| DATOS PERSONALES | FIRMA |
| --- | --- |
| |  |  | | --- | --- | | Nombre: Erick Wilfredo | DNI:2270018930101 | |  |
| Apellidos: Diaz Saborio |

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

| FECHA | MODELO | CIUDAD DEL EXAMEN |
| --- | --- | --- |
| 09-11/07/2021 | Modelo - C |  |

| Etiqueta identificativa |
| --- |
|  |

**INSTRUCCIONES GENERALES**

1. Ten disponible tu documentación oficial para identificarte, en el caso de que se te solicite.
2. Rellena tus datos personales en todos los espacios fijados para ello y lee atentamente todas las preguntas antes de empezar.
3. Las preguntas se contestarán en la lengua vehicular de esta asignatura.
4. Si tu examen consta de una parte tipo test, indica las respuestas en la plantilla según las características de este.
5. 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.
6. 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”.
7. 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.

**Puntuación**

**Examen**

* 10 10.00 puntos

El examen consta de cuatro preguntas.

Las puntuaciones son:

Pregunta 1: 1.5 puntos.

Pregunta 2: 1.5 puntos.

Pregunta 3: 2 puntos.

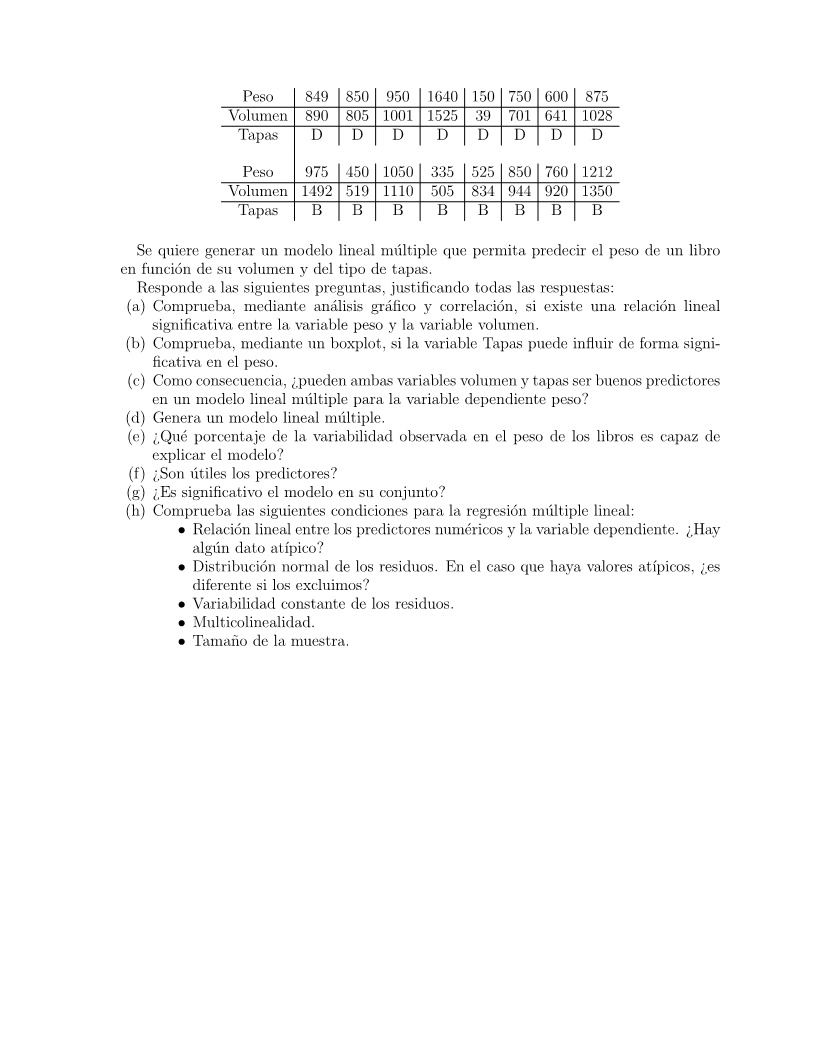
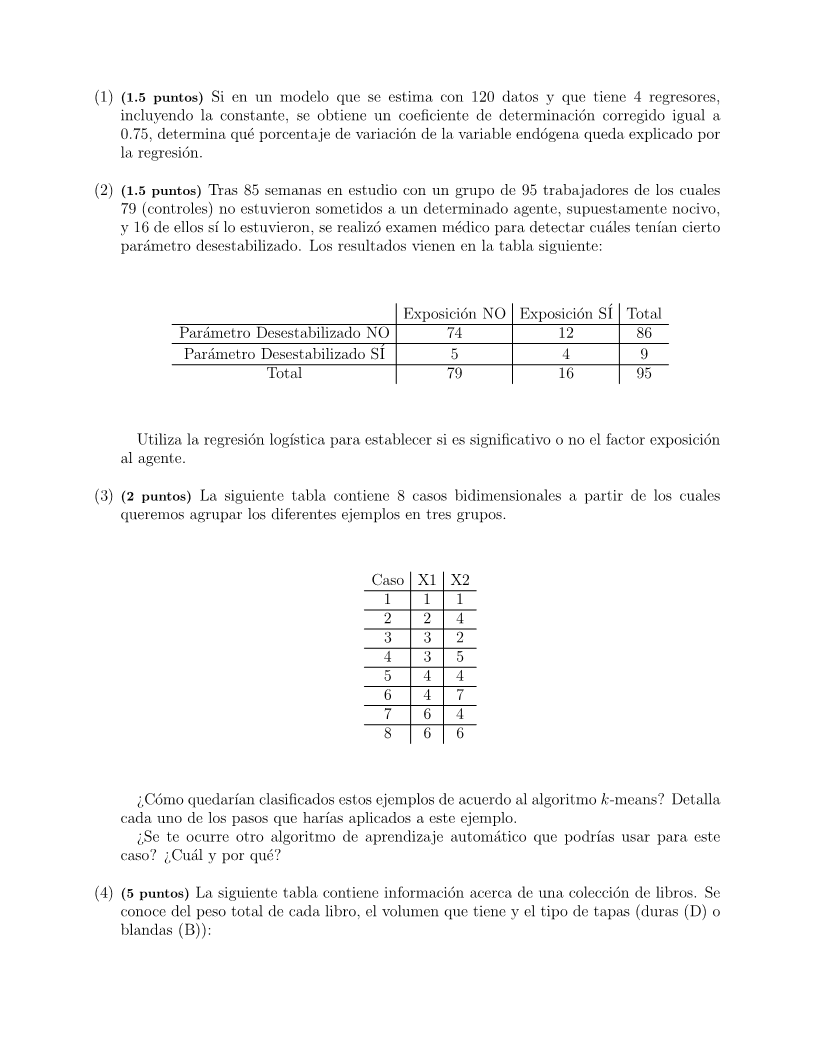
Pregunta 4: 5 puntos.

Responde a las preguntas en el espacio indicado entre las páginas 3 y 15.

Encontrarás las preguntas del examen a partir de la página 16.

¡Suerte!

**1.** Pregunta 1 (Responder en 1 caras)   
  
  
**2.** Pregunta 2 (Responder en 1 caras)   
  
  
**3.** Pregunta 3 (Responder en 2 caras)   
  
  
**4.** Pregunta 4 (Responder en 5 caras)



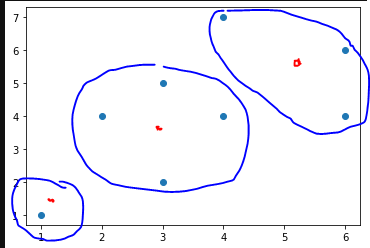
# Pregunta 1

El 75% de la variación de la variable es explicada por el modelo

# Pregunta 2

# Pregunta 3

Si queremos dividir los datos en 3 grupos con el algoritmo de k-means el primer paso de el algoritmo es seleccionar 3 puntos aleatorios en el espacio y de forma iterativa los va a ir ajustando minizando la distancia entre los puntos. Se realizo en Python un grafico de los puntos por lo que es probable que k-means agrupe de esta forma los puntos y en rojo se coloco la posible ubicación de el centroide:



Estos puntos son linealmente separable por lo que si tuviésemos las categorías seria muy simple utilizar un modelo de árbol.

# Pregunta 4

Creamos el datafatame de pandas

d = {

"peso": [849, 850, 950, 1640, 150, 750, 600, 875, 975, 450,1050, 335, 525, 850, 760, 1212],

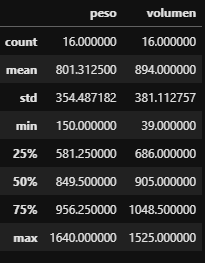
"volumen": [890, 805, 1001, 1525, 39, 701, 641, 1028, 1492, 519, 1110, 505,834, 944, 920, 1350],

"tapas": ['D', 'D', 'D', 'D', 'D', 'D', 'D', 'D', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B']

}

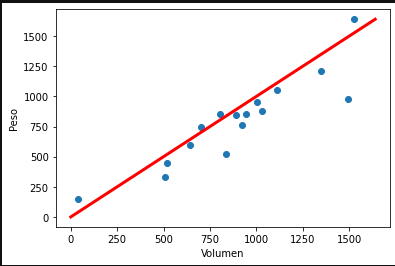
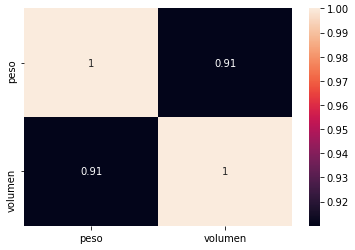
data = pd.DataFrame(d)

Obtenemos una descripción de los datos con el la función describe data.describe()



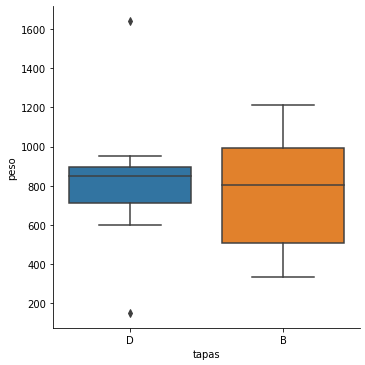
1. **Comprobación de relación lineal entre peso y Volumen**

Generamos la matriz de correlación para validar si existe relación entre la variable volumen y peso.



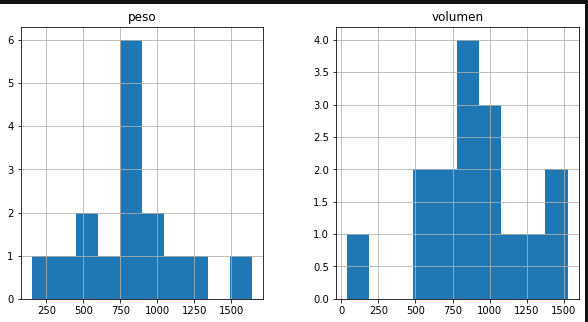
Podemos verificar con la matriz de correlación que existe una fuerte correlación entre la variable peso y la variable volumen y también podemos verificar de forma grafica que hay una relación lineal entra las 2 variables.

1. **Análisis de la variable tapas**



El rango intercuartil de los libros de pasta blande es mayor al de los de pasta dura y sus valores medios son bastante similares, se puede ver que los libros de pasta dura están en un rango alto de precios, pero los libros de pasta blanda cubren valores menores y mayores.

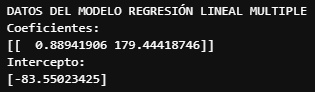
1. **Comprobación de variables predictoras**



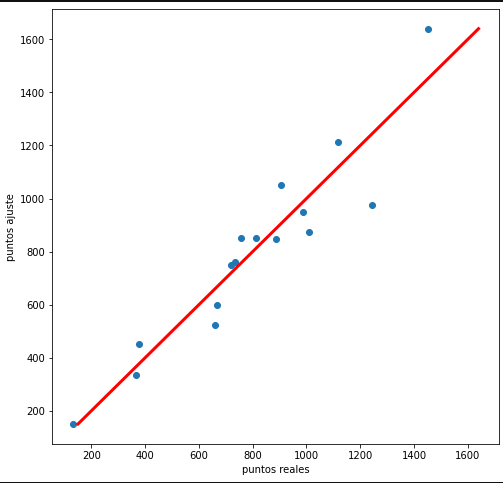
Obtenemos el histograma de las variables y parece ser que están distribuidas de forma normal, por lo que podemos utilizarlas para un modelo de regresión lineal, pareciera ser que el volumen tiene un data atípico, pero en general según los análisis gráficos y de correlación hechos en los incisos previos podemos utilizar ambas variables como predictoras.

1. **Modelo Lineal**

Se creo un modelo de regresión múltiple, utilizando las variables volumen y tapa, debido a que tapa es una variable categórica la encodeamos blanda = 0 y dura = 1.



Modelo lineal obtenido : con un =0.8943 podemos concluir que es un buen modelo por el valor de , tenemos un error medio cuadrático de 12452.55

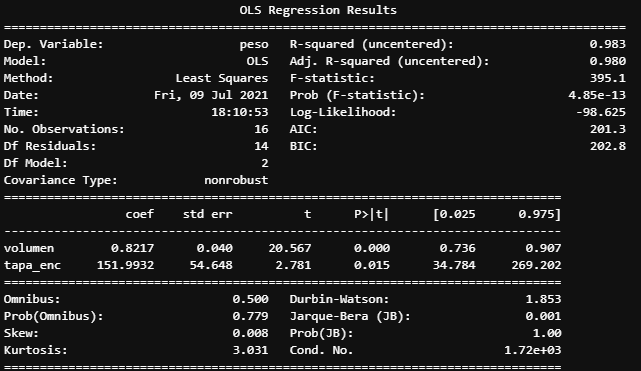


Podemos ver en la gráfica que los valores están muy cercanos a la línea por lo que las predicciones de el modelo son muy acertadas.

1. **Porcentaje de variabilidad en el peso de los libros**

Tenemos un valor de =0.8943 por lo que el modelo explica el 89% de la variabilidad en el peso de los libros.

1. **Son útiles los predictores**



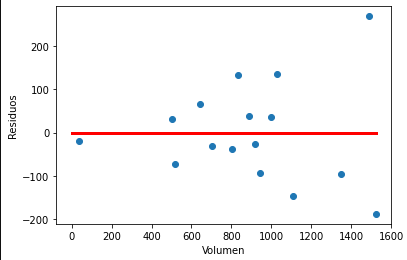
Podemos ver en la tabla que el F-statistic es = 4.58e-13 por lo que aceptamos el modelo.

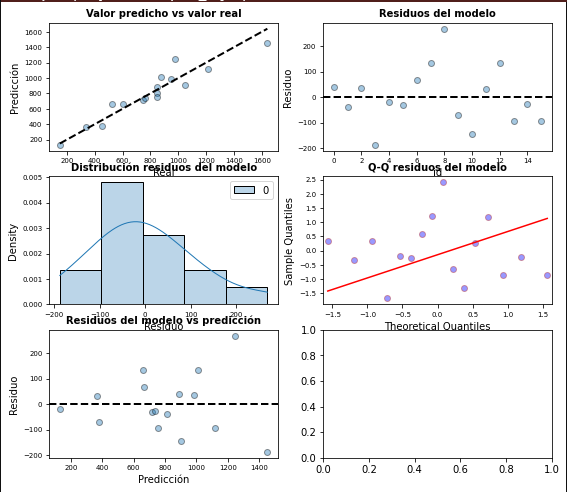
Tanto el volumen como la tapa son menores al p-value por lo que podemos decir que ambas son variables significativas para el modelo.

1. **Es significativo el modelo en su conjunto**

Podemos ver en la tabla que el F-statistic es = 4.58e-13 por lo que aceptamos el modelo ya que es significativo.

1. **Comprobación de condiciones**
   1. Relación Linealidad
   2. Distribución de residuos : podemos ver de forma grafica que los residuos se distribuyen sin un tendencia per si existen ciertos valores atípicos en los volúmenes mayores a 1200





* 1. Variabilidad de residuos : Corremos la prueba de Shapiro-Wilk para comprobar la normalidad y la prueba de Breush-Pagan para comprobar la homogeniedad de la varianza

Test Shapiro-Wilk, p.valor: 0.79126

El resultado del test Breusch-Pagan es: p.valor = 0.024

Con los resultados de los test comprobamos la normalidad de los residudos pero no la homogeneidad de la varianza.

* 1. Multicolinealidad: Se observa que la variable peso y volumen tienen valores VIF elevados por lo que podemos tener un problema de colinealidad

