INSTITUTO TECNOLOGICO SUPERIOR DE SAN ANDRES TUXTLA

DOCENTE:

Pablo Promotor Campechano

ALUMNO:

Julio Cesar Hernández Domínguez

PROGRAMA EDUCATIVO:

Ingeniería Industrial

AIGNATURA:

Estadística Inferencial I

PRODUCTO ACADEMICO:

EXAMEN UI

FECHA: 01-03-2024

GRUPO: 401C

En un experimento para determinar los factores que afectan el ahorro de combustible en camiones, se midió el consumo de combustible en millas por galón (mi/gal), el peso del camión en toneladas y la lectura del odómetro en miles de millas en 15 camiones.

En la siguiente tabla se muestran los datos del experimento.

MPG	Peso	Odómetro
7.28	10.5	15
5.63	23.0	71
5.26	27.5	36
6.58	14.5	113
5.01	30.5	39
6.73	14.0	97
5.37	21.0	195
7.28	8.5	8
4.85	26.0	84
5.08	26.5	25
5.51	15.0	124
4.75	30.0	25
6.03	15.0	75
5.26	22.5	192
5.60	16.0	139

PORCENTAJE OBTENIDO 50%

Determine

a) La ecuación de la regresión lineal múltiple

b) Pronostique las millas por galón para un camión que pesa diez toneladas y tiene una lectura del

odómetro de 50000 millas.

c) El error estándar

d) El coeficiente de determinación

e) El coeficiente de correlación

Numero de muestras	y	x_1	x_2	x_1^2	x_2^2	$x_1 \cdot x_2$	$x_1 \cdot y$	$x_2 \cdot y$
1	7.28	10.5	15	110.25	225	157.5	76.44	109.2
2	5.63	23	71	529	5041	1633	129.49	399.73
3	5.25	27.5	36	756.25	1296	990	144.65	189.36
4	6.58	14.5	113	210.25	12769	1638.5	95.41	743.59
5	5.01	30.5	39	930.25	1521	1189.5	152.805	195.39
6	5.73	14	97	196	9409	1358	74.22	652.81
7	5.37	21	195	441	38025	4095	112.77	1047.15
8	7.28	8.5	8	72.25	64	68	61.88	58.24
9	4.85	26	84	676	7056	2184	126.1	407.4
10	5.08	26.5	25	702.25	625	662.5	134.62	127
11	5.51	15	124	225	15376	1860	82.65	683.24
12	4.75	30	25	900	625	750	142.5	118.75
13	6.03	15	75	225	5625	1125	90.95	452.25
14	5.26	22.5	192	506.25	36864	4320	118.35	1009.92
15	5.6	16	139	256	19321	2224	89.6	778.4
Total =	86.22	300.5	1238	6735.75	153842	24255	1651.935	6972.38

J

$\sum y = 86.22$
$\sum x_1 = 300.5$
$\sum x_2 = 1238$
$\sum x_1^2 = 6735.75$
$\sum x_2^2 = 153842$
$\sum x_1 \cdot x_2 = 24255$
$\sum x_1 \cdot y = 1651.935$
$\sum x_2 \cdot y = 6972.38$

$\sum y = na + b_1 \sum x_1 + b_2 \sum x_2$ G1
$\sum x_1 y = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 \cdot x_2$ G2
$\sum x_2 y = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2$ G3

$86.22 = 15a + 300.5b_1 + 1238b_2$
$1651.935 = 300.5a + 6735.75b_1 + 24255b_2$
$6972.38 = 1238a + 24255b_1 + 153842b_2$

n=15

Resolviendo el sistema de ecuaciones:

$a = 8.2406$
$b_1 = -0.1082$
$b_2 = -0.0039$

a) $\hat{y} = a + b_1 x_1 + b_2 x_2$
$\hat{y} = 8.24 - 0.1082 x_1 - 0.0039 x_2$

b) $\hat{y} = 8.24 - 0.1082(10) - 0.0039(50)$
$\hat{y} = 6.963$ millas por galón

Numero de muestras	y	x_1	x_2	$\hat{y}$ 2.000	SCE $(y_i - \bar{y})^2$	SCR $(y - \hat{y})^2$	SCT $(y_i - \hat{y})^2$
1	7.28	10.5	15	7.04	0.659	1.68	2.34
2	5.53	28	71	5.47	0.025	0.079	0.013
3	5.26	22.5	36	5.12	0.018	0.388	0.238
4	6.58	14.5	113	6.23	0.121	0.233	0.692
5	5.01	20.0	39	4.78	0.049	0.920	0.544
6	6.78	14	97	6.34	0.146	0.359	0.964
7	5.37	21	195	5.20	0.026	0.291	0.142
8	7.28	8.5	8	7.29	9.46	2.376	2.347
9	4.85	26	84	5.09	0.062	0.420	0.806
10	5.08	26.5	25	5.27	0.038	0.222	0.446
11	5.81	15	124	6.13	0.389	0.118	0.056
12	4.75	30	25	4.89	0.021	0.724	0.996
13	6.03	15	75	6.32	0.087	0.337	0.070
14	5.26	22.5	192	5.05	0.041	0.477	0.2381
15	5.6	16	139	5.96	0.137	0.098	0.021
	86.22	300.5	1238	86.26	1.21	8.704	9.93

SCE = 1.2154

SCR = 8.70

SCT = 9.93

c)

$$Sc = \sqrt{\frac{SCE}{n-3}} \qquad Sc = \sqrt{\frac{1.21}{15-3}} \qquad Sc = 0.31 \text{ MPS}$$

d)

$$R^2 = \frac{SCR}{SCT} \qquad R^2 = \frac{8.70}{9.93} \qquad R^2 = 0.87$$

e)

$$\sqrt{R^2} = R = \sqrt{0.87} \quad R = 0.935$$

LISTA DE COTEJO DE INVESTIGACIÓN DOCUMENTAL

DOCENTE: PABLO PROMOTOR CAMPECHANO	ASIGNATURA: ESTADISTICA INFERENCIAL II	
DATOS GENERALES DEL PROCESO DE EVALUACIÓN		
NOMBRE DEL ALUMNO: HERNANDEZ DOMINGUEZ JULIO CESAR	UNIDAD: I	
PERIODO: FEBRERO-JUNIO 2024	GRUPO: 401 C	FECHA DE ENTREGA: 23/02/2024
INSTRUCCIONES		
Revisar las actividades que se solicitan y marque en los apartados "SI" cuando la evidencia se cumple; en caso contrario marque "NO". En la columna "OBSERVACIONES" indicaciones que puedan ayudar al alumno a saber cuáles son las condiciones no cumplidas, si fuese necesario.		

VALOR DEL REACTIVO	CARACTERÍSTICA A CUMPLIR (REACTIVO)	CUMPLE		OBSERVACIONES
		SI	NO	
3%	**PRESENTACIÓN:** la investigación cumple con los requisitos de: **a.** Buena presentación **b.** letra legible **c.** Limpieza y orden **d.** Ortografía (El documento es redactado de forma correcta sin faltas de ortografía)	√		
2%	Maneja el lenguaje técnico apropiado y presenta en todo el documento coherencia y secuencia entre párrafos	√		
2%	**INTRODUCCIÓN**: Da una idea clara y objetiva de lo que tratara el tema (motivando al lector a continuar con su lectura y revisión), fundamentando con las referencias bibliográficas que se utilizaron para su redacción.	√		
8%	**CONTENIDO**: Maneja un lenguaje técnico apropiado y presenta en todo el documento coherencia, secuencia entre párrafo, es digerible a todo público y presenta una metodología **COHERENCIA Y COHESIÓN:** Maneja un lenguaje técnico apropiado y presenta en todo el documento coherencia, secuencia entre párrafo y es digerible a todo público coherente.	√		
3%	**Conclusiones**: Las conclusiones son claras y acordes con el objetivo esperado.	√		
2%	**Responsabilidad**: Entregó la investigación documental en la fecha señalada.	√		
20%	CALIFICACIÓN	20%		

LISTA DE COTEJO DE PROBLEMARIO

DOCENTE: ING. PABLO PROMOTOR CAMPECHANO		ASIGNATURA: ESTADISTICA INFERENCIAL II
DATOS GENERALES DEL PROCESO DE EVALUACIÓN		
NOMBRE DEL ALUMNO (A): HERNANDEZ DOMINGUEZ JULIO CESAR		Problemario de la Unidad: 1
PERIODO: FEBRERO-JUNIO 2024	GRUPO: 401 C	FECHA DE ENTREGA: 29/02/2024
INSTRUCCIONES		
Revisar las actividades que se solicitan y marque en los apartados "SI" cuando la evidencia se cumple; en caso contrario marque "NO". En la columna "OBSERVACIONES" indicaciones que puedan ayudar al alumno a saber cuáles son las condiciones no cumplidas, si fuese necesario.		

VALOR DEL REACTIVO	**CARACTERÍSTICA A CUMPLIR (REACTIVO)**	**CUMPLE**		**OBSERVACIONES**
		SI	**NO**	
5 %	**PRESENTACIÓN:** El trabajo cumple con los requisitos de **a.** Buena presentación **b.** No tiene faltas de ortografía **c.** Ordenado y limpio	√		
5 %	**FORMATO DE ENTREGA**: Hoja de presentación (asignatura, unidad, tema de estudio, docente, fecha, nombre del alumno	√		
10 %	**DESARROLLO DE EJERCICIOS**: Identifica los principios, leyes, normas e incluso técnicas y metodologías apropiadas, si el ejercicio lo permite, debe de presentar datos, fórmula, sustitución y resultado.	√		
5 %	**RESULTADO:** El alumno llega al resultado correcto, con sus respectivas unidades	√		
5 %	**RESPONSABILIDAD**: Entregó el problemario en la fecha y hora señalada.	√		
30 %	**CALIFICACIÓN**	30%		

INGENIERIA INDUSTRIAL
AEF - 1025

INSTITUTO TECNOLOGICO SUPERIOR DE SAN ANDRES TUXTLA

MATERIA: ESTADISTICA INFERENCIAL II

DOCENTE: PABLO PROMOTOR CAMPECHANO

INTEGRANTES:

MARIA FERNANDA ALEMAN GONZALEZ
KARLA LISSETTE FILIDOR DOMINGUEZ
JOSE GABRIEL FISCAL MEMECHI
JULIO CESAR HERNANDEZ DOMINGUEZ
IVAN JAIR MIXTEGA ANOTA

CARRERA: INGENIERIA INDUSTRIAL

GRUPO: 401-C SEMESTRE: CUARTO SEMESTRE

UNIDAD 1: REGRESION LINEAL MULTIPLE

ACTIVIDAD: INVESTIGACION DOCUMENTAL

FECHA DE ENTREGA: 23/02/2024

SAN ANDRES TUXTLA, VER

Índice

INTRODUCCIÓN

La regresión lineal múltiple es una poderosa técnica estadística utilizada para modelar la relación entre una variable dependiente y múltiples variables independientes.

Es una extensión de la regresión lineal simple que permite analizar y predecir el comportamiento de una variable dependiente en función de dos o más variables independientes.

Los intervalos de confianza y predicción juegan un papel crucial en la evaluación de la precisión de los parámetros del modelo y en la estimación de la incertidumbre en las predicciones futuras. Los intervalos de confianza proporcionan un rango dentro del cual es probable que se encuentre el verdadero valor de un parámetro, mientras que los intervalos de predicción nos brindan un rango de valores dentro de los cuales es probable que se encuentren las futuras observaciones de la variable dependiente.

Para ello analizaremos estos temas que son importantes en la estadística y haremos una retroalimentación de ello

1.1 regresión lineal múltiple

La regresión lineal múltiple es una técnica estadística que se utiliza para modelar la relación entre una variable dependiente y dos o más variables independientes. Algunas características de la regresión lineal múltiple incluyen:

Múltiples Variables Independientes: A diferencia de la regresión lineal simple que tiene una sola variable independiente, la regresión lineal múltiple involucra varias variables independientes que pueden afectar a la variable dependiente.

Modelo Lineal: El modelo asume que la relación entre las variables sigue una forma lineal, lo que significa que los cambios en las variables independientes están linealmente relacionados con los cambios en la variable dependiente.

Coeficientes de Regresión: La regresión lineal múltiple calcula coeficientes de regresión para cada variable independiente, que representan la magnitud del cambio en la variable dependiente asociada con un cambio unitario en la variable independiente, manteniendo las otras variables constantes.

Interpretación de Coeficientes: La interpretación de los coeficientes en la regresión lineal múltiple es importante, ya que permite entender cómo cada variable independiente contribuye al modelo y cómo interactúan entre sí.

Supuestos: Al igual que en la regresión lineal simple, la regresión lineal múltiple se basa en varios supuestos, incluyendo linealidad, independencia de errores, homocedasticidad (varianza constante de errores) y normalidad de errores.

Evaluación del Modelo: Se utilizan diversas métricas, como el coeficiente de determinación (R^2), el error estándar de la estimación y las pruebas de significancia de los coeficientes, para evaluar la bondad de ajuste y la validez del modelo.

La regresión lineal múltiple puede manejar variables categóricas utilizando técnicas como la codificación de variables dummy o la codificación de efectos contrastantes para representar la información de variables categóricas en el modelo.

En algunos casos, puede ser beneficioso aplicar transformaciones a las variables independientes o dependientes para cumplir con los supuestos de la regresión lineal, como transformaciones logarítmicas o de raíz cuadrada.

La selección de variables es un aspecto importante en la regresión lineal múltiple para evitar el sobreajuste del modelo. Técnicas como la selección hacia adelante, hacia atrás o bidireccional, así como métodos de regularización como la regresión de Ridge o Lasso, pueden ayudar en la selección de variables relevantes.

La multicolinealidad ocurre cuando dos o más variables independientes están altamente correlacionadas entre sí. Puede conducir a estimaciones inestables de los coeficientes de regresión y afectar la interpretación del modelo. Se deben tomar medidas para detectar y abordar la multicolinealidad si está presente.

La validación cruzada es una técnica importante para evaluar la capacidad predictiva del modelo y evitar el sobreajuste. Métodos como la validación cruzada k-fold o la validación cruzada leave-one-out pueden utilizarse para evaluar la capacidad de generalización del modelo.

El análisis de los residuos del modelo es fundamental para evaluar la validez de los supuestos de la regresión lineal múltiple. Los residuos deben ser independientes, homocedásticos y seguir una distribución normal. Se pueden realizar gráficos de residuos y pruebas estadísticas para verificar estos supuestos.

La regresión lineal múltiple se utiliza en una amplia variedad de campos, incluyendo ciencias sociales, ciencias naturales, negocios, ingeniería y medicina, para analizar y predecir relaciones entre variables.

1.1.1 pruebas de hipótesis en regresión lineal múltiple

Las pruebas de hipótesis se utilizan para evaluar la significancia de los coeficientes de regresión, es decir, para determinar si las variables independientes tienen un efecto significativo sobre la variable dependiente. Esto implica formular una hipótesis nula y una alternativa sobre el coeficiente de interés, y luego utilizar estadísticas como el valor t o el valor p para tomar decisiones sobre si rechazar o no la hipótesis nula. Estas pruebas ayudan a determinar qué variables independientes son estadísticamente significativas en la predicción de la variable dependiente en el modelo de regresión.

Las pruebas de hipótesis en regresión lineal múltiple son herramientas estadísticas utilizadas para evaluar la significancia de los coeficientes de regresión y determinar si existe una relación significativa entre las variables independientes y la variable dependiente en un modelo de regresión

Hipótesis nula y alternativa

En este contexto, la hipótesis nula (H0) generalmente afirma que no hay relación entre las variables independientes y la variable dependiente, mientras que la hipótesis alternativa (H1) sugiere lo contrario.

Prueba de significancia individual: Estas pruebas evalúan la significancia de cada coeficiente de regresión individualmente. Se utilizan pruebas t como la t de Student para determinar si un coeficiente es significativamente diferente de cero.

Prueba F global : Esta prueba evalúa la significancia conjunta de todos los coeficientes en el modelo de regresión. Se utiliza una prueba F para determinar si el modelo en su conjunto explica una cantidad significativa de la variabilidad en la variable dependiente.

Estadísticos de prueba y valores p: Los estadísticos de prueba (como las estadísticas t y F) se utilizan junto con los valores p asociados para tomar decisiones sobre la aceptación o rechazo de la hipótesis nula. Un valor p bajo sugiere que hay evidencia suficiente para rechazar la hipótesis nula.

Los intervalos de confianza: Además de las pruebas de hipótesis, los intervalos de confianza para los coeficientes de regresión también son importantes para evaluar la precisión de las estimaciones y la magnitud de los efectos.

Supuestos : Es crucial que se cumplan ciertos supuestos para que las pruebas de hipótesis en regresión lineal múltiple sean válidas, como la linealidad, la independencia de errores, la homocedasticidad y la normalidad de los errores.

1.1.2 intervalos de confianza y predicción en regresión múltiple

En el contexto de la regresión múltiple, los intervalos de confianza y de predicción son herramientas estadísticas importantes para evaluar la precisión de las estimaciones de los parámetros del modelo y para predecir valores futuros de la variable dependiente, respectivamente.

1. Intervalos de Confianza (IC)

- Representan un rango dentro del cual es probable que se encuentre el valor verdadero del parámetro de interés (por ejemplo, un coeficiente de regresión).

- Los intervalos de confianza típicamente se construyen alrededor de los estimadores de los parámetros, como el coeficiente de regresión, y proporcionan una medida de la incertidumbre asociada con la estimación del parámetro.

Un intervalo de confianza del 95%, por ejemplo, indica que si repitiéramos el experimento muchas veces y calculáramos un intervalo de confianza para cada experimento, aproximadamente el 95% de esos intervalos contendrían el valor verdadero del parámetro.

Construcción: Los intervalos de confianza se construyen utilizando la distribución t de Student y se basan en el error estándar del estimador.

Utilidad: Proporcionan una medida de la precisión de la estimación del parámetro y ayudan a determinar si un parámetro es significativamente diferente de cero.

2. Intervalos de Predicción (IP)

- Son utilizados para estimar el rango dentro del cual se espera que se encuentren los valores futuros de la variable dependiente para un conjunto dado de valores de las variables independientes.

- Los intervalos de predicción tienen en cuenta tanto la incertidumbre asociada con la estimación del parámetro como la variabilidad inherente de los datos.

- Por lo general, los intervalos de predicción son más amplios que los intervalos de confianza, ya que también tienen en cuenta la variabilidad adicional que puede surgir en los nuevos datos.

- Son especialmente útiles para evaluar la precisión de las predicciones del modelo y para tomar decisiones basadas en esas predicciones.

Estos intervalos son herramientas valiosas para interpretar y evaluar modelos de regresión múltiple, ya que proporcionan información sobre la precisión y la incertidumbre asociada con las estimaciones de los parámetros y las predicciones del modelo.

Diferencias entre IC e IP:

Ancho: Los intervalos de predicción son generalmente más amplios que los intervalos de confianza porque también tienen en cuenta la variabilidad adicional de los datos futuros.

Objetivo: Los intervalos de confianza se centran en la estimación de parámetros, mientras que los intervalos de predicción se centran en la incertidumbre asociada con las predicciones futuras.

Aplicación: Los intervalos de confianza se utilizan para evaluar la precisión de los parámetros del modelo, mientras que los intervalos de predicción se utilizan para evaluar la precisión de las predicciones del modelo.

1.2 Regresión no lineal.

La regresión no lineal es un método estadístico utilizado para modelar la relación entre una variable dependiente y una o más variables independientes cuando esa relación no es lineal. En lugar de ajustarse a una línea recta, como en la regresión lineal, la regresión no lineal se ajusta a una función no lineal.

Es importante elegir una función no lineal adecuada que se ajuste a los datos. Esto puede implicar probar varias funciones y evaluar su ajuste a través de métricas como el coeficiente de determinación (R^2) o la suma de los errores cuadrados (SSE).

En la regresión no lineal, los parámetros de la función no lineal deben inicializarse adecuadamente antes de ajustar el modelo a los datos. Esto puede afectar significativamente la convergencia y el rendimiento del modelo

A diferencia de la regresión lineal, donde se pueden obtener soluciones analíticas, la regresión no lineal a menudo requiere métodos iterativos para encontrar los mejores parámetros del modelo que minimicen la función de pérdida.

Interpretar los parámetros en un modelo de regresión no lineal puede ser más complicado que en un modelo de regresión lineal debido a la naturaleza no lineal de la función.

Además de las métricas mencionadas anteriormente es importante considerar otras medidas de ajuste del modelo, como el error cuadrático medio (RMSE) o el error absoluto medio (MAE), especialmente si los errores no están distribuidos normalmente.

Al igual que en la regresión lineal, la multicolinealidad puede ser un problema en la regresión no lineal. Esto ocurre cuando hay alta correlación entre las variables independientes, lo que puede dificultar la interpretación de los coeficientes del modelo.

Es crucial validar el modelo de regresión no lineal utilizando técnicas como la validación cruzada, para asegurarse de que el modelo generalice bien a nuevos datos y no esté sobre ajustado

En ocasiones, puede ser beneficioso aplicar transformaciones a las variables independientes o dependientes para lograr una relación más lineal antes de ajustar un modelo no lineal.

La interpretación de un modelo de regresión no lineal puede ser más desafiante que en la regresión lineal debido a la complejidad de la función no lineal. En algunos casos, técnicas como el análisis de sensibilidad o el análisis de efectos marginales pueden ayudar a comprender el impacto de las variables en la respuesta.

Referencias bibliográficas

https://www.bioestadistica.uma.es/apuntesMaster/regresi%C3%B3n-lineal-m%C3%BAltiple.html

https://cienciadedatos.net/documentos/25_regresion_lineal_multiple

https://www.studocu.com/es-mx/document/instituto-tecnologico-de-veracruz/seguridad-funcionamiento-y-riesgo-ambiental/1121/17379505

https://www3.uji.es/~gregori/docencia/mt1021-2122/mt1021-2122-la-4-reglin.html

INGENIERIA INDUSTRIAL
AEF - 1025

INSTITUTO TECNOLOGICO SUPERIOR DE SAN ANDRES TUXTLA

MATERIA: ESTADISTICA INFERENCIAL II

DOCENTE: PABLO PROMOTOR CAMPECHANO

INTEGRANTES:

MARIA FERNANDA ALEMAN GONZALEZ
KARLA LISSETTE FILIDOR DOMINGUEZ
JOSE GABRIEL FISCAL MEMECHI
JULIO CESAR HERNANDEZ DOMINGUEZ
IVAN JAIR MIXTEGA ANOTA

CARRERA: INGENIERIA INDUSTRIAL

GRUPO: 401-C   SEMESTRE: CUARTO SEMESTRE

UNIDAD 1: REGRESION LINEAL MULTIPLE

ACTIVIDAD: INVESTIGACION DOCUMENTAL

FECHA DE ENTREGA: 23/02/2024

SAN ANDRES TUXTLA, VER

Problema 1.

La siguiente tabla muestra los pesos Y, las estaturas X_1 y las edades X_2 de 12 muchachos. El peso se mide en libras y la estatura en pulgadas, la edad en años.

Peso Y	Estatura X_1	Edad X_2
64	57	8
71	59	10
53	49	6
67	62	11
55	51	8
58	50	7
77	55	10
56	48	9
56	52	10
51	42	6
76	61	12
68	57	9

$N = 12$

$\sum x_1 = 57+59+49+62+51+50+55+48+52+42+61+57 = 629$

$\sum x_2 = 8+10+6+11+8+7+10+9+10+6+12+9 = 94$

$\sum Y = 64+71+53+67+55+58+77+56+56+51+76+68 = 719$

$\sum (x_1^2) = 57^2+59^2+49^2+62^2+51^2+50^2+55^2+48^2+52^2+42^2+61^2+57^2 = 37814$

$\sum (x_2^2) = 8^2+10^2+\dots\dots+9^2 = 870$

$\sum (x_1 x_2) = (57\times8)+(59\times10)+(49\times6)+(62\times11)+(51\times8)+(50\times7)+(55\times10)+(48\times9)+(52\times10)+(42\times6)+(61\times12)+(57\times9) = 5470$

$\sum(X_1 Y) = (57 \times 64) + (59 \times 71) + \cdots + (57 \times 68) = 41193$

$\sum(X_2 Y) = (8 \times 64) + (10 \times 71) + \cdots + (9 \times 68) = 5404$

$$\bar{X}_1 = \frac{\sum X_1}{N} = \frac{629}{12} = 52.42$$

$$\bar{X}_2 = \frac{\sum X_2}{N} = \frac{94}{12} = 7.83$$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{719}{12} = 59.92$$

$$b_1 = \frac{(12 \times 41193) - (629 \times 719)}{(12 \times 37811) - (629^2)} = 1.5217$$

$$b_2 = \frac{(12 \times 5404) - (94 \times 719)}{(12 \times 870) - (94^2)} = -0.4281$$

$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$

$b_0 = 59.92 - 1.5217 \times 52.42 - (-0.4281) \times 7.83$

$b_0 = 3.4292$

Por lo tanto, la ecuación de regresión lineal múltiple es:

$Y = 3.4292 + 1.5217X_1 - 0.4281X_2$

$Y = 3.4292 + 1.5217(58) - 0.4281(11)$

$Y = 92.69$

$$SE = \sqrt{\frac{\sum(Y - \hat{Y})^2}{N-3}}$$

$SE = 3.99$

$$r = \frac{\sum(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{\sqrt{\sum(X_1 - \bar{X}_1)^2 \sum(X_2 - \bar{X}_2)^2}}$$

$r = 0.9017$

$R^2 = r^2$

$R^2 = 0.8134$

Problema 3

Una empresa de desarrollo de software establece relacionar sus ventas en función del número de pedidos de los tipos de software que desarrolló, para atender 10 proyectos en el presente año.

En la siguiente tabla Y representa las ventas en miles de pesos, X es el número de pedidos del software de sistemas y W representa el número de pedidos del software de aplicaciones educativas. Determine la ecuación de regresión múltiple

Y	440	455	470	510	506	480	460	500	490	450
X	50	40	35	45	51	55	53	48	38	44
W	105	140	110	130	125	115	100	103	118	98

Determine el error estandar, el coeficiente de correlación y el coeficiente de determinación

Ecuación de regresión múltiple

$$\bar{Y} = a + b_1 X_1 + b_2 X_2$$

Pedidos	Y	X_1	X_2	$(X_1)^2$	$(X_2)^2$	$(X_1)(X_2)$	$(X_1)Y$	$(X_2)Y$
1	440	50	105	2500	11025	5250	22000	46200
2	455	40	140	1600	19600	5600	18200	63700
3	470	35	110	1225	12100	3850	16450	51700
4	510	45	130	2025	16900	5850	22950	66300
5	506	51	125	2601	15625	6375	25806	63250
6	480	55	115	3025	13,225	6325	26400	55200
7	460	53	100	2809	16,000	5300	24380	46000
8	500	48	103	2304	10,609	4944	24000	51500
9	490	38	118	1444	13,924	4484	18620	57820
10	450	44	98	1936	9604	4312	19800	44100

$\Sigma_Y = 4,761$ $\Sigma_{X_1} = 459$ $\Sigma_{X_2} = 1144$ $\Sigma = 138,612$ $\Sigma = 218,606$

$\Sigma_{X^2} = 20,929$ $\Sigma = 52290$ $\Sigma = 595770$

$$\sum y = na + b_1 \sum x_1 + b_2 \sum x_2$$

$$\sum x_1 y = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2$$

$$\sum x_2 y = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2$$

$4761 = 10a + b_1(459) + b_2(1144)$ Ec 1

$218606 = a(459) + b_1(29{,}929) + b_2(52{,}290)$ Ec 2

$545{,}270 = a(1144) + b_1(52290) + b_2(138{,}612)$ Ec 3

Se ordena el sistema

$4761 = 10a + 459b_1 + 1144b_2$ Ec 1
$218606 = 459a + 29929b_1 + 52290b_2$ Ec 2
$545{,}270 = 1144a + 52290b_1 + 138612b_2$ Ec 3

Resolviendo por el sistema de Cramer

$a = 498.66$
$b_1 = -0.643$
$b_2 = 0.060$

Sustituyendo estos valores en la ecuación:

$\bar{y} = a + b_1x_1 + b_2x_2$

$\bar{y} = 498.66 - 0.643 + 0.060$

Error estandar

$$S_e = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-3}}$$

$$S_e = \sqrt{\frac{SCE}{n-3}}$$

$$S_e = \sqrt{\frac{6{,}078.5}{10-3}}$$

$$S_e = \sqrt{\frac{6{,}078.5}{7}}$$

$$S_e = 11.13$$

Pedido	$Y = Y_i$	X_1	X_2	$\hat{Y}_i = \bar{Y} = 498.66 - 0.643X_1 + 0.060X_2$	$(Y_i - Y_i)^2$	$(\hat{Y} - \bar{Y})^2$	$(Y_i - \bar{Y})^2$
1	440	50	105	472.81	1076.4	103.8	1849
2	455	40	140	481.34	693.7	2.75	784
3	470	35	110	482.755	162.5	0.060	169
4	510	45	130	547.725	1421	4189	729
5	506	51	125	473.33	1069.2	93.50	529
6	480	55	115	470.165	96.82	164.73	9
7	460	53	100	470.55	110.25	155.00	529
8	500	48	103	473.94	679.12	82.08	289
9	490	38	118	481.27	81	2.99	49
10	450	44	98	476.24	688.5	45.69	1089
				483,0.1	SCE	SCR	SCT
					6,078.5	4842.6	60343

Problema 4

En la facultad de ingeniería de sistemas se quiere entender los factores de aprendizaje de los alumnos que cursan la asignatura de PHP, para lo cual se escoge al azar una muestra de 15 alumnos y ellos registran notas promedias en las asignaturas de algoritmos y base de datos, como se muestra en el siguiente cuadro.

1

Alumno	PHP	Algoritmos	Base de Datos
1	13	15	15
2	13	14	13
3	13	16	13
4	15	20	14
5	16	18	18
6	15	16	17
7	12	13	15
8	13	16	14
9	13	15	14
10	13	14	13
11	11	12	12
12	14	16	11
13	15	17	16
14	15	19	14
15	15	13	15

Determine la ecuación de regresión lineal múltiple.

$$Y = 2.551_1 + 0.583_2 + 0.373_3 + -0.242_4$$

Determine el error estándar, el coeficiente de correlación y el coeficiente de determinación.

$$Y = a + b_1x_1 + b_2x_2 + b_3x_3$$

	Coeficiente	Error Típico	Estadístico T	P-Valor	Inferior	Superior
Intercepción	2.551	2.369	1.077	0.305	-2.663	7.766
Algoritmos	0.583	0.267	2.186	0.051	-0.004	1.169
Bases de datos	0.373	0.144	2.589	0.025	0.056	0.691
Programación	-0.242	0.27	-0.893	0.391	-0.837	0.354

Alumno	PHP	Algoritmos	Base de datos	$(x1)^2$	$(x2)^2$	$(x1)^{\square}(x2)^{\square}$	$(x1)^y$	$(x2)^y$
1	13	15	15	225	225	50625	195	195
2	13	14	13	196	169	33124	182	169
3	13	16	13	256	169	43264	208	169
4	15	20	14	400	196	78400	300	210
5	16	18	18	324	324	104976	288	288
6	15	16	17	256	289	73984	240	255
7	12	13	15	169	225	38025	156	180
8	13	16	14	256	196	50176	208	182
9	13	15	14	225	196	44100	195	182
10	13	14	13	196	169	33124	182	169
11	11	12	12	144	144	20736	132	132
12	14	16	11	256	121	30976	224	154
13	15	17	16	289	256	73984	255	240
14	15	19	14	361	196	70756	285	210
15	15	13	15	169	225	38025	195	225
Total de	206	234	214	3722	3100	784275	3245	2960

$$r^2 = \frac{18.7737874}{26.9333333} = 0.69704656$$

Estadística	
Coeficiente de correlación múltiple	0.83489314
Coeficiente de determinación R'2	0.69704656
R'2 ajustado	0.6144229
Error típico	0.86126471
Observaciones	15

Problema 5

Se piensa que la energía eléctrica consumida mensualmente por una industria se relaciona con la temperatura promedio X_1 en °F, y el número de días laborales del mes X_2. Se cuenta con los datos del último año, los cuales se presentan en la tabla siguiente.

y	x_1	x_2
240	25	24
236	31	21
290	45	24
274	60	25
301	65	25
316	72	26
300	80	25
296	84	25
267	75	24
276	60	25
288	50	25
261	38	23

a) Determine la ecuación de regresión lineal múltiple.

b) Estime el consumo de electricidad para un mes en el que la temperatura promedio es de 74 °F y el número de días laborales es de 24.

c) Determine el error estándar, el coeficiente de correlación y el coeficiente de determinación.

y	x1	x2	$(x1)^2$	$(x2)^2$	$(x1)(x2)$	$(x1)y$	$(x2)y$
240	25	24	625	576	360000	6000	5760
236	31	21	961	441	423801	7316	4956
290	45	21	2025	441	893025	13050	6090
274	60	25	3600	625	2250000	16440	6850
301	65	25	4225	625	2640625	19565	7525
316	72	26	5184	676	3504384	22752	8216
300	80	26	6400	676	4326400	24000	7800
296	84	25	7056	625	4410000	24864	7400
267	75	24	5625	576	3240000	20025	6408
276	60	25	3600	625	2250000	16560	6900
288	50	23	2500	529	1322500	14400	6624
261	38	23	1444	529	763876	9918	6003
Total de			43245	6944	26384611	194890	80532

$b_{0}=-102{,}713$, $b_{1}=0{,}605$, $\beta_{2}=8.923$

$\hat{y}=-102{,}713+0{,}605X_1+8{,}923X_2$

$R_2=0{,}5988$

Problema 2

Los siguientes datos describen el experimento para determinar la influencia de:

X_1: Temperatura del agua

X_2: Cantidad de blanqueador de cloro

Sobre una medida adecuada de blancura del rayón (y).

Temperatura del agua X_1	Cantidad del blanqueador de cloro X_2	Blancura del rayón Y
82	0.3	76.5
82	0.5	76
88	0.5	79.9
88	0.3	83.5
82	0.5	89.5
82	0.3	54.2
88	0.3	85.7
88	0.5	99.5
82	0.5	89.4
82	0.3	97.5
88	0.3	103.2
88	0.5	108.7

$\sum x_1 = 82+82+88+88+82+82+88+88+82+82+88+88 = 1020$

$\sum x_2 = 0.3+0.5+0.5+0.3+0.5+\cdots+0.3+0.5 = 5.2$

$\sum Y = 76.5+76+\cdots+103.2+108.7 = 1027.2$

$\sum (x_1^2) = 82^2+82^2+88^2+88^2+82^2+82^2+88^2+88^2+82^2+82^2+88^2+88^2 = 13168$

$\sum (x_2^2) = 0.3^2+0.5^2+\cdots+0.3^2+0.5^2 = 1.66$

$\sum (x_1 x_2) = (82\times0.3)(82.0.5)+\cdots+(88\times0.3)(88\times0.5) = 671.2$

$\sum (x_1 Y) = (82\times76.5)(82\times76)+\cdots(88\times103.2)+(88\times108.7) = 83366.7$

$\sum (x_2 Y) = (0.3\times76.5)+(0.5\times76)+\cdots+(0.3\times103.2)+(0.5\times108.7) = 5383.3$

$$\bar{X}_1 = \frac{\sum x_1}{N} = \frac{1008}{12} = 84$$

$$\bar{X}_2 = \frac{\sum X_2}{N} = \frac{5.2}{12} = 0.4333$$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{1027.2}{12} = 85.6$$

$$b_1 = \frac{(12 \times 83366.1) - (1008 \times 1027.2)}{(12 \times 131664) - (1008)^2} = 0.5340$$

$$b_2 = \frac{(12 \times 5383.3) - (5.2 \times 1027.2)}{(12 \times 1.66) - (5.29} = -8322.85$$

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

$$b_0 = 85.6 - (-0.5340) \times 84 - (-8322.85) \times 0.4333$$

$$b_0 = 135.2662$$

Por lo tanto la ecuación de regresión estimada es:

$$Y = 135.2662 - 0.5340 X_1 - 8322.85 X_2$$

$$Y = 135.2662 - 0.5340(87) - 8322.85(0.4)$$

$$\underline{Y = 90.49}$$