

NOMBRE: Moreno Villarreal Carlos Alejandro

FECHA: 28/09/2022

INDICACIONES: contestar en el documento de Word, debajo de la pregunta debe de ir la respuesta, utilizar color rojo o azul para diferenciar las respuestas, en cada ejercicio se requiere que AGREGES LAS OPERACIONES que justifiquen las respuestas.

Esta EVIDENCIA se sube en el apartado de TAREAS en TEAMS, BUSCAR la liga correspondiente a esta actividad y NO OLVIDES dar clic en ENVIAR.

PARTE TEORICA CUESTIONARIO:

1. ¿Qué es el análisis de correlación?

R= Grupo de técnicas para medir la asociación entre dos variables.

2. ¿Para qué sirve el coeficiente de correlación?

R= Número que indica la medida de la fuerza de la relación lineal entre dos variables.

3. ¿Cuáles son las características del coeficiente de correlación?

R=

1) El coeficiente de correlación de la muestra se identifica con la letra minúscula r .

2) Muestra la dirección y fuerza de la relación lineal (recta) entre dos variables en escala de intervalo o en escala de razón.

3) El valor de r varía entre -1 y 1 .

4) Un valor cercano a 0 indica que hay poca asociación entre las variables.

5) Un valor cercano a 1 indica una asociación directa o positiva entre las variables

6) Un valor cercano a -1 indica una asociación inversa o negativa entre las variables.

4. ¿Qué es el análisis de regresión?

R= La técnica para desarrollar la ecuación para representar la relación lineal entre dos variables y proporcionar las estimaciones se denomina **ANÁLISIS DE REGRESIÓN.**

5. ¿Qué es la ecuación de regresión?

R= La ecuación de regresión, es la ecuación que expresa la relación lineal entre dos variables..

6. ¿Qué es el principio de mínimos cuadrados?

R El principio de mínimos cuadrados determina una ecuación de regresión al minimizar la suma de los cuadrados de las distancias verticales entre los valores reales de Y , y los valores pronosticados de Y .

7. ¿Qué es el error estándar de estimación?
R= **El error estándar de estimación es una medida para describir cuan preciso es el pronóstico de Y de la recta de regresión con base en un valor de X dado.**
8. ¿Cuándo se utiliza el intervalo de confianza?
R= **Se utiliza cuando la ecuación de regresión se emplea para predecir el valor medio de Y para una X dada**
9. ¿Cuándo se utiliza el intervalo de predicción?
R= **Se utiliza cuando la ecuación de regresión se emplea para predecir una E individual (n=1) para una X dada**
10. ¿Con que función de Excel se calcula el valor del coeficiente de correlación?
R= **COEF.DE.CORREL(matriz1;matriz2)**
11. ¿Con que función de Excel se calcula el valor de b?
R= **Pendiente ()**
12. ¿Con que función de Excel se calcula el valor de a?
R= **INTERSECCION.EJE ()**
13. ¿Con que función de Excel se calcula el Error de correlación?
R= **Error.Tipico.XY()**
14. ¿Qué procedimiento se hace para generar el análisis de regresión en Excel?
R= **Ir a DATOS, seleccionar ANALISIS DE DATOS, seleccionar REGRESIÓN, seleccionar valores de Y, valores de X y rango de salida.**
15. ¿Cuál es la fórmula para la regresión lineal múltiple?
R=
$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$$
16. ¿Cómo se les llama a los valores que acompaña a las variables que forman la ecuación de regresión lineal múltiple?
R= **Coeficientes de regresión**
17. ¿Cómo se le llama al valor de a que aparece en la ecuación de regresión lineal múltiple?
R= **Intersección**
18. ¿Qué indica un signo negativo en el coeficiente de regresión para una variable?
R= **Una relación inversa entre las variables**
19. ¿Coeficiente que determina el porcentaje de la variación de la variable dependiente explicada por la variable independiente?
R= **Coeficiente de determinación**

20. ¿Cuáles son las características del coeficiente de determinación?

R= Ayuda a representar si en realidad mi ecuación se ajusta a los datos, se representa en porcentaje (que afecta y que no afecta)

1. Se representa por una letra R (mayúscula) al cuadrado. En otras palabras, se escribe como R² debido a que se comporta como el cuadrado de un coeficiente de correlación.
2. Puede variar de 0 a 1. Un valor cercano a 0 indica poca asociación entre el conjunto de variables independientes y la variable dependiente; un valor cercano a 1 significa una asociación fuerte.
3. No puede adoptar valores negativos. Ningún número que se eleva al cuadrado (o a la segunda potencia) puede ser negativo.
4. Es fácil de interpretar. Como R² es un valor entre 0 y 1, es fácil de interpretar, compararlo y comprender.

21. ¿Cuál es la fórmula para calcular el error estándar de la estimación múltiple?

R= $S_{y.1,2,3...k} = \sqrt{MSE}$ error de la media al cuadrado o MSE

22. ¿Qué es el clustering?

R= Es una técnica de Análisis Exploratorio de Datos para resolver problemas de clasificación, que intenta identificar, de forma automática, agrupaciones de elementos (también llamados conglomerados o clústeres homogéneos) de acuerdo con una medida de distancia o similitud entre ellos..

23. ¿A qué disciplina pertenece el clustering?

R= Inteligencia Artificial

24. ¿En que se basa el análisis de clustering?

R= se basa en la idea de distancia entre las observaciones.

25. ¿Cuáles son los 2 tipos diferentes de técnicas de Clustering?

R= Clustering de Partición y Clustering Jerárquico

26. ¿Cuál es el objetivo del método de partición K medias?

R= generar una partición de un conjunto de n observaciones en k grupos

27. ¿Cómo se determinan los primeros centros de método k medias?

R= Se eligen los que tú quieras de tu base de datos

28. ¿Cómo se determinan los nuevos centros del método k means?

R= Calculando el promedio de los datos de cada grupo

29. ¿Cómo se determina a que grupo pertenecen los elementos en el método k medias?

R= Se calculan las distancias de los centros a cada uno de los datos, se busca la menor distancia y se escribe el grupo al que pertenece la menor distancia.

30. ¿Cuándo termina el método k means?

R= Cuando las últimas dos columnas de los elementos que dan iguales, es decir, que no haya cambio

31. ¿Qué es y para qué sirve el método del codo?

R= Método gráfico que utiliza la mayor distancia de separación entre los elementos de los grupos, para determinar la cantidad ideal de grupos

PARTE PRÁCTICA EJERCICIOS:

1. El fabricante de equipo de cómputo desea estudiar la relación entre el número de meses desde la compra de un aparato y el tiempo que se utilizó el aparato la semana pasada.

Persona	Meses con el equipo	Horas de uso
Rodolfo	12	4
Homero	2	10
Benito	6	8
Lorenzo	9	5
Felipe	7	5
Melisa	2	8
Saúl	8	3
Karla	4	8
Marcela	10	2
Verónica	5	5

Cuadrante IV (4)

Cuadrante I (0)

Cuadrante III (1)

Cuadrante II (5)

a) Trace el diagrama de dispersión.

b) En base al diagrama de dispersión, Explique si existe o no alguna relación entre el número de meses con el aparato y el tiempo de uso.

R= Como la mayoría de los puntos están en los cuadrantes II Y IV, entonces la relación entre las variables Meses con el equipo y Horas de uso es NEGATIVA.

c) Calcule el coeficiente de correlación (r).

Coeficiente de correl= -0.82693964

d) En base al coeficiente de correlación indique que tipo de relación hay entre las variables.

R= Como r= -0.8269 y es un valor cercano a -1, entonces es una asociación INVERSA O NEGATIVA entre las variables.

x media=	6.5
y media=	5.8

e) Determina la ecuación de regresión. **Ecuación de Regresión:** $\hat{Y} = a + b(x)$

Sx= desviacion estandar de x	3.34165628	
Sy= desviacion estandar de y	2.57336788	
PENDIENTE	-0.63681592	
INTERSECCION.EJE	9.93930348	
Ecuacion de regresion= y prom = 9.9393 - 0.6368(x)		

f) Utiliza la ecuación de regresión para calcular las horas de uso que se estiman para un aparato que se compró hace 6 meses, un año y 1 ½ años.

$$R = \hat{Y} = 9.9393 - 0.6368(6) = 6.11$$

$$\hat{Y} = 9.9393 - 0.6368(12) = 2.29$$

$$\hat{Y} = 9.9393 - 0.6368(18) = -1.52$$

g) Determine el intervalo de confianza del 95% para la media pronosticada cuando una persona tiene 8 meses con el equipo.

INTERVALOS			
	n=	10	
	x media=	6.5	
	SUMA(x-xmedia)²=	100.5	
	gl=n-2=	8	
	nivel de confianza=	0.95=95%	
	valor de t=	2.306	
	Error de correl Sxy=	1.53475491	
	x=	8	
	y prom=	4.84477612	

Intervalo de confianza						
	0.02238806	0.12238806	1.2381346			
Lim.Inferior	3.60664152					
Lim.Superior	6.08291072					

El intervalo de confianza varia de 3.60 a 6.08

Interpretación: si una computadora que se compro hace 8 meses se utilizo entre 3 y 6 horas durante la semana pasada

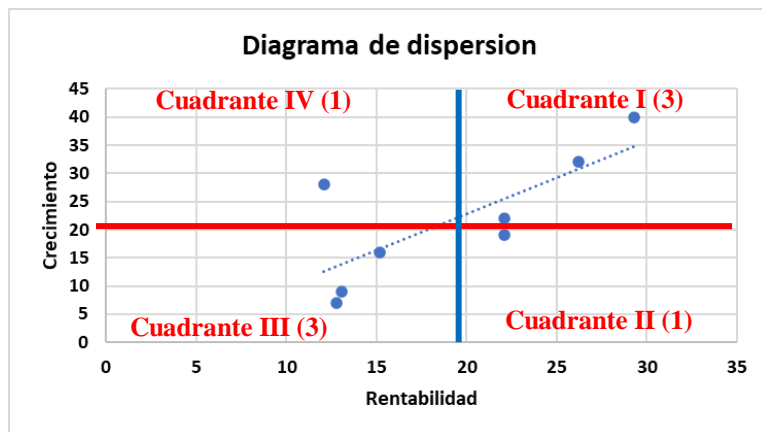
h) Establezca el intervalo de predicción del 95% para la media pronosticada cuando una persona tiene 8 meses con el equipo.

Intervalo de prediccion						
	0.02238806	1.12238806	3.74946975			
Lim.Inferior	1.09530637					
Lim.Superior	8.59424587					
El intervalo de prediccion varia del 1.09 y 8.59						
Todas las computadoras que se compraron hace 8 meses se utilizaron entre 1 y 8 horas durante la semana pasada.						

2. En la siguiente tabla se muestra el interés porcentual anual del capital y el crecimiento porcentual anual medio de las ventas de ocho tiendas y mueblerías que venden línea blanca. La rentabilidad de una compañía depende de entre otras cosas del crecimiento anual.

Compañía	Rentabilidad (porcentaje)	Crecimiento (porcentaje)
Electra	22.1	19
Muebleria Standar	15.2	16
Coppel	26.2	32
Muebleria Elizondo	13.1	9
Sears	12.1	28
Hemsa	12.8	7
Soriana	29.3	40
Famsa	22.1	22

- a) Trace el diagrama de dispersión.



x media=	19.1125
y media=	21.625

- b) En base al diagrama de dispersión, Explique si existe o no alguna relación entre el porcentaje de rentabilidad y el porcentaje de crecimiento.

R= Como la mayoría de los puntos están en los cuadrantes I Y III, entonces la relación entre las variables Rentabilidad y Crecimiento es POSITIVA.

- c) Calcule el coeficiente de correlación (r).

Coeficiente de correl	0.75444871
-----------------------	------------

- d) En base al coeficiente de correlación indique que tipo de relación hay entre las variables.

R= Como $r = 0.7544$ y es un valor cercano a 1, entonces es una asociación DIRECTA O POSITIVA entre las variables.

e) Determina la ecuación de regresión. (\hat{Y}) **Ecuación de Regresión:** $\hat{Y} = a + b(x)$

Sx= desviacion estandar de x	6.67991392	
Sy= desviacion estandar de y	11.32554	
PENDIENTE	1.2791391	
INTERSECCION.EJE	-2.82254611	
Ecuacion de regresion= y prom = -2.8225 + 1.2791(x)		

f) Utiliza la ecuación de regresión para calcular: ¿Cuál es el porcentaje de rentabilidad para una compañía con un porcentaje de crecimiento del 10% , 29% y 30%.

$$R = \hat{Y} = 2.8225 + 1.2791(10) = 9.96\%$$

$$\hat{Y} = 2.8225 + 1.2791(29) = 34.27\%$$

$$\hat{Y} = 2.8225 + 1.2791(30) = 35.55\%$$

g) Determine el intervalo de confianza del 95% para una compañía que tenga un porcentaje de crecimiento del 35%.

INTERVALOS			
	n=	8	
	x media=	19.1125	
	SUMA(x-xmedia)2=		312.34875
	gl=n-2=	6	
	nivel de confianza=	0.95=95%	
	valor de t=	2.447	
	Error de correl Sxy=		8.02922745
	x=	35	
	y prom=	41.94732	

Intervalo de confianza									
	0.80811163	0.93311163	18.9790526						
Lim.Inferior	22.9682699								
Lim.Superior	60.9263751								
El intervalo de confianza varia del 22.96 al 60.92									
Cuando una compañía tiene un porcentaje de crecimiento del 35%, su rentabilidad anual tiende estar entre el 23% y 61%.									

h) Establezca el intervalo de predicción del 95% para una compañía que tenga un porcentaje de crecimiento del 35%.

Intervalo de prediccion							
	0.80811163	1.93311163	27.3172009				
Lim.Inferior	14.63012						
Lim.Superior	69.2645234						
El intervalo de prediccion varia del 14.63 y 69.26							
Cuando todas las compa�as tiene un 35% de crecimiento, su rentabilidad anual se encuentra entre 15% y 69%.							

3.- Un estudio quiere generar un modelo que permita predecir la esperanza de vida media en a os de los habitantes de una ciudad en funci n de diferentes variables. Se dispone de informaci n sobre: la esperanza de vida, el n mero de habitantes de la ciudad, el n mero de asesinatos, el nivel de estudios universitarios (a os de estudio) y el n mero de heladas que ha sufrido a la ciudad.

X_1 =N mero de habitantes
 X_2 =N mero de asesinatos
 X_3 =A os de estudio
 X_4 =N mero de Heladas

La ecuaci n de regresi n lineal m ltiple para determinar la esperanza de vida que depende de la cantidad de habitantes, el n mero de asesinatos, el nivel de estudios y el n mero de heladas que mejor se ajusta a los datos es la siguiente:

$$\hat{y} = 71.03 + 0.00005014 x_1 - 0.3001 x_2 + 0.04658 x_3 - 0.00594 x_4$$

a)  Cu al es el nombre completo de la ecuaci n?

R= Ecuaci n de Regresi n M ltiple.

b) Interprete el n mero 71.03

R= El 71.03 representa la intersecci n con el eje de las Y.

c)  Cu al es la esperanza de vida para las personas que viven en una ciudad con 1,953,000 habitantes, 305 asesinatos registrados, un promedio de 9 a os de estudio por persona y 12 heladas hasta la fecha?

$$\hat{y} = 71.03 + 0.00005014 (1953000) - 0.3001(305) + 0.04658(9) - 0.00594(12)$$

$$\hat{y} = 71.03 + 97.92342 - 91.5305 + 0.41922 - 0.07128$$

$$\hat{y} = 77.77086$$

R= La esperanza de vida de los habitantes de una ciudad es de 78 a os.

d)  Cu al es el cambio que sufre la esperanza de vida cuando el n mero de asesinatos aumenta 60 unidades?

$$\hat{y} = 71.03 + 0.00005014 (1953000) - 0.3001(365) + 0.04658(9) - 0.00594(12)$$

$$\hat{y} = 71.03 + 97.92342 - 109.5365 + 0.41922 - 0.07128$$

$$\hat{y} = 59.76486$$

R= La esperanza de vida de los habitantes de una ciudad disminuye a 60 a os, por lo tanto, la esperanzade vida disminuyo 18 a os

e)  C mo se le denomina al n mero 0.04658?

R= Coeficiente de regresión de nivel de estudios de los habitantes

f) Conforme aumenta la población en miles de personas, ¿cuánto aumenta la esperanza de vida?

R= Conforme aumenta la población en miles de personas el estimado de esperanza de vida aumenta 0.050

4.- Se hizo un estudio para recoger información sobre el diámetro (en pulgadas), la altura (en pies) y el volumen (en pies cúbicos) del tronco de distintos cerezos en una determinada región.

Se encontró la ecuación de regresión lineal múltiple que determina la altura de los árboles en relación con el diámetro y el volumen del tronco

$$\hat{y} = 85.34 + 3.231 x_1 - 6.135 x_2$$

a) ¿Cuál es el nombre completo de la ecuación?

R= Ecuación de Regresión Múltiple

b) Interprete el número 85.34

R= El 85.34 representa la intersección con el eje de las Y.

c) ¿Cuál es la altura promedio de un árbol con un diámetro de 23 pulgadas y un volumen de 20 pies cúbicos?

$$\hat{y} = 85.34 + 3.231 (23) - 6.135(20)$$

$$\hat{y} = 85.34 + 74.313 - 122.7$$

$$\hat{y} = 36.953$$

d) ¿Cuál es el cambio que sufre la altura de un árbol cuando el diámetro de su tronco reduce un 25%?

$$\hat{y} = 85.34 + 3.231 (17.25) - 6.135(20)$$

$$\hat{y} = 18.37475$$

e) ¿Cómo se le denomina al número - 6.135?

R= Coeficiente de regresión del volumen

f) Conforme aumenta el volumen en un pie cubico ¿Cuánto es la disminución de la altura de los cerezos?

R= Conforme aumenta el volumen en un pie cubico, la altura de los cerezos disminuye 6.135 metros.

5.- La siguiente tabla incluye información acerca del peso (en kg.), la altura (en cm.), contorno de cintura (en cm.) y la edad de 12 individuos.

peso	altura	cintura	edad
74	168	62	25
92	196	75	31
63	170	60	29
72	175	71	64
58	162	66	44
78	169	62	41
85	190	79	37
85	186	74	35
73	176	70	34
62	170	66	29
80	176	71	19
72	179	69	50

- a) Determine la ecuación de regresión que mejor se ajuste a los datos proporcionados.

Coefficientes

Intercepción	-78.0301678
altura	0.9362915
cintura	-0.13261301
edad	-0.09671544

y=	Peso
X1=	Altura
X2	Cintura
X3	Edad

$$y = -78.03 + 0.9362 x_1 - 0.1326 x_2 - 0.0967 x_3$$

- b) Interprete cada coeficiente que compone a la ecuación de regresión lineal.

R=

- El numero -78.03 representa la intersección con el eje de las Y.
- El numero 0.9362 es el coeficiente de regresión de la variable x1 Altura por ser un valor positivo la relación que se tiene entre x1 y Y que es el peso es POSITIVO o DIRECTA, quiere decir que entre más altura tenga, el peso aumentara y si aumenta un cm, en la variable peso aumenta 0.93kg.
- El numero -0.1326 es el coeficiente de regresión de la variable x2 Cintura por ser un valor negativo la relación que se tiene entre x2 y Y es INVERSO o INDIRECTO, significa que no afecta, esto quiere decir que si aumenta cm de cintura la variable peso disminuye 0.13kg.
- El numero -0.0967 es el coeficiente de regresión de la variable x3 que es Edad se tiene que la relación de x3 y Y es INVERSA o NEGATIVA ya que su valor es negativo, es decir si la edad aumenta entonces el peso disminuye 0.09kg, por lo tanto, no afecta la edad para que aumente su peso.

- c) Encuentra el valor del error estándar de estimación múltiple e interprétalo.

Error típico	6.02432526
--------------	------------

En la ecuación el error es de 6.024 de kg de peso

se espera que los residuos sean aproximados a una distribución más o menos normal, por lo que alrededor de 74% de ellos estará dentro de 6.02 y cerca de 95% dentro de $\pm 2(6.02) = 12.04$

- d) Encuentra el valor de coeficiente de determinación e interprétalo

Coeficiente de determ

TOTAL	100%
% COEF.DETERM	74.64%
	25.36%

$R = 0.7464 * 100 = 74.64\%$

El 74.64% del valor de Y Peso se debe a la Altura, Cintura y Edad entonces hay un 25.36% que se debe a otros factores, error aleatorio o variables no incluidas en el análisis.

- e) ¿Son eficaces estas tres variables para explicar la variación del peso de las personas?

$R =$ Las 3 variables afectan de forma distinta el peso de las personas, pero solo un 74.64% significa que hay que considerar otras variables que también afectan el aumento o disminución del peso

6.- Se pretende estudiar la posible relación lineal entre el precio de pisos en miles de euros, en una conocida ciudad española y variables como la superficie en m² y la antigüedad del inmueble en años. Para ello, se realiza un estudio, en el que se selecciona de forma aleatoria una muestra estratificada representativa de los distintos barrios de la ciudad. Los datos aparecen en la siguiente tabla.

Precio	Superficie	Antigüedad
200	100	20
120	70	15
155	120	30
310	150	20
320	90	12
400	227	400
100	75	100
80	65	80
75	80	75
169	150	169
110	120	110
210	100	210
200	125	200
180	137	180
140	90	140
95	110	95

Precio	Superficie	Antigüedad
200	100	20
120	70	15
155	120	30
310	150	20
320	90	12
400	227	400
100	75	100
80	65	80
75	80	75
169	150	169
110	120	110
210	100	210
200	125	200
180	137	180
140	90	140
95	110	95

- a) Ajustar un modelo de regresión lineal múltiple.

	<i>Coeficientes</i>
Intercepción	-19.05123892
Superficie	1.924121322
Antigüedad	-0.168058

y=	Precio
X1=	Superficie
X2	Antigüedad

$$\hat{y} = -19.05 + 1.9241x_1 - 0.1680x_2$$

- b) Obtener una estimación de los parámetros del modelo y su interpretación

- El -19.05 representa la intersección con el eje de las Y
- El numero 1.924 es el coeficiente de regresión de la variable x1 Superficie por ser un valor positivo la relación que se tiene entre x1 y Y que es el precio del piso es POSITIVO o DIRECTA, quiere decir que entre más superficie tenga, el precio del piso va a aumentar, es decir si aumenta un m2 en la variable x1, en la variable del precio aumenta 1.92 miles de euros.
- El numero -0.1680 es el coeficiente de regresión de la variable x2 que es la Antigüedad se tiene que la relación de x2 y Y es INVERSA o NEGATIVA ya que su valor es negativo, es decir si tienen muchos más años de antigüedad aumenta entonces el precio del piso disminuye 0.16 miles de euros.

- c) ¿Cuánto será el precio estimado del piso en una conocida ciudad española para una superficie en 130 m² y 35 años de antigüedad?

$$\hat{y} = -19.05 + 1.9241(130) - 0.1680(35)$$

$$\hat{y} = 225.20$$

El precio estimado del piso es de 225.20 euros

- d) ¿Puede eliminarse alguna variable del modelo? Razona la respuesta

R= Coeficiente de determinación (Todas las variables)

<i>Estadísticas de la regresión</i>	
Coeficiente de correlación múltiple	0.717627464
Coeficiente de determinación R ²	0.514989178
R ² ajustado	0.440372128
Error típico	70.1976492
Observaciones	16

Coeficiente de determinación (Sin la variable Antigüedad)

<i>Estadísticas de la regresión</i>	
Coeficiente de correlación múltiple	0.70548913
Coeficiente de determinación R ²	0.497714913
R ² ajustado	0.461837407
Error típico	68.83822356

Observaciones	16
---------------	----

Coeficiente de determinación (Sin la variable Superficie)

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.392449673
Coeficiente de determinación R^2	0.154016746
R^2 ajustado	0.09358937
Error típico	89.33782923
Observaciones	16

Incluiría la variable independiente superficie, y quitaría la variable Antigüedad, ya que en el coeficiente de determinación que no aportaba mucho y no hacía mucha diferencia dentro de la ecuación en comparación de las demás.

e) Calcula el valor del error estándar de estimación múltiple e interprétalo

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados
Regresión	2	68019.7706	34009.8853
Residuos	13	64060.2294	4927.709954
Total	15	132080	

Error típico	70.1976492
--------------	------------

En esta ecuación se emplea un error de 70.19 en el precio de pisos en miles de euros se espera que los residuos sean aproximados a una distribución normal, por lo que alrededor de 51% de ellos estará dentro de 70.19 y cerca de 95% dentro de $\pm 2(70.19) = 140.38$ miles de euros.

f) Calcula el coeficiente de determinación y da su interpretación.

Coeficiente de determinación R^2	0.514989178	TOTAL	100%
		% COEF.DETERM	51.49%
			48.51%

$$0.5149 \times 100 = 51.49\%$$

El 51.49% del valor de Y (Precio) se debe a la Superficie y a la Antigüedad del Inmueble en años entonces hay un 48.51% que se debe a otros factores más importantes.

7.- Utiliza los datos que se muestran en la tabla para determinar por el método de K means los elementos que deben formar 3 grupos de frutas. NO olvides convertir en variables numéricas el color y el sabor de las frutas.

EJERCICIO 7									
	fruta	nombre	color	Peso promedio kg	altura promedio cm	ancho promedio cm	sabor	nivel de azúcar por cada 100 gramos de fruta	INTERACCION 1
	1	sandía	1	30	40	30	1	6.2	1
	2	melón	2	2.5	18	20	1	7.86	2
	3	papaya	2	2.7	25	18	1	5	2
	4	manzana	3	0.2	10	10	1	10.31	3
	5	pera	1	0.22	12	10	1	9.5	3
	6	naranja	2	0.18	10	10	1	9	3
	7	fresa	3	0.02	6	6	1	4.89	3
	8	uva	4	0.01	2	1.5	1	16.25	3
	9	toronja	2	0.35	15	16	2	7	2
	10	limón	1	0.018	4	4	2	2.5	3
PASO 1									
K=3	3 GRUPOS								
PASO 2									
Centros									
Centro 1	1	30	40	30	1	6.2	1		
Centro 2	1	0.22	12	10	1	9.5	5		
Centro 3	1	0.018	4	4	2	2.5	10		
PASO 3									
Centro 1	Centro 2	Centro 3	PASO 4						
			INTERACCION 1	INTERACCION 2	INTERACCION 3	INTERACCION 4			
1	0	62.57	89.68	1	1	1	1	1	
2	62.16	5.56	27.52	2	2	2	2	2	INTERPRETACION
3	56.5	8.47	35.58	2	2	2	2	2	Se tiene que
4	85.91	24.01	8.24	2	2	3	3	3	Grupo 1: Sandía
5	81.08	21.18	10.12	2	2	2	3	3	Grupo 2: Melon, Papaya y Toronja
6	83.62	21.72	6.58	2	2	3	3	3	Grupo 3: Manzana, pera, naranja, fresa, uva y limon
7	91.29	30.23	7.02	3	3	3	3	3	
8	109.54	47.64	20.19	3	3	3	3	3	
9	71.45	8.88	19.90	2	2	2	2	2	
10	96.682	36.95	14.75	3	3	3	3	3	

8.- El fabricante de equipo de cómputo desea agrupar una muestra de clientes para programar mantenimiento de sus equipos en relación entre el número de meses desde la compra de un aparato y el tiempo que se utilizó el aparato la semana pasada.

Utiliza los datos que se muestran en la tabla para determinar, por el método de K means los elementos que deben pertenecer a 4 grupos de personas.

Tomar como centros: Benito, Felipe, Saúl y Marcela.

EJERCICIO 8

	Persona	Meses con el equip	Horas de uso	INTERACCION
1	Rodolfo	12	4	1
2	Homero	2	10	2
3	Benito	6	8	2
4	Lorenzo	9	5	3
5	Felipe	7	5	3
6	Melisa	2	8	2
7	Saúl	8	3	4
8	Karla	4	8	2
9	Marcela	10	2	4
10	Verónica	5	5	3

PASO 1

K=4 4 Grupos

PASO 2

Centros

Centro 1	12	4	1
Centro 2	6	8	3
Centro 3	7	5	5
Centro 4	8	3	7

PASO 5
PASO 3

Centro 1	Centro 2	Centro 3	Centro 4
1	0	13	6
2	16	3	10
3	10	3	4
4	4	9	2
5	6	7	0
6	14	2	8
7	5	10	3
8	12	1	6
9	4	13	6
10	8	5	2

PASO 4

INTERACCION 1	INTERACCION 2
1	1
2	2
2	2
3	3
3	3
2	2
4	4
2	2
4	4
3	3

INTERPRETACION

Para el grupo 1 quedaron: Rodolfo
 Grupo2: Homero, Benito, Melisa y Karla
 Grupo3: Lorenzo, Felipe y Veronica
 Grupo4: Saul y Marcela

PASO 6

Centro 1	12	4
Centro 2	3.5	8.5
Centro 3	7	5
Centro 4	9	2.5