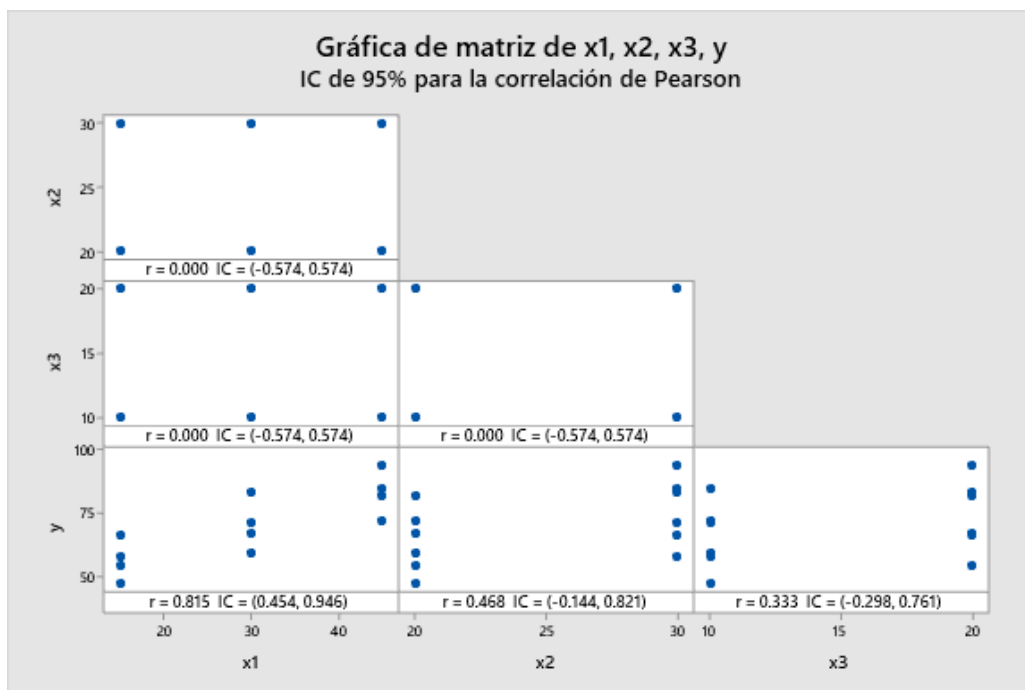


NOMBRES: Valeria Guadalupe García Salazar y Jennifer Priscila de León Flores

Considere los datos Analgésicos correspondientes al porcentaje de efectividad de un analgésico y las cantidades de tres diferentes medicamentos (en miligramos) presentes en cada cápsula. Use un alfa de 0.10

medicamento A	medicamento B	medicamento C	eficacia (%)
x1	x2	x3	y
15	20	10	47
15	20	20	54
15	30	10	58
15	30	20	66
30	20	10	59
30	20	20	67
30	30	10	71
30	30	20	83
45	20	10	72
45	20	20	82
45	30	10	85
45	30	20	94

- A)** Realice la matriz de dispersión y comente respecto a la posible relación entre par de variables



Las gráficas presentan dispersión en las variables x1, x2 y x3 donde se observa que algunas son equidistantes a excepción de x1 vs y ya que se presentan unos puntos mas juntos, pero de igual manera equidistantes unos con otros.

B) Revise problemas de multicolinealidad

Coefficientes

Término	Coef	EE del coef.	IC de 95%	Valor T	Valor p	FIV
Constante	-2.33	2.20	(-7.41, 2.74)	-1.06	0.320	
x1	0.9000	0.0281	(0.8353, 0.9647)	32.08	0.000	1.00
x2	1.2667	0.0687	(1.1082, 1.4251)	18.43	0.000	1.00
x3	0.9000	0.0687	(0.7415, 1.0585)	13.10	0.000	1.00

Se puede observar en la tabla que los FIV's son menores a 10, por lo que no existen problemas de multicolinealidad.

C) Ajustar la regresión lineal e indicar la ecuación correspondiente.

De la tabla del inciso anterior tenemos los coeficientes:

Coefficientes

Término	Coef	EE del coef.	IC de 95%	Valor T	Valor p	FIV
Constante	-2.33	2.20	(-7.41, 2.74)	-1.06	0.320	
x1	0.9000	0.0281	(0.8353, 0.9647)	32.08	0.000	1.00
x2	1.2667	0.0687	(1.1082, 1.4251)	18.43	0.000	1.00
x3	0.9000	0.0687	(0.7415, 1.0585)	13.10	0.000	1.00

Por lo que la ecuación nos queda de la siguiente manera:

Ecuación de regresión

$$y = -2.33 + 0.9000 x_1 + 1.2667 x_2 + 0.9000 x_3$$

D) Probar la significancia del modelo e indicar el ajuste obtenido

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	PRESS	R-cuad. (pred)	AICc	BIC
1.19024	99.48%	99.29%	23.6247	98.92%	53.37	45.79

Como nos indica la columna R-cuad, el modelo tiene un ajuste del 99.48%

Análisis de Varianza

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	3	2182.33	99.48%	2182.33	727.44	513.49	0.000
x1	1	1458.00	66.46%	1458.00	1458.00	1029.18	0.000
x2	1	481.33	21.94%	481.33	481.33	339.76	0.000
x3	1	243.00	11.08%	243.00	243.00	171.53	0.000
Error	8	11.33	0.52%	11.33	1.42		
Total	11	2193.67	100.00%				

Prueba de significancia del modelo:

H0: B1=B2=B3=0 vs **H1:** Bi dif 0 para i=1,2,3

Rechazo H0 si p valor=0.00< alfa=0.10

Rechazo H0

Las variables se relacionan de manera significativa.

E) Analizar mejores subconjuntos e indicar el modelo elegido

REGRESION DE LOS MEJORES SUB CONJUNTOS y VS. X1, x2, x3:

la respuesta es y

Vars	R-cuad.	R-cuad. (ajust)	R-cuad. (pred.)	Cp de Mallows	x 1	x 2	x 3
1	66.5	63.1	52.0	511.3	8.5771	X	
1	21.9	14.1	0.0	1200.7	13.086	X	
2	88.4	85.8	79.6	173.5	5.3159	X	X
2	77.5	72.6	60.3	341.8	7.3987	X	X
3	99.5	99.3	98.9	4.0	1.1902	X	X

Ecuación de regresión

$$y = 0.970 x_1 + 1.616 x_2$$

El modelo elegido fue el de la variable 2 ya que es el más alto dejando fuera el de la variable 3.

F) Calcular los intervalos de confianza e indicar si existe evidencia de regresión al origen

Coeficientes

Término	Coef	EE del coef.	IC de 90%	Valor T	Valor p	FIV
Constante	11.17	8.68	(-4.75, 27.08)	1.29	0.230	
x1	0.900	0.125	(0.670, 1.130)	7.18	0.000	1.00
x2	1.267	0.307	(0.704, 1.829)	4.13	0.003	1.00

Existe evidencia estadística que afirma que hay regresión al origen, es decir, el intervalo incluye al 0.

G) En caso de existir, volver a ajustar el modelo y realizar la prueba e indicar el ajuste

Ecuación de regresión

$$y = 0.970 x_1 + 1.616 x_2$$

Coefficientes

Término	Coef	EE del coef.	IC de 90%	Valor T	Valor p	FIV
x1	0.970	0.117	(0.758, 1.181)	8.32	0.000	5.69
x2	1.616	0.148	(1.347, 1.884)	10.90	0.000	5.69

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	PRESS	R-cuad. (pred)	AICc	BIC
5.48720	99.50%	99.40%	440.675	99.27%	81.72	80.18

El modelo tiene un ajuste del 99.50%, es muy bueno para estos datos.

Análisis de Varianza

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	2	60412.9	99.50%	60412.9	30206.5	1003.22	0.000
x1	1	56833.1	93.61%	2083.6	2083.6	69.20	0.000
x2	1	3579.8	5.90%	3579.8	3579.8	118.89	0.000
Error	10	301.1	0.50%	301.1	30.1		
Falta de ajuste	4	50.1	0.08%	50.1	12.5	0.30	0.868
Error puro	6	251.0	0.41%	251.0	41.8		
Total	12	60714.0	100.00%				

Prueba de significancia del modelo:

H0: B1=B2=0 vs **H1:** Bi dif 0 para i=1,2

Rechazo H0 si p valor=0.00 < alfa=0.10

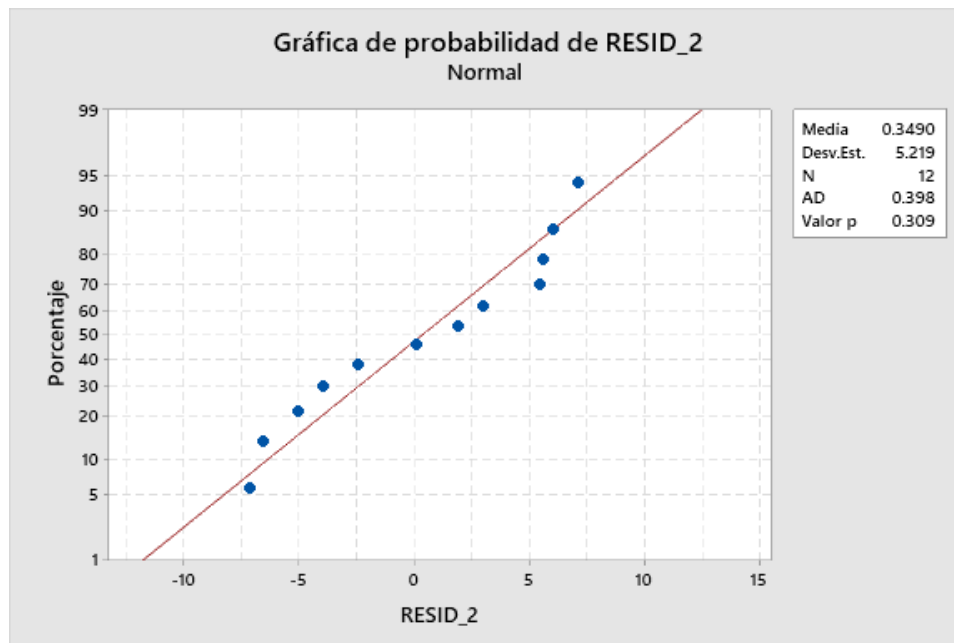
Rechazo H0

Las variables se relacionan de manera significativa.

H) Analice el cumplimiento de supuestos de los residuales incluyendo gráficas y comentarios al respecto



Supuestos:



Prueba de normalidad

Rechazo H_0 si $p \text{ valor} < \alpha$

$0.309 < 0.10$

NO RECHAZO H_0

Los residuales siguen una distribución normal

Media cero

Media: 0.3490

Podemos decir que la media es relativamente pequeña.

Varianza constante

Para la gráfica de vs ajustes por lo que consideramos una varianza constante

Incorrelación

H_0 : Residuales incorrelacionados

H_1 : Residuales correlacionados

RESID_1	Resid^2	Restas	Restas^2
0.140625	0.01977539		
7.140625	50.9885254	-7	49
-5.015625	25.1564941	12.15625	147.774414
2.984375	8.90649414	-8	64
-2.40625	5.79003906	5.390625	29.0588379
5.59375	31.2900391	-8	64
-6.5625	43.0664063	12.15625	147.774414
5.4375	29.5664063	-12	144
-3.953125	15.6271973	9.390625	88.1838379
6.046875	36.5646973	-10	100
-7.109375	50.5432129	13.15625	173.086914
1.890625	3.57446289	-9	81
	301.09375		1087.87842

$EP = 1087.8784 / 301.09375 = 3.6130$

Datos: $n=12$ $\alpha=0.05$

n	dL	dU
12		
15	0.95	1.57

Región de rechazo:

Rechazo H_0 si $3.61 < dL$

No Rechazo H_0 si $3.61 > dU$

No se rechaza H_0

Con un nivel de significancia del 5% existe evidencia estadística que afirma que los residuales están incorrelacionados.

Se cumplen todos los supuestos.

- I) Calcule los residuales y residuales estandarizados del modelo elegido e indique la presencia de datos atípicos.

RESID_2	RESIDEST_2
0.140625	0.02672944
7.140625	1.35726177
-5.015625	-1.08407287
2.984375	0.64504024
-2.40625	-0.45543929
5.59375	1.05874847
-6.5625	-1.27854029
5.4375	1.05936196
-3.953125	-0.84217458
6.046875	1.28822752
-7.109375	-1.41707765
1.890625	0.37684922

No hay presencia de datos atípicos ya que los valores de los residuales estandarizados se encuentran en el intervalo de $(-3,3)$