

Probabilidad y Estadística

Unidad 3. Regresión

Periodo: Ago – Dic 2021

Profesor: M. en C. Paul Ramírez De la Cruz

Nombre: _____

**Tarea 7.2. Análisis de
Regresión Lineal Múltiple**

Puntaje total: 10 puntos

Fecha de entrega:

Calificación: _____

1. Instrucciones. Resuelva los siguientes ejercicios de regresión múltiple haciendo uso de R

1. Se investigó el efecto del peso del vehículo (en toneladas) y la temperatura ambiente promedio (en °F) sobre el rendimiento del combustible, en una muestra aleatoria de 10 vehículos. Los datos se muestran a continuación:

Peso (X_1)	Temperatura (X_2)	Rendimiento (Y)
1.35	90	17.9
1.9	30	16.5
1.7	80	16.4
1.8	40	16.8
1.3	35	18.8
2.05	45	15.5
1.6	50	17.5
1.8	60	16.4
1.85	65	15.9
1.4	30	18.3

- a. Obtenga una ecuación de mínimos cuadrados para el modelo de regresión lineal que explique el Rendimiento de combustible del vehículo con base en su Peso y la Temperatura ambiente promedio.
 - b. ¿Hay pruebas suficientes que indiquen que el modelo resulta útil para predecir y ? Pruebe con $\alpha = 0.10$ usando el contraste de hipótesis del Análisis de Varianza.
 - c. Calcule un intervalo de predicción de 95% para el rendimiento observado en un automóvil con un peso de 1.35 ton si la temperatura ambiente es de 40° F. Interprete el intervalo
2. [Lind et al (2004). Estadística para administración y economía. Ejercicio 14.19, p. 535] El señor Steve Douglas fue contratado como gerente en capacitación por una importante empresa financiera. Como primer proyecto, se le pidió que estudiara la utilidad bruta en la industria química. ¿Qué factores afectan las utilidades en esa industria? Steve seleccionó una muestra aleatoria de 16

compañías y obtuvo datos respecto a la cantidad de empleados (EMPLEADOS), el número de dividendos consecutivos pagados en acciones comunes (DIVIDENDOS), el valor total del inventario al inicio del presente año (INVENTARIO) y la ganancia bruta (GANANCIA) de cada empresa.

Ganancia	Empleados	Dividendos	Inventario
2800	140	12	1800
1300	65	21	320
1230	130	42	820
1600	115	80	76
4500	390	120	3600
5700	670	64	8400
3150	205	43	508
640	40	14	870
3400	480	88	5500
6700	810	98	9875
3700	120	44	6500
6440	590	110	9130
1280	440	38	1200
4160	280	24	890
3870	650	60	1200
980	150	24	1300

- Ajuste un modelo de regresión lineal múltiple a los datos y estime los coeficientes de regresión utilizando todas las variables
 - Realice el contraste de hipótesis del Análisis de Varianza y diga si al menos una de las variables utilizadas es adecuada para pronosticar el valor de la Ganancia
 - Identifique las variables que son útiles para pronosticar Y para ajustar un modelo corregido. Proporcione este nuevo modelo.
3. [Ejercicios 12.1. 43 y ss.] Se estudió el agrietamiento de la pintura de látex en estructura de madera. La preocupación principal del estudio es investigar el efecto de la permeabilidad al agua (X_1) y la energía de fractura (X_2 , la energía necesaria para que se propague una grieta en la película de pintura) en la tasa de agrietamiento de la pintura (Y). La investigación arrojó los datos siguientes

Agrietamiento	Permeabilidad	Energía de Fractura
2	2.1	4.31
9	8.4	22.11
5	5.1	11.4
10	14.5	24.15
3	4.4	6.21
3	6.2	5.65
8	12.5	9.71
7	7	12
8	17.2	14.25
5	7.1	8.63

- Elabore una matriz de dispersión. ¿Parece existir una relación lineal entre las variables?

- b. Estime puntualmente los coeficientes del modelo de regresión lineal múltiple, considerando todas las variables explicativas e interprételos en el contexto del problema.
 - c. Proporcione una estimación por intervalo del 90% de confianza para β_0 . Interprete el resultado
 - d. Proporcione una estimación por intervalo del 90% de confianza para beta β_1 . Interprete el resultado
 - e. Calcule el valor de r y r^2 e interprete ambos resultados. ¿Corresponden a su apreciación en el inciso a)?
 - f. Estime puntualmente el valor del agrietamiento si se tiene una permeabilidad de 5.5 y una energía de fractura de 10.2
 - g. Estime con un intervalo de confianza de 95% el número medio
 - h. Estime con un intervalo de confianza de 95% el número observado de operaciones de E/S (en miles) que se tendría si se procesa un archivo de 444 (miles de) registros
4. [Problema 14.18 p. 534 Lind et al.] Las tiendas de Fran's Convenience Marts están localizadas en el área metropolitana de Erie, Pennsylvania (EUA). A Fran, la dueña, le agradaría expandirse a otras comunidades del noroeste de Pennsylvania y el suroeste de Nueva York, tales como Jamestown, Corry, Meadville y Warren. Como parte de su presentación al banco local, le gustaría entender mejor los factores que hacen que una tienda particular sea lucrativa. La propietaria debe hacer todo el trabajo sola, así que no podrá analizar todos sus establecimientos. Selecciona una muestra aleatoria de 15 tiendas y de cada una registra el promedio de las ventas diarias (Y), la superficie (área), el número de cajones de estacionamiento, y el ingreso económico medio de familias en esa región

Venta Diaria	Área de la Tienda	Cajones	Ingresos
1840	532	6	44
1746	478	4	51
1812	530	7	45
1806	508	7	46
1792	514	5	44
1825	556	6	46
1811	541	4	49
1803	513	6	52
1830	532	5	46
1827	537	5	46
1764	499	3	48
1825	510	8	47
1763	490	4	48
1846	516	8	45
1815	482	7	43

- a. Obtenga el mejor modelo de regresión lineal múltiple utilizando el método de selección de variables *paso a paso*.
- b. Realice un análisis completo de residuos para el modelo obtenido.

Referencias

1. **Lind, Marchal & Mason.** *Estadística Para Administración Y Economía*. Alfaomega. México, 2004.
2. **Mendenhall, William & Sincich, Terry.** *Probabilidad y estadística para ingeniería y ciencias*. Prentice-Hall Hispanoamericana. México, 1995
3. **Milton, J. Susan & Arnold, Jesse C.** *Probabilidad y estadística con aplicaciones para ingeniería y ciencias computacionales*. McGraw-Hill. México, 2004