

REGRESION LINEAL MULTIPLE



- DR. PORFIRIO GUTIERREZ GONZALEZ
- CORREO ELECTRONICO: pgutierrezglez@gmail.com
- PAGINA WEB: www.seraace.com

Regresión Lineal Múltiple

En muchos problemas existen dos o más variables que están relacionadas y puede ser importante modelar y explorar esta relación. Por ejemplo, **el rendimiento de una reacción química** puede depender de la **temperatura, presión y concentración** del catalizador. En este caso se requiere al menos un modelo de regresión con tres variables.

El problema general consiste en ajustar el modelo de primer orden

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

O en ajustar el modelo de segundo orden

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

HIPOTESIS DE LOS PARAMETROS DE LA REGRESION

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_{11} = \beta_{22} = \beta_{33} = \beta_{12} = \beta_{13} = \beta_{23} = 0$$

H_a : al menos una $\beta_{ij} \neq 0$

ESTADISTICO DE PRUEBA T -STUDENT

HIPOTESIS DEL MODELO DE REGRESION

H₀: EL MODELO NO ES SIGNIFICATIVO

H_a: EL MODELO ES SIGNIFICATIVO

Estadístico de prueba Tabla de Anova (F-Fisher)

AJUSTE DEL MODELO – R-CUADRADO

R-CUADRADO > 70%

Métodos de obtención del modelo

- ☐ Método de todas las variables
- ☐ Método Forward (Hacia adelante)
- ☐ Método Backward (Hacia atrás)

El mejor modelo debe contener

- ☐ Los parámetros del modelo son significativos
- ☐ El ANOVA del modelo es significativo
- ☐ Se tiene un R-cuadrado aceptable

Ejemplo Regresión Múltiple

Un ingeniero químico se encuentra investigando el rendimiento de un proceso, del cual le interesan tres variables: temperatura, presión y concentración porcentual. Cada variable puede estudiarse a dos niveles, bajo y alto, y el ingeniero decide correr un diseño 2^3 con estas tres variables. El experimento y los rendimientos resultantes se muestran en la siguiente tabla,

X1=TEMPERATURA	X2=PRESION	X3=CONCENTRACION	Y=RENDIMIENTO
50	100	10	32
50	100	20	36
50	200	10	57
100	100	10	46
100	200	10	65
50	200	20	57
100	100	20	48
100	200	20	68

METODO DE TODAS LAS VARIABLES

Ecuación de regresión

$Y(\text{RENDIMIENTO}) = -11.1 + 0.315 X_1(\text{TEMPERATURA}) + 0.2875 X_2(\text{PRESION}) + 0.375 X_3(\text{CONCENTRACION}) - 0.000700 X_1(\text{TEMPERATURA}) * X_2(\text{PRESION}) + 0.00100 X_1(\text{TEMPERATURA}) * X_3(\text{CONCENTRACION}) - 0.00150 X_2(\text{PRESION}) * X_3(\text{CONCENTRACION})$

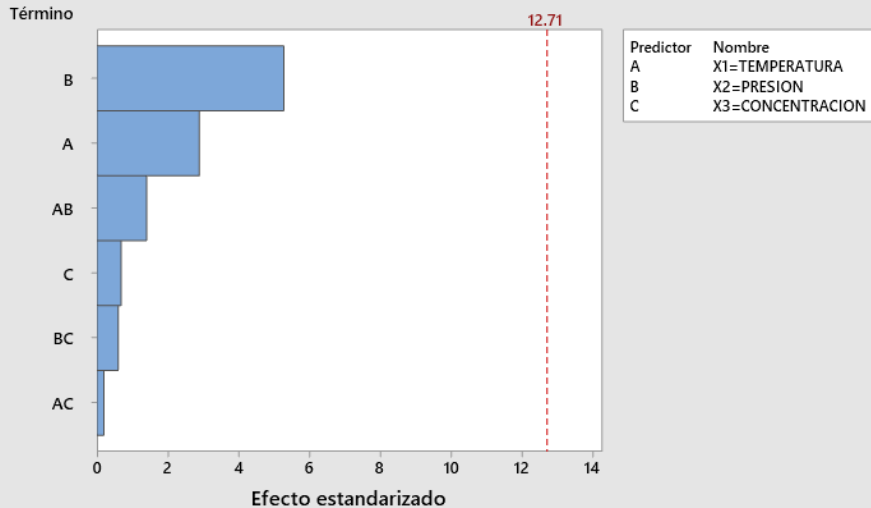
Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p
Constante	-11.1	10.3	-1.08	0.475
X1(TEMPERATURA)	0.315	0.109	2.89	0.212
X2(PRESION)	0.2875	0.0545	5.28	0.119
X3(CONCENTRACION)	0.375	0.545	0.69	0.616
X1(TEMPERATURA)*X2(PRESION)	-0.000700	0.000500	-1.40	0.395
X1(TEMPERATURA)*X3(CONCENTRACION)	0.00100	0.00500	0.20	0.874
X2(PRESION)*X3(CONCENTRACION)	-0.00150	0.00250	-0.60	0.656

Ningún termino es significativo

METODO DE TODAS LAS VARIABLES

Diagrama de Pareto de efectos estandarizados
(la respuesta es Y=RENDIMIENTO, $\alpha = 0.05$)



Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	6	1173.75	195.625	62.60	0.096
X1=TEMPERATURA	1	26.11	26.112	8.36	0.212
X2=PRESION	1	87.01	87.007	27.84	0.119
X3=CONCENTRACION	1	1.48	1.480	0.47	0.616
X1=TEMPERATURA*X2=PRESION	1	6.12	6.125	1.96	0.395
X1=TEMPERATURA*X3=CONCENTRACION	1	0.12	0.125	0.04	0.874
X2=PRESION*X3=CONCENTRACION	1	1.12	1.125	0.36	0.656
Error	1	3.12	3.125		
Total	7	1176.88			

Resumen del modelo

	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
	1.76777	99.73%	98.14%
			83.01%

El modelo no es significativo. El R-cuadrado de 99.73 esta subestimado.

METODO FORWARD (HACIA ADELANTE)

Ecuación de regresión

$$Y(\text{RENDIMIENTO}) = -8.87 + 0.3300 X1(\text{TEMPERATURA}) + 0.2650 X2(\text{PRESION}) \\ + 0.2250 X3(\text{CONCENTRACION}) - 0.000700 X1(\text{TEMPERATURA}) * X2(\text{PRESION})$$

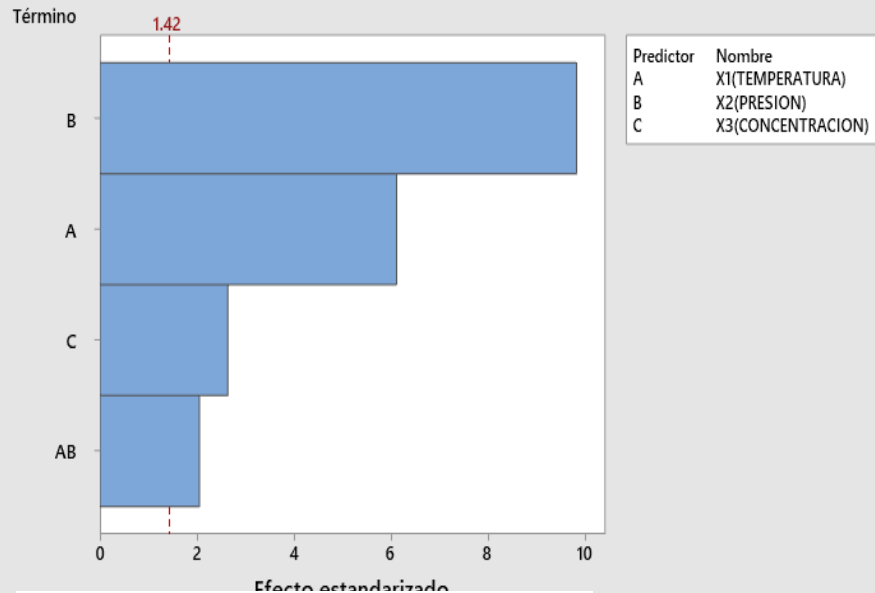
Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	FIV
Constante	-8.87	4.46	-1.99	0.141	
X1(TEMPERATURA)	0.3300	0.0540	6.11	0.009	10.00
X2(PRESION)	0.2650	0.0270	9.81	0.002	10.00
X3(CONCENTRACION)	0.2250	0.0854	2.63	0.078	1.00
X1(TEMPERATURA)*X2(PRESION)	-0.000700	0.000342	-2.05	0.133	19.00

El termino constante no es significativo. Los parámetros X1, X2, X3, y X1*X2, son significativos para un alfa de 0.05.

METODO FORWARD (HACIA ADELANTE)

Diagrama de Pareto de efectos estandarizados
(la respuesta es Y(RENDIMIENTO), $\alpha = 0.25$)



Resumen del modelo

	R-cuad.	R-cuad.
	S R-cuad. (ajustado)	(pred)
	1.20761	99.63%
	99.13%	97.36%

Análisis de Varianza

Fuente	GL	SC	Ajust.	MC	Ajust.	Valor F	Valor p
Regresión	4	1172.50		293.125		201.00	0.001
X1(TEMPERATURA)	1	54.45		54.450		37.34	0.009
X2(PRESION)	1	140.45		140.450		96.31	0.002
X3(CONCENTRACION)	1	10.13		10.125		6.94	0.078
X1(TEMPERATURA)*X2(PRESION)	1	6.12		6.125		4.20	0.133
Error	3	4.37		1.458			
Total	7	1176.88					

El modelo es significativo para un alfa de 0.05. El R-cuadrado de 99.63, el cual es aceptable.

METODO BACKWARD (HACIA TRAS)

Ecuación de regresión

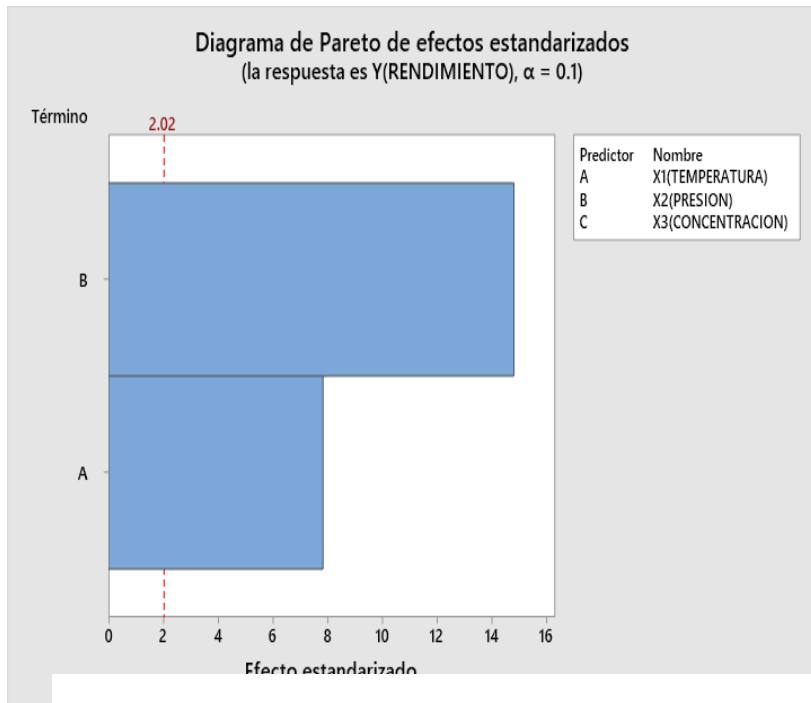
$$Y(\text{RENDIMIENTO}) = 2.38 + 0.2250 X1(\text{TEMPERATURA}) + 0.2125 X2(\text{PRESION})$$

Coeficientes

Término	EE del		Valor T	Valor p	FIV
	Coef	coef.			
Constante	2.38	3.13	0.76	0.482	
X1(TEMPERATURA)	0.2250	0.0287	7.83	0.001	1.00
X2(PRESION)	0.2125	0.0144	14.80	0.000	1.00

El termino constante no es significativo. Los parámetros X1, X2, son significativos para un alfa de 0.05.

METODO BACKWARD (HACIA TRAS)



Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	2	1156.25	578.125	140.15	0.000
X1(TEMPERATURA)	1	253.13	253.125	61.36	0.001
X2(PRESION)	1	903.13	903.125	218.94	0.000
Error	5	20.62	4.125		
Total	7	1176.88			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
2.03101	98.25%	97.55%	95.51%

El modelo es significativo para un alfa de 0.05. El R-cuadrado de 98.25, el cual es aceptable.

METODO DE TODAS LAS VARIABLES SIN CONSTANTE

Ecuación de regresión

$$\begin{aligned} Y(\text{RENDIMIENTO}) = & -0.019 X1(\text{TEMPERATURA}) + 0.2875 X2(\text{PRESION}) + 0.375 X3(\text{CONCENTRACION}) \\ & + 0.00223 X1(\text{TEMPERATURA}) * X1(\text{TEMPERATURA}) \\ & - 0.000700 X1(\text{TEMPERATURA}) * X2(\text{PRESION}) \\ & + 0.00100 X1(\text{TEMPERATURA}) * X3(\text{CONCENTRACION}) \\ & - 0.00150 X2(\text{PRESION}) * X3(\text{CONCENTRACION}) \end{aligned}$$

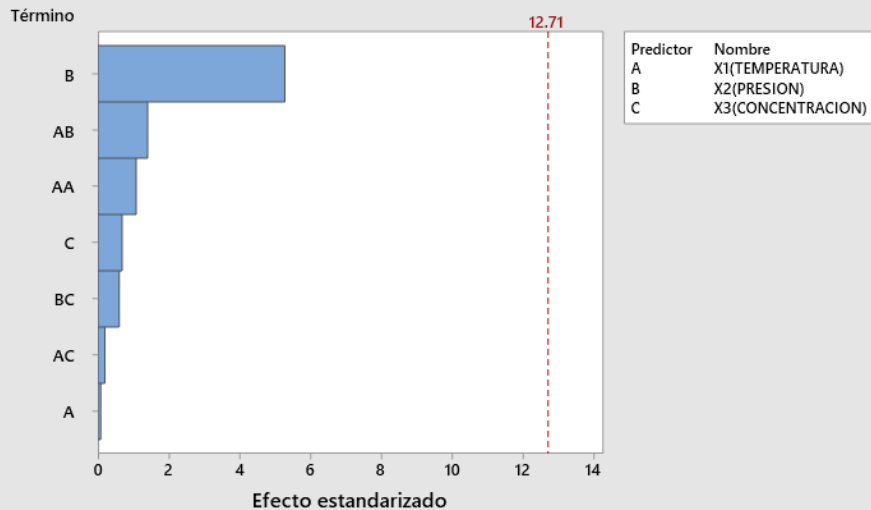
Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	FIV
X1(TEMPERATURA)	-0.019	0.232	-0.08	0.949	859.38
X2(PRESION)	0.2875	0.0545	5.28	0.119	190.00
X3(CONCENTRACION)	0.375	0.545	0.69	0.616	190.00
X1(TEMPERATURA)*X1(TEMPERATURA)	0.00223	0.00206	1.08	0.475	575.88
X1(TEMPERATURA)*X2(PRESION)	-0.000700	0.000500	-1.40	0.395	100.00
X1(TEMPERATURA)*X3(CONCENTRACION)	0.00100	0.00500	0.20	0.874	100.00
X2(PRESION)*X3(CONCENTRACION)	-0.00150	0.00250	-0.60	0.656	100.00

Ningún termino es significativo para un alfa de 0.05

METODO DE TODAS LAS VARIABLES SIN CONSTANTE

Diagrama de Pareto de efectos estandarizados
(la respuesta es Y(RENDIMIENTO), $\alpha = 0.05$)



Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	7	22083.9	3154.84	1009.55	0.024
X1(TEMPERATURA)	1	0.0	0.02	0.01	0.949
X2(PRESION)	1	87.0	87.01	27.84	0.119
X3(CONCENTRACION)	1	1.5	1.48	0.47	0.616
X1(TEMPERATURA)*X1(TEMPERATURA)	1	3.7	3.65	1.17	0.475
X1(TEMPERATURA)*X2(PRESION)	1	6.1	6.13	1.96	0.395
X1(TEMPERATURA)*X3(CONCENTRACION)	1	0.1	0.12	0.04	0.874
X2(PRESION)*X3(CONCENTRACION)	1	1.1	1.13	0.36	0.656
Error	1	3.1	3.12		
Total	8	22087.0			

Resumen del modelo

	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
S	1.76777	99.99%	99.89%
			99.09%

El modelo es significativo para un alfa de 0.05. El R-cuadrado de 99.99, el cual es aceptable. Pero es un modelo con parámetros no significativos.

METODO FORWARD (HACIA ADELANTE) SIN CONSTANTE

Ecuación de regresión

$$Y(\text{RENDIMIENTO}) = 0.2278 X1(\text{TEMPERATURA}) + 0.21388 X2(\text{PRESION}) + 0.00776 X3(\text{CONCENTRACION}) * X3(\text{CONCENTRACION})$$

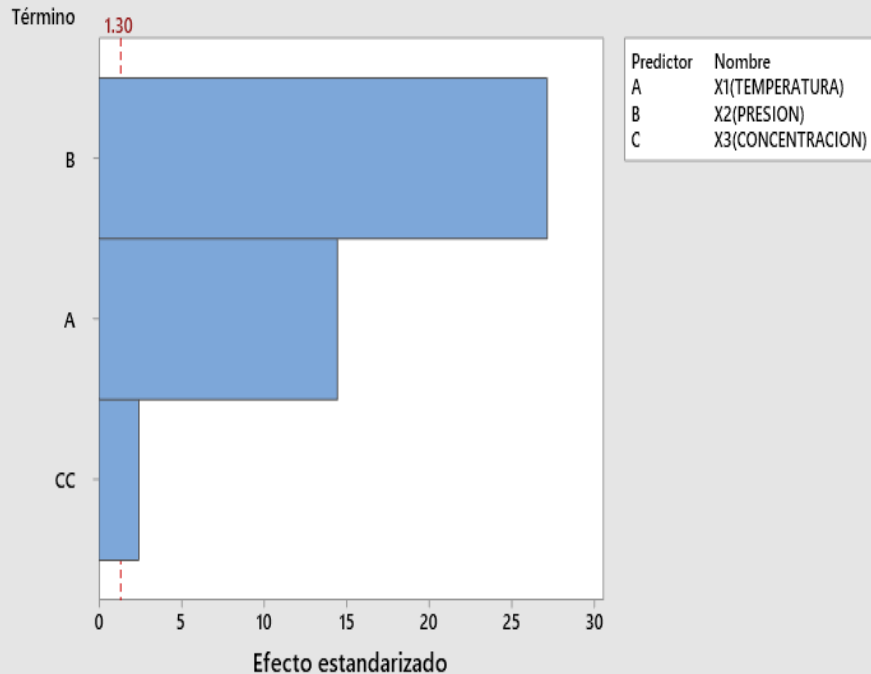
Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	FIV
X1(TEMPERATURA)	0.2278	0.0158	14.45	0.000	5.87
X2(PRESION)	0.21388	0.00788	27.13	0.000	5.87
X3(CONCENTRACION)*X3(CONCENTRACION)	0.00776	0.00320	2.42	0.060	3.30

Los parámetros X1, X2, X3*X3, son significativos para un alfa de 0.05.

METODO FORWARD (HACIA ADELANTE) SIN CONSTANTE

Diagrama de Pareto de efectos estandarizados
(la respuesta es Y(RENDIMIENTO), $\alpha = 0.25$)



Análisis de Varianza

Fuente	GL	SC	Ajust. MC	Ajust. Valor F	Valor p
Regresión	3	22076.4	7358.80	3473.81	0.000
X1(TEMPERATURA)	1	442.0	442.04	208.67	0.000
X2(PRESION)	1	1559.3	1559.26	736.07	0.000
X3(CONCENTRACION)*X3(CONCENTRACION)	1	12.4	12.41	5.86	0.060
Error	5	10.6	2.12		
Total	8	22087.0			

Resumen del modelo

	S	R-cuad. (ajustado)	R-cuad. (pred)
	1.45546	99.95%	99.92%
			99.86%

El modelo es significativo para un alfa de 0.05. El R-cuadrado de 99.99, el cual es aceptable.

METODO BACKWARD (HACIA TRAS) SIN CONSTANTE

Ecuación de regresión

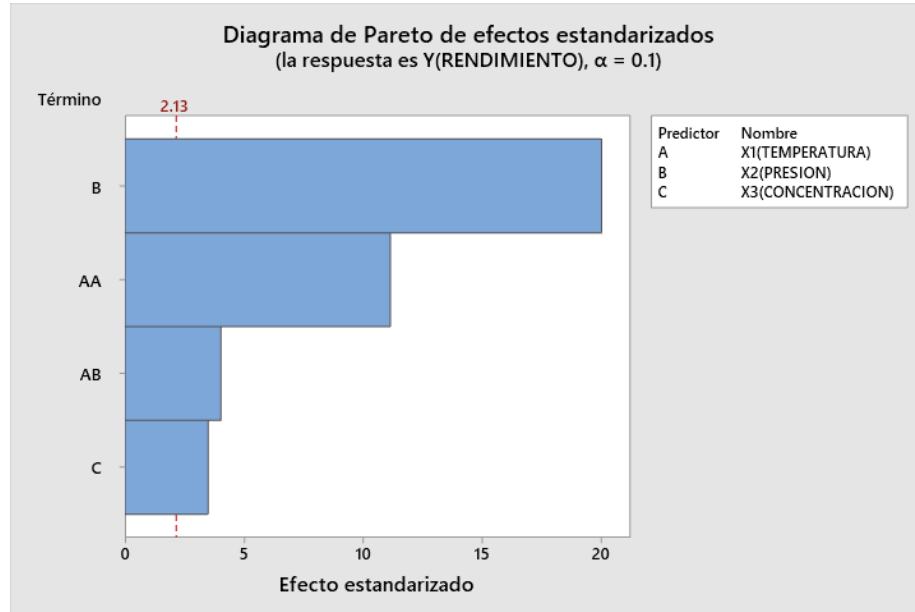
$$Y(\text{RENDIMIENTO}) = 0.2814 X_2(\text{PRESION}) + 0.2523 X_3(\text{CONCENTRACION}) \\ + 0.002402 X_1(\text{TEMPERATURA}) * X_1(\text{TEMPERATURA}) \\ - 0.000882 X_1(\text{TEMPERATURA}) * X_2(\text{PRESION})$$

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	FIV
X2(PRESION)	0.2814	0.0141	20.01	0.000	30.74
X3(CONCENTRACION)	0.2523	0.0721	3.50	0.025	8.08
X1(TEMPERATURA)*X1(TEMPERATURA)	0.002402	0.000216	11.14	0.000	15.34
X1(TEMPERATURA)*X2(PRESION)	-0.000882	0.000219	-4.03	0.016	46.56

Los parámetros X2, X3, X1*X1, X1*X2, son significativos para un alfa de 0.05.

METODO BACKWARD (HACIA TRAS) SIN CONSTANTE



Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	4	22081.9	5520.46	4290.01	0.000
X2(PRESION)	1	515.1	515.09	400.28	0.000
X3(CONCENTRACION)	1	15.8	15.76	12.25	0.025
X1(TEMPERATURA)*X1(TEMPERATURA)	1	159.8	159.79	124.17	0.000
X1(TEMPERATURA)*X2(PRESION)	1	20.9	20.87	16.22	0.016
Error	4	5.1	1.29		
Total	8	22087.0			

Resumen del modelo

	R-cuad.	R-cuad.
S	R-cuad. (ajustado)	(pred)
1.13438	99.98%	99.95%
		99.90%

El modelo es significativo para un alfa de 0.05. El R-cuadrado de 99.98, el cual es aceptable.

El mejor modelo

- ❑ El mejor modelo es con el método de Backward (hacia atrás).
- ❑ Todos los parámetros son significativos
- ❑ El ANOVA del modelo es significativo
- ❑ Se tiene un R-Cuadrado aceptable.

Ecuación de regresión

$$\begin{aligned} Y(\text{RENDIMIENTO}) = & 0.2814 X2(\text{PRESION}) + 0.2523 X3(\text{CONCENTRACION}) \\ & + 0.002402 X1(\text{TEMPERATURA}) * X1(\text{TEMPERATURA}) \\ & - 0.000882 X1(\text{TEMPERATURA}) * X2(\text{PRESION}) \end{aligned}$$