| DATOS PERSONALES | FIRMA |
| --- | --- |
| |  |  | | --- | --- | | Nombre: | DNI: | |  |
| Apellidos: |

| 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.

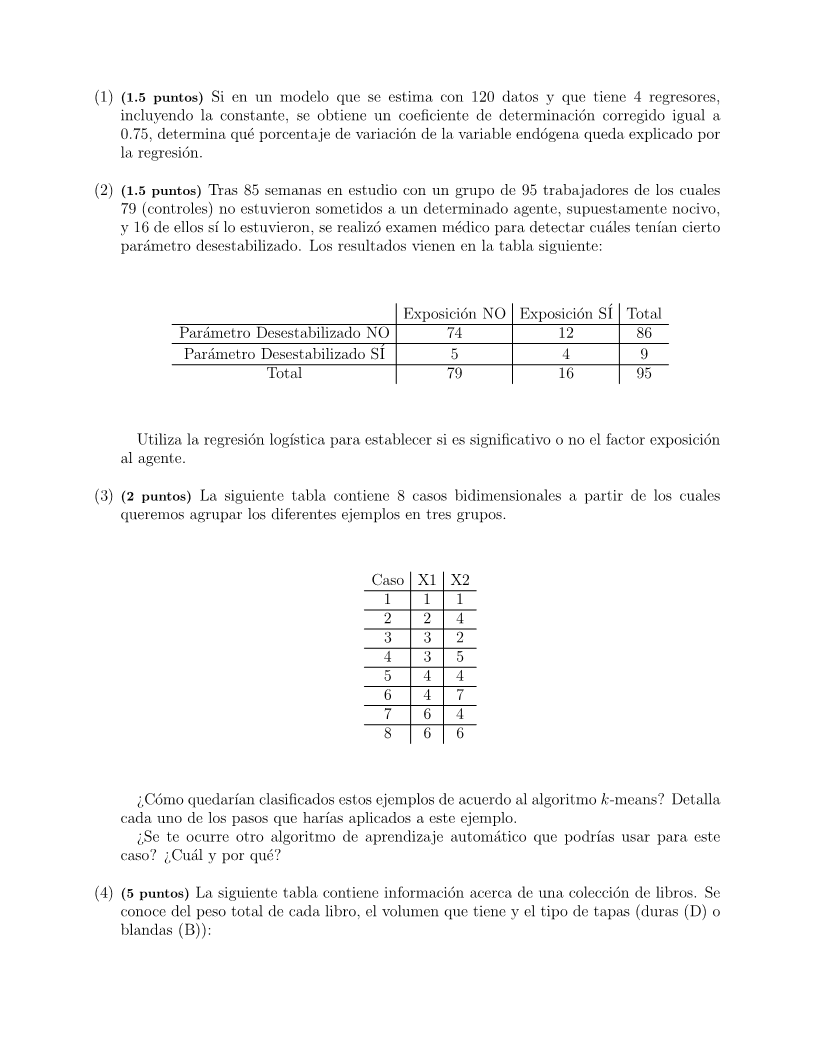
Pregunta 4: 5 puntos.

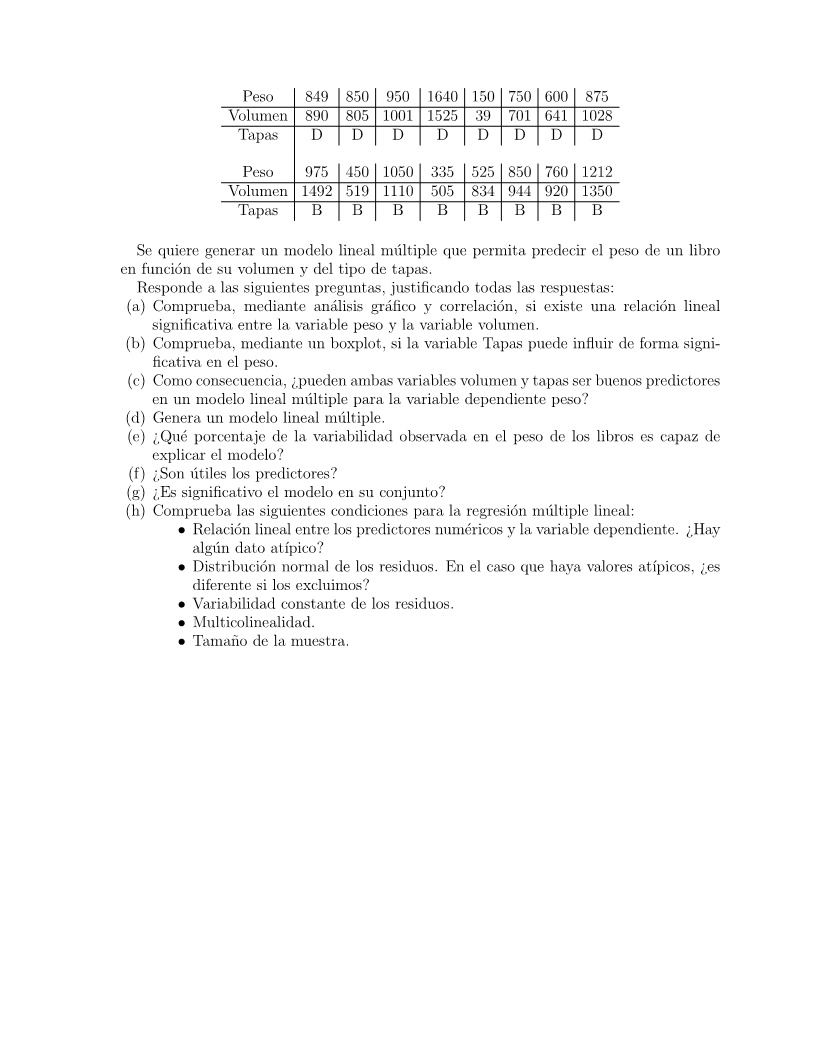
Responde a las preguntas en el espacio indicado entre las páginas 3 y 15.

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)





**EJERCICIO 1**

El modelo cuenta de 4 regresores, lo cual significa que tiene 4 variables de regresion. Al tener un coeficiente de determinacion corregido R2, se representa la proporcion de la varianza de la variable a predecir, o de la variable de la cual queremos obtener la regresion. Este coeficiente se puede representar como un cociente de la suma de cuadrados de la regresion, respecto a la suma total de estos mismos. Este coeficiente toma valores entre 0 y 1, por lo cual la medida de la variabilidad de los valores que se ajustan a la media muestral, es muy cercana a la media de la misma variabilidad de la muestra de la variable a predecir.

Un valor de R muy cercano a cero, indicaria que muy pocos valores se estan pegando a la respuesta de la prediccion, o en otras palabras, a la regresion. Un valor de R cercano a 1, indicaria un posible overfitting.

**EJERCICIO 3**

El metodo de creacionde clusters K Means es de tipo no supervisado, por lo cual por cada dato agregado se recalcula de nuevo el centroide.

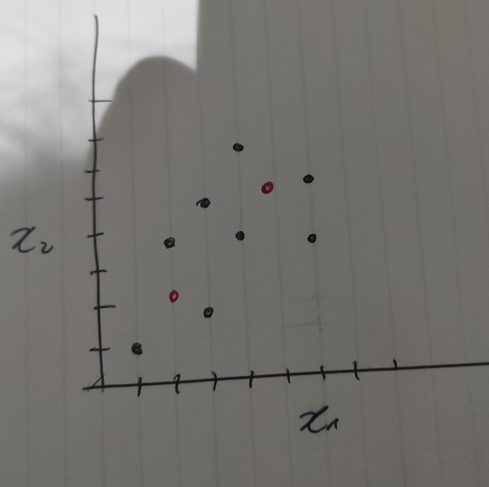
Si queremos clasificar en 3 grupos, debemos configurar 3 centroides, por lo cual cada centroide se obtendra der cada conjunto de datos obtenido.

K-means funciona separando los datos en clusters, por lo cual, la inercia obtenida es en relacion a “la distancia” que se encuentra el centroide de los datos. Entre mas cerca esten los centroides a un conjunto de datos, la inercia baja poco a poco. Para determinar el valor optimo de categorias, se obtiene el “Codo” visualizado en un grafico.

Para crear los clusters, partimos de centroides, o bien separando regiones siguiendo el proceso:

Generar K centroides, los cuales determinaran cuantas categorias tendremos. Luego asignar a cada observacion(o muestra, o fila)

, el centroide mas cercano. Se calculan de nuevo los centroides por cada dato agregado dentro del cluster(Esto tiende a mover un poco la posicion del centroide, y aproximarlo a una region mas precisa), y por ultimo, se vuelven a generar centroides, asignandolos a cada observacion, hasta que las observaciones ya no varian del cluster.



A consecuencia del tiempo del examen, no me ha dado tiempo de graficar k-means, pero los datos quedan distribuidos de la manera en que se observa en la fotografia. Se podria decir que los centroides posiblemente estan aproximados o cercanos a donde se encuentran los puntos rojos

Se podria utilizar otro algoritmo como KNN, solo que en este caso es supervisado, y hay que tener en cuenta que depende de la seleccion de K, es la precision del algoritmo.

**EJERCICIO 4**

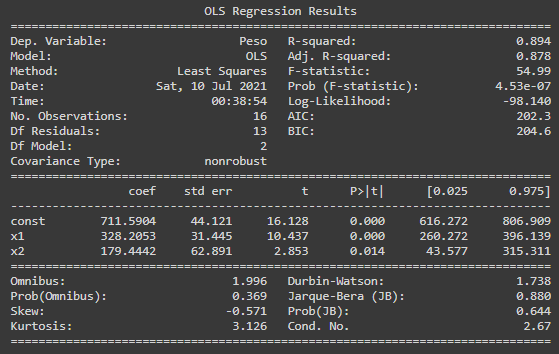
El codigo del ejercicio 4, he utilizado el codigo desarrollado en clase, con explicacion en el notebook y separandolo en el notebook. Sin embargo se ha desarrollado con las mismas variables utilizadas anteriormente de “housing” por el poco tiempo del examen. Se han desarrollado todos los pasos del notebook a pesar de no pedirse de la misma manera por el poco tiempo del examen. Gracias de antemano por su comprension al tema..

A) Si existe una relacion entre la variable peso y la variable volumen. El valor de correlacion obtenido es alto, cercano a 1, y el metodo empleado tiene como valor maximo 1, por lo cual se estima que la correlacion entre ambas variables es alta.

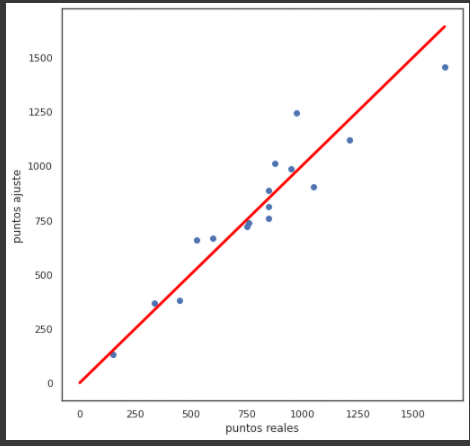
B) La variable Tapas puede influir significativamente en el peso, pero es mas significativo el volumen del libro hacia el peso correlativamente.

C) Ambas variables pueden ser buenos predictores, sin embargo en el caso de elegir un solo predictor, se optaria por una prediccion con el volumen del libro.

D) El modelo generado de regresion lineal multiple se ha hecho con los ejempos de clase, por lo cual las variables tienen el mismo nombre para ahorrar tiempo en el desarrollo.



Se observa que la variable categorica tiene un p valor aceptado menor a 0.05, sin embargo es mucho mejor la variable cuantitativa del volumen relacionado al peso. La regresion Lineal ha quedado de la siguiente manera:



F) Si son utiles los predictores. En el grafico de la regresion lineal se observa que ningun dato tiene outliers, se puede decir que cerca del 1000 en puntos reales de los datos, se tiene algo parecido a un outlier, pero no esta muy alejado del valor predicho. En este caso, puede ser posible que la estadistica robusta no funcione bien, sin embargo, son muy pocos valores para asegurar eso.

G) Si es significativo

H) En caso de excluir valores atipicos, se estan dejando de considerar muestras, por lo cual si es diferente si se dejan de considerar. Puede no ser tan notoria la variacion si nos basamos en estadistica robusta.