



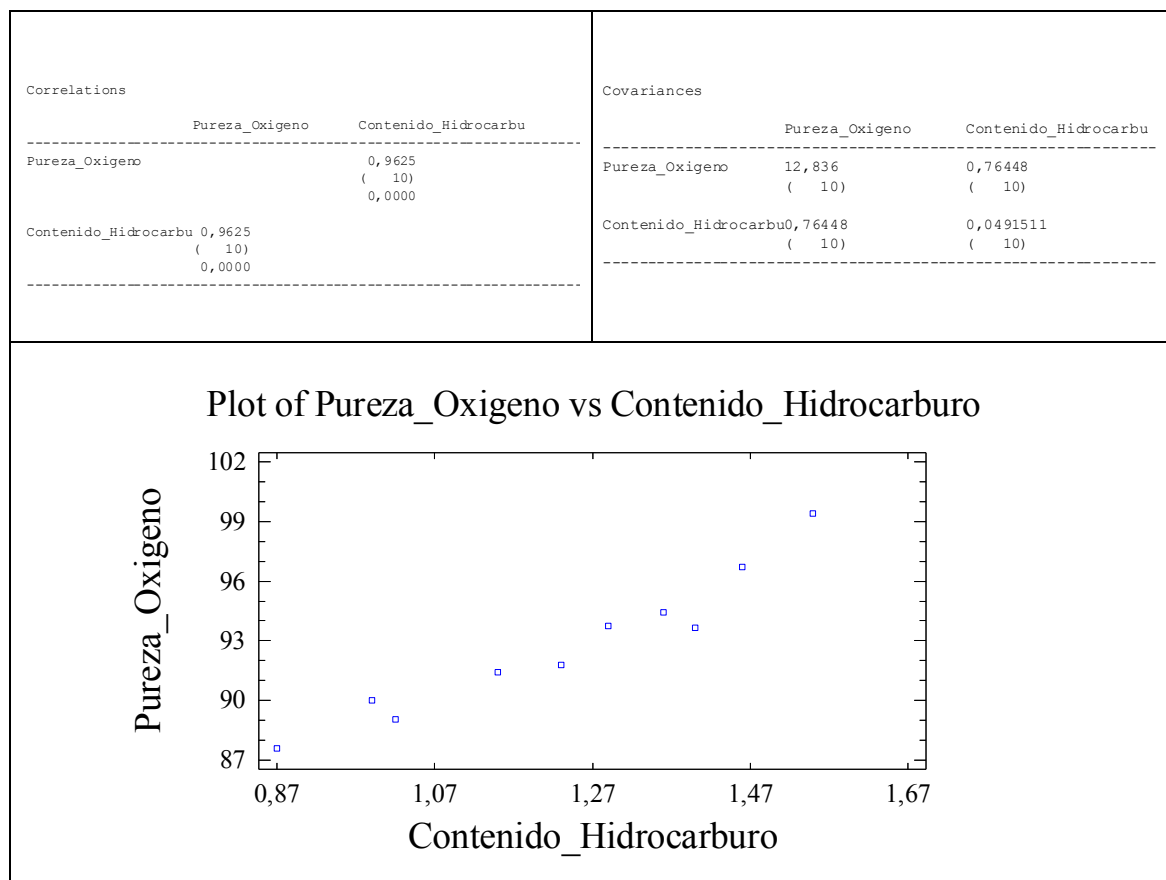
EJERCICIOS DE REGRESIÓN

Ejercicio1

La empresa **DESTILER S.A.** está interesada en saber si existe algún tipo de relación entre la pureza de oxígeno obtenido en un proceso de destilación y la presencia de hidrocarburos en el condensador principal de la columna de destilación. Para ello ha realizado 10 pruebas con los resultados que se muestran en la siguiente tabla.

	1	2	3	4	5	6	7	8	9	10
Pureza_Oxígeno (%)	90.0	89.0	91.4	93.7	96.7	94.4	87.5	91.7	99.4	93.6
	1	5	3	4	3	5	9	7	2	5
Contenido_Hidrocarburo (%)	0.99	1.02	1.15	1.29	1.46	1.36	0.87	1.23	1.55	1.40

A partir de los datos anteriores la empresa ha elaborado las siguientes herramientas estadísticas:





Regression Analysis - Linear model: $Y = a + b \cdot X$

Dependent variable: Pureza_Oxigeno

Independent variable: Contenido_Hidrocarburo

Parameter	Estimate	Standard Error	T Statistic	P-Value
Intercept	73,6219	1,93808	37,987	0,0000
Slope	15,5537	1,55068	10,0302	0,0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	107,014	1	107,014	100,61	0,0000
Residual	8,50964	8	1,0637		
Total (Corr.)	115,524	9			

Correlation Coefficient = 0,962465

R-squared = 92,6339 percent

Standard Error of Est. = 1,03136

Predicted Values

X	Predicted Y	95,00% Prediction Limits		95,00% Confidence Limits	
		Lower	Upper	Lower	Upper
0,87	87,1536	84,3433	89,9639	85,6565	88,6507
1,55	97,7301	94,9887	100,471	96,3667	99,0934
1,5	96,9524	94,2802	99,6246	95,7342	98,1706
0,89	87,4646	84,6866	90,2427	86,0289	88,9004
1,43	95,8636	93,2707	98,4566	94,8307	96,8966

Si la empresa te pasase toda esta información estadística, interpreta cada una de las herramientas elaboradas indicando, justificadamente, qué finalidad tiene cada una de ellas y la conclusión que se puede extraer. Empieza identificando quién sería la variable dependiente y quién la independiente.

Ejercicio 2

La empresa **PINK S.A.** ha diseñado unas pruebas para medir el efecto de un cierto tipo de aditivo en el tiempo de secado de pintura obteniéndose los siguientes resultados:

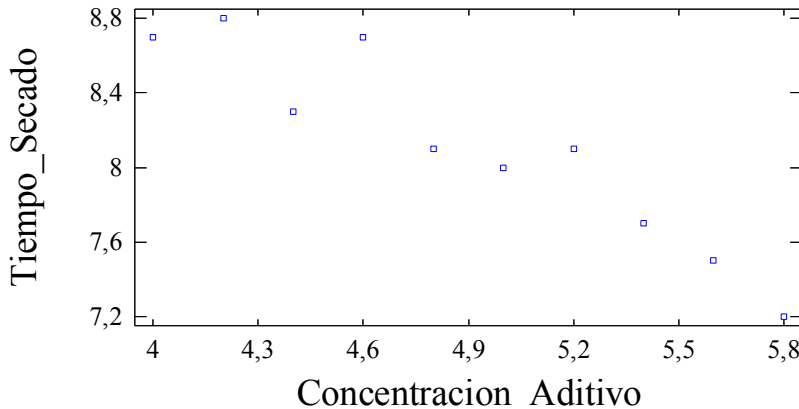
	1	2	3	4	5	6	7	8	9	10
Concentración_Aditivo (%)	4.0	4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8
Tiempo_Secado (h)	8.7	8.8	8.3	8.7	8.1	8.0	8.1	7.7	7.5	7.2



A partir de los datos anteriores la empresa ha elaborado las siguientes herramientas estadísticas:

Correlations		Covariances	
Concentracion_AditivTiempo_Secado		Concentracion_AditivTiempo_Secado	
Concentracion_Aditiv	-0,9408 (10) 0,0000	Concentracion_Aditiv	0,366667 (10)
Tiempo_Secado	-0,9408 (10) 0,0000	Tiempo_Secado	-0,305556 (10)

Plot of Tiempo_Secado vs Concentracion_Aditivo



Concentracion_Aditivo	Tiempo_Secado
4,0	8,7
4,3	8,8
4,4	8,3
4,6	8,7
4,8	8,1
5,0	8,0
5,2	8,1
5,4	7,7
5,6	7,5
5,8	7,2

Regression Analysis - Linear model: Y = a + b*X				
Dependent variable: Tiempo_Secado				
Independent variable: Concentracion_Aditivo				
Parameter	Estimate	Standard Error	T Statistic	P-Value
Intercept	12,1933	0,523576	23,2885	0,0000
Slope	-0,833333	0,106126	-7,85234	0,0000

Analysis of Variance					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	2,29167	1	2,29167	61,66	0,0000
Residual	0,297333	8	0,0371667		
Total (Corr.)	2,589	9			

Correlation Coefficient = -0,940827

R-squared = 88,5155 percent

Standard Error of Est. = 0,192787

Predicted Values

X	Predicted Y	95,00% Prediction Limits		95,00% Confidence Limits	
		Lower	Upper	Lower	Upper
4,3	8,61	8,12116	9,09884	8,40671	8,81329
4,5	8,44333	7,9669	8,91976	8,27202	8,61464
4,9	8,11	7,64373	8,57627	7,96942	8,25058
5,1	7,94333	7,4745	8,41216	7,79447	8,09219
5,3	7,77667	7,30024	8,2531	7,60536	7,94798

Si la empresa te pasase toda esta información estadística, interpreta cada una de las herramientas elaboradas indicando, justificadamente, qué finalidad tiene cada una de ellas y la conclusión que se puede extraer. Empieza identificando quién sería la variable dependiente y quién la independiente.

Problema 3

La factoría **ACEROX** quiere controlar el consumo de energía diario, medido en termias, (utilizada en su mayor parte en la climatización de las naves) Se ha establecido el siguiente modelo de regresión lineal para la variable consumo en función de la temperatura. Interpreta los parámetros estimados.

Análisis de Regresión - Modelo Lineal $Y = a + b \cdot X$					
Variable dependiente: Consumo					
Variable independiente: Temperatura					
Parámetro	Error Estimación	Estadístico estándar	T	P-Valor	
Ordenada	449,227	7,63321	58,8516	0,0000	
Pendiente	-18,4226	0,626782	-29,3924	0,0000	
Análisis de la Varianza					
Fuente	Suma de cuadrados	GL	Cuadrado medio	Cociente-F	P-Valor
Modelo	552226,0	1	552226,0	863,91	0,0000
Residuo	35156,9	55	639,216		
Total (Corr.)	587383,0	56			
Coeficiente de Correlación = -0,969612					
R-cuadrado = 94,0147 porcentaje					
R-cuadrado (ajustado para g.l.) = 93,9058 porcentaje					
Error estándar de est. = 25,2827					

- ¿Podemos suponer que la variable temperatura es significativa desde el punto de vista estadístico? Explica qué implica ese hecho.
- Interpreta en este contexto los parámetros del modelo y propón la mejor estimación posible para ellos.



- c) ¿Qué porcentaje de la variabilidad del consumo diario de energía está explicado por otras variables distintas a la temperatura?
- d) ¿Qué consumo energético medio podríamos esperar si la temperatura fuese igual a 3 grados? ¿Podríamos predecir el consumo energético medio para una temperatura de 23 grados, si sabemos que en la muestra la temperatura máxima ha sido de 19.5 grados?, ¿por qué?
- e) Si suponemos que la variable Consumo diario de energía cuando la temperatura es igual a 3 grados se comporta como una normal, estima la media y la desviación para esa variable. Calcula entre qué valores se encontrará la variable con una probabilidad del 95%.