

Considere los siguientes datos relativos al uso mensual de agua de una planta de producción (galones) a su producción mensual (toneladas), la media de la temperatura ambiente mensual (°F), y el número mensual de días de operación de la planta durante un periodo de 12 meses

Uso de Agua y	Producción x1	Media de Temperatura x2	días de operación x3
2228	98.5	67.4	19
2609	108.2	70.3	20
3088	109.6	82.1	21
2378	101	69.2	21
1980	83.3	64.5	19
1717	70	63.7	21
2723	144.7	58	19
2031	84.4	58.1	20
1902	97.4	36.6	17
1721	131.8	49.6	23
2254	82.1	44.3	18
2522	64.5	44.1	19

- a) Analice el modelo completo, incluya la prueba de significancia y comente sobre el ajuste obtenido (10 puntos)

Prueba de significancia:

### Análisis de Varianza

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	3	814951	41.80%	814951	271650	1.92	0.206
x1	1	155472	7.97%	166523	166523	1.17	0.310
x2	1	360920	18.51%	608274	608274	4.29	0.072
x3	1	298559	15.31%	298559	298559	2.10	0.185
Error	8	1134716	58.20%	1134716	141839		
Total	11	1949666	100.00%				

### Prueba de significancia

$H_0$ : La prueba no es significativa,  $B_1=0$

$H_a$ : La prueba es significativa,  $B_1$  diferente de cero

Región de rechazo:

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

$\alpha = 0.05$  y  $p\text{-valor} = 0.206$

No rechazamos  $H_0$  dado que 0.206 no es menor que el valor de  $\alpha$ , con 95% de confianza podemos decir que las betas son iguales a cero a excepción del valor de  $B_0$

b) ¿Existe evidencia de multicolinealidad? Justifique (10 puntos)

VIFS	Regresión	Valor VIF
X1	X1-X2,X3	1.1
X2	X2-X1,X3	1.28
X3	X3-X1-X2	1.36

El valor de todos los VIF es menor a 10, entonces podemos decir que no hay evidencia de problemas de multicolinealidad en ninguna de las variables.

c) Indique el mejor modelo generado mediante mejores subconjuntos; comente el modelo final escogido, la prueba de significancia y el ajuste obtenido (15 puntos)

la respuesta es y

		R-cuadrado	R-cuadrado	Cp de	x	x	x	
Vars	R-cuadrado	(ajust)	(pred.)	Mallows	S	1	2	3
1	22.4	14.6	0.0	2.7	389.03	X		
1	8.0	0.0	0.0	4.6	423.58	X		
2	33.3	18.4	0.0	3.2	380.24		X	X
2	26.5	10.1	0.0	4.1	399.06	X	X	
3	41.8	20.0	0.0	4.0	376.62	X	X	X

El modelo escogido sería  $Y = X_2, X_3$  con un ajuste del 18.4%

## Análisis de Varianza

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	2	648428	33.26%	648428	324214	2.24	0.162
x2	1	436205	22.37%	638381	638381	4.42	0.065
x3	1	212223	10.89%	212223	212223	1.47	0.257
Error	9	1301238	66.74%	1301238	144582		
Total	11	1949666	100.00%				

### Prueba de significancia:

$H_0$ : La prueba no es significativa,  $B_1 = 0$

$H_a$ : La prueba es significativa,  $B_1$  diferente de cero

Región de rechazo:

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

$p\text{-valor} = 0.162 < \alpha = 0.05$

No rechazo  $H_0$ , con 95% de confianza podemos decir que la prueba no es significativa para las variables  $X_2$  y  $X_3$ .

- d) Calcule los IC y en caso de existir evidencia de regresión al origen del modelo elegido en inciso anterior realice el análisis e incluya la prueba de significancia correspondiente (15 puntos)

### Ecuación de regresión

$$y = 2981 + 20.61 x_2 - 97.9 x_3$$

### Coefficientes

Término	Coef	EE del coef.	IC de 95%	Valor T	Valor p	FIV
Constante	2981	1427	(-246, 6208)	2.09	0.066	
$x_2$	20.61	9.81	(-1.58, 42.80)	2.10	0.065	1.28
$x_3$	-97.9	80.8	(-280.8, 84.9)	-1.21	0.257	1.28

El intervalo de confianza con el 95% es (-246,6208), podemos ver que el intervalo incluye al cero entonces tenemos evidencia de regresión al origen

Volviendo a hacer el análisis del modelo elegido, pero ahora sin incluir el término de constante en el modelo, obtenemos:

### Ecuación de regresión

$$y = 18.2 x_2 + 59.2 x_3$$

### Coefficientes

Término	Coef	EE del coef.	IC de 95%	Valor T	Valor p	FIV
$x_2$	18.2	11.3	(-6.9, 43.3)	1.62	0.136	28.68
$x_3$	59.2	34.3	(-17.3, 135.6)	1.72	0.115	28.68

Intervalos de confianza con 95%:

$$-6.9 < B_1 < 43.3$$

$$-17.3 < B_2 < 135.6$$

### Análisis de Varianza

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	2	61457671	96.95%	61457671	30728836	159.02	0.000
$x_2$	1	60882906	96.04%	507093	507093	2.62	0.136
$x_3$	1	574766	0.91%	574766	574766	2.97	0.115
Error	10	1932446	3.05%	1932446	193245		
Total	12	63390117	100.00%				

**Prueba de significancia:**

$H_0$ : La prueba no es significativa,  $B_1=0$

$H_a$ : La prueba es significativa,  $B_1$  diferente de cero

Región de rechazo:

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

$p\text{-valor}=0 < \alpha=0.05$

Rechazo  $H_0$ , con 95% de confianza podemos decir que la prueba es significativa para este modelo

e) Calcule los residuales del modelo ¿Existe evidencia de puntos atípicos? Justifique (10 puntos)

Para el modelo  $Y-X_2, X_3$  los residuales son:

RESID	RESIDEST
-126.166712	-0.31126111
142.743261	0.35361778
347.241839	0.97043746
-127.35288	-0.30887472
-321.246145	-0.77685466
-687.986288	-1.64435951
540.368919	1.27985644
-212.625388	-0.50618594
228.225126	0.57864084
-545.021762	-1.54356554
380.542437	0.9274371
593.022671	1.49282766

No existe evidencia de datos atípicos ya que ningún valor de los residuales estandarizados es mayor a 2 o menor que -2.

f) Estime el uso de agua de la planta durante un mes cuando su producción es 90 toneladas, la media de la temperatura ambiente es 65°F, y opera por 20 días (10 puntos)

Del a ecuación de la regresión obtenida en el mejor modelo

$$Y=18.2x_2+59.3x_3$$

Sustituimos los valores en la ecuación

$$Y=18.2(65^\circ\text{F})+59.3(20\text{días})$$

$$Y=2369$$

-->Se estima un uso de agua mensual de 2369 galones

- g) Del análisis de mejores subconjuntos identifique la variable con mejor desempeño y analice si un modelo polinómico puede mejorar el ajuste del modelo que involucra dicha variable con la respuesta. Ajuste un modelo cuadrático (10 puntos) y un modelo cúbico, incluya en ambos casos la prueba de significancia y comente sobre el ajuste obtenido (10 puntos)

la respuesta es y

Vars	R-cuadrado	R-cuadrado (ajust)	R-cuadrado (pred.)	Cp de Mallows	S	x 1	x 2	x 3
1	22.4	14.6	0.0	2.7	389.03		X	
1	8.0	0.0	0.0	4.6	423.58	X		
2	33.3	18.4	0.0	3.2	380.24		X	X
2	26.5	10.1	0.0	4.1	399.06	X	X	
3	41.8	20.0	0.0	4.0	376.62	X	X	X

Podemos ver que la variable X2 sería la mejor para un modelo polinómico ya que tiene un ajuste del 14.6%

Análisis para el modelo polinómico con Y-X2

Ecuación:  $y = 1373.11 + 15.07X_2$

Tiene un ajuste del 22.373%

ANÁLISIS DE VARIANZA					
	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	436204.849	436204.849	2.88216699	0.12041416
Residuos	10	1513461.4	151346.14		
Total	11	1949666.25			

Prueba de significancia:

$H_0$ : La prueba no es significativa,  $B_1 = 0$

$H_a$ : La prueba es significativa,  $B_1$  diferente de cero

Región de rechazo:

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

$p\text{-valor} = 0.240 < \alpha = 0.05$

No rechazamos  $H_0$ , con 95% de confianza podemos decir que la prueba no es significativa

### Análisis para el modelo cuadrático con $Y-X_2, X_2^2$

Ecuación:  $y=4433.55-95.19x_2+0.946x_2^2$

Tiene un ajuste del 39.048%

#### ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	2	761296.08	380648.04	2.88279902	0.107761
Residuos	9	1188370.17	132041.13		
Total	11	1949666.25			

#### Prueba de significancia:

$H_0$ : La prueba no es significativa,  $B_1=0$

$H_a$ : La prueba es significativa,  $B_1$  diferente de cero

Región de rechazo:

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

$p\text{-valor}=0.2155<\alpha=0.05$

No rechazo  $H_0$ , con 95% de confianza podemos decir que la prueba no es significativa

### Análisis para el modelo cubico con $Y-X_2, X_2^2, X_2^3$

Ecuación:  $y=-5049.49+413.67x_2-7.80x_2^2+0.048x_2^3$

Tiene un ajuste del 46.52%

#### ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	3	907079.062	302359.687	2.32007215	0.15180642
Residuos	8	1042587.19	130323.398		
Total	11	1949666.25			

#### Prueba de significancia:

$H_0$ : La prueba no es significativa,  $B_1=0$

$H_a$ : La prueba es significativa,  $B_1$  diferente de cero

Región de rechazo:

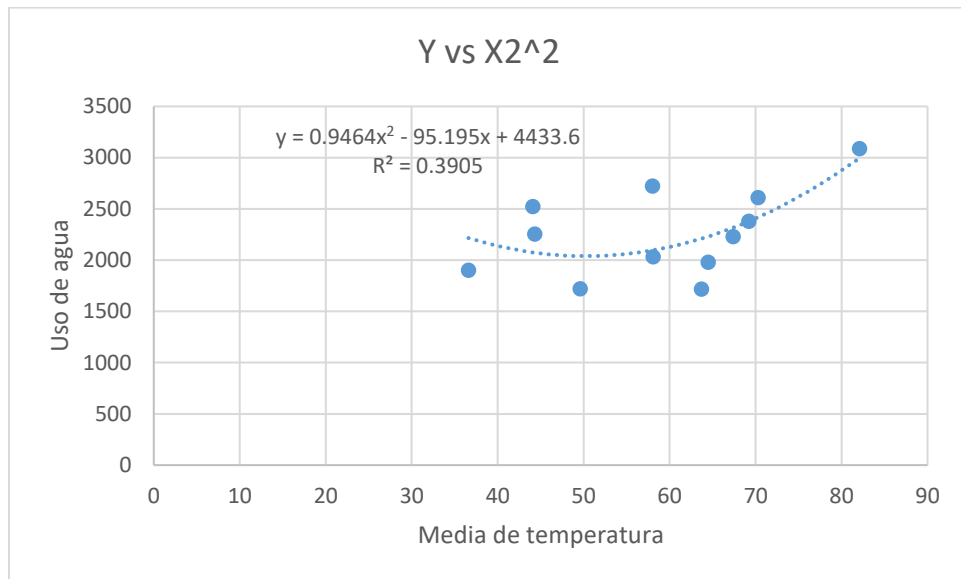
Rechazo  $H_0$  si  $p\text{-valor}<\alpha$

$p\text{-valor}=0.30<\alpha=0.05$

No rechazo  $H_0$ , con 95% de confianza podemos decir que la prueba no es significativa

- h) Incluya la gráfica de dispersión de x empleada vs y incluyendo la comparativa de ambos modelos polinómicos ajustados. ¿Cuál modelo elegiría como el mejor? (10 puntos)

Gráfica del modelo cuadrático



Gráfica del modelo cubico

