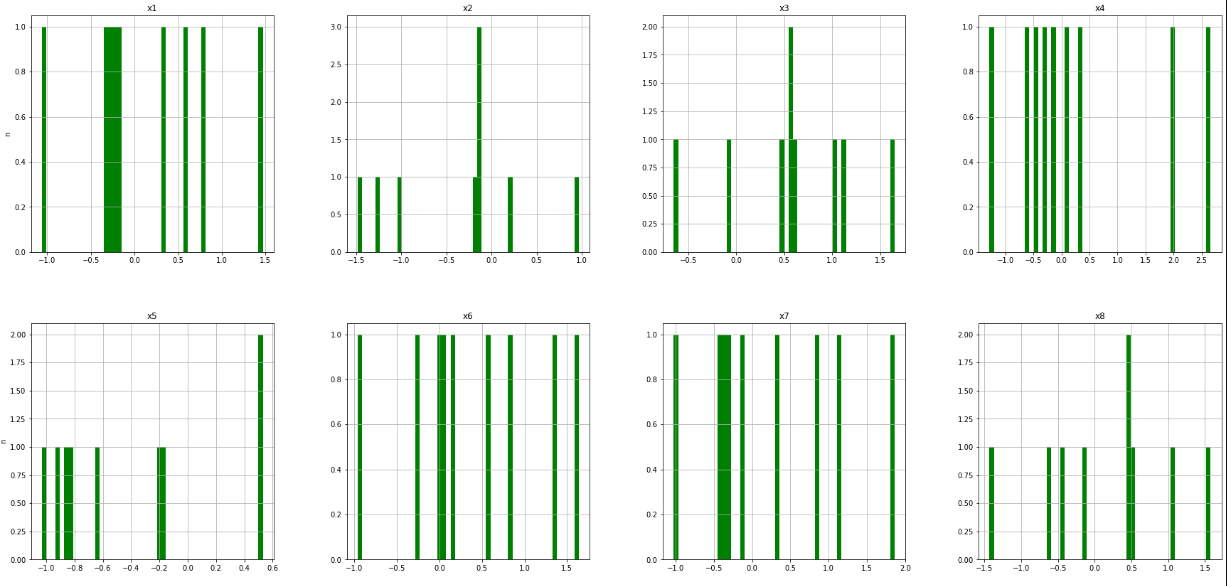
df.hist(bins=50, ax=axes, color="green");

plt.setp(axes[-1, :], xlabel = 'Valor x');

plt.setp(axes[:,0],ylabel ='n');

#mostrar todas las figuras

plt.show();



pregunta 2

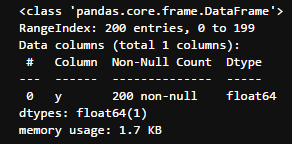
Busqueda de la variable y

nom\_y=["Y"]

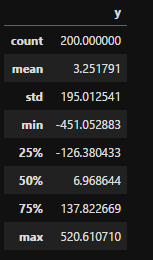
busqueda del data frame para la regresión de la variable y

dt\_fy= pd.DataFrame(data = y.flatten(), columns = nom\_y)

dt\_fy.info()



dt\_fy.describe()



plt.bar([0, 1], [105, 95], color= "g");

plt.xticks([0, 1]);

plt.xlabel("Valor de y");

plt.ylabel("n");

plt.show();

import matplotlib

import matplotlib.pyplot as plt

nom\_var = ["x%d" % i for i in range(1, predictores + 1)]

dt\_fX = pd.DataFrame(data= X[0:,0:], columns= nom\_var)

#funcion para la descripcion

dt\_fX.info()

from sklearn.linear\_model import LinearRegression

#ajustar el modelo

lm\_1= LinearRegression()

lm\_1.fit(X,Y)

LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=None, normalize=False)

print(lm\_1.intercept\_)

PREGUNTA3

Regresion lineal simple

import matplotlib

import matplotlib.pyplot as plt

nom\_var = ["x%d" % i for i in range(1, predictores + 1)]

dt\_fX = pd.DataFrame(data= X[0:,0:], columns= nom\_var)

#funcion para la descripcion

dt\_fX.info()

from sklearn.linear\_model import LinearRegression

#ajustar el modelo

lm\_1= LinearRegression()

lm\_1.fit(X,Y)

LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=None, normalize=False)

print(lm\_1.intercept\_)

#Cargar funciones

import statsmodels.api as sm

# intercepto

X1 = sm.add\_constant(dt\_fX)

#Comprobar df

X1.head()

#Ajustar el modelo

est\_x1= sm.OLS(Y,X1)

#Ver Ajuste

est\_1 = est\_x1.fit()

print(est\_1.summary())

impresion del resumen de datos del ajuste



cols=['x11','x12','x13','x14','x15','x16','x17','x18','x19','x20']

X4=X1.drop(cols,axis=1)

X4.head()



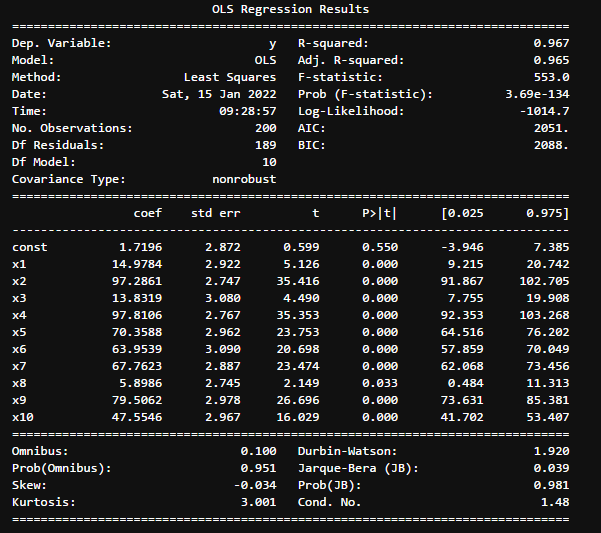
#Ajuste

est\_X4=sm.OLS(Y,X4)

est\_4=est\_X4.fit()

#resumen

print(est\_4.summary())



PREGUNTA4

#Aplicando el método Lasso y así obtener los valores esperados

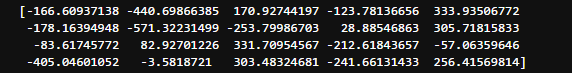
from sklearn.linear\_model import Lasso

lasso = Lasso(alpha = 1)

lasso.fit(Xval, Yval)

y\_pred1 = lasso.predict(Xval)

print(y\_pred1)



mean\_squared\_error = np.mean((y\_pred1 - Yval)\*\*2)

print(mean\_squared\_error)

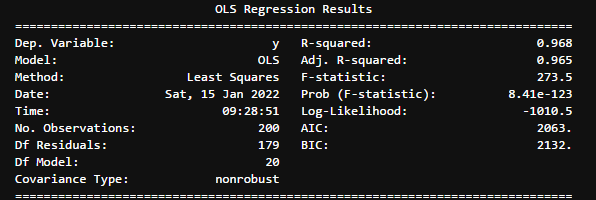


PREGUNTA 5

Análisis

De los ajustes realizados el que mejor resultado brindado fue el stepwise pues nos permite revisar paso a paso los datos y obntener un R2 ajustado mas real a las estimaciones en comparción con el métodos Lasso por lo que stepwise sería mejor aplicación

SE indica en la siguiente imagen



PREGUNTA TEÓRICA

Datos anómalos

Los datos anómalos son aquellos puntos que están fuera de los rangos de la regresión, este tipo de datos lleva a los resultados estadísticos a presentar fallas. Para identificar este tipo de datos es necesario representarlos de manera gráfica a todos los puntos de la muestra y demostrar si están agrupados de manera constante, es decir uno detrás de otro, de forma lineal, curva, etc, y de esta manera determinar los puntos excluidos. Para trabajar con estos datos es necesario se requiere aplicar modelos de regresión como la lineal, el cual determina los puntos anómalos y los elimina, si no se eliminan es necesario utilizar otro modelo de regresión, al puntos anómalos al ser trabajados con datos multivariantes son difíciles de detectar y eliminar, aunque una opción es el uso de regresión penalizada como las del tipo Lasso, red elástica, en caso de no poder eliminarlos con ningún modelo descrito anteriormente es necesario anularlos de forma manual.