

# Transformaciones y Ponderación para corregir Inadecuaciones del Modelo

Carlos Enrique Ponce Villagran

Facultad de Ciencias Físico-Matemáticas

25 de noviembre de 2019



# 1 Contexto del Problema



# Contexto del Problema

La siguiente tabla muestra la presión de vapor del agua para diversas temperaturas.

n	Temperatura(K)	Presión de vapor (mm Hg)
1	273	4.6
2	283	9.2
3	293	17.5
4	303	31.8
5	313	55.3
6	323	92.5
7	333	149.4
8	343	233.7
9	353	355.1
10	363	525.8
11	373	760



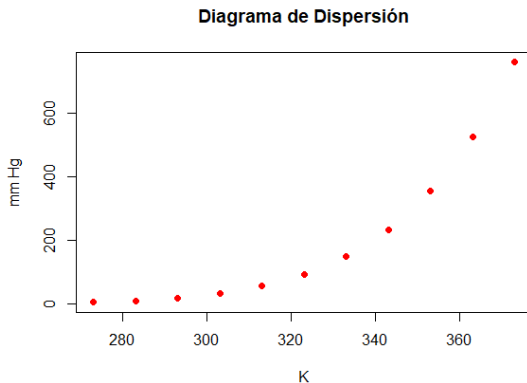
- a) Trazar un diagrama de dispersión. ¿Parece que será adecuado un modelo de línea recta?
- b) Ajustar el modelo rectilíneo. Calcular los estadísticos de resumen y las gráficas de residuales. ¿Cuáles son las conclusiones acerca de la adecuación del modelo?
- c) Según la ecuación de Clausius-Clapeyron, de la química física,

$$\ln(p_v) \propto -\frac{1}{T}$$

Repetir la parte b usando la transformación adecuada basada en esta información.



a) El diagrama de dispersión es el siguiente:



**Figura:** Diagrama de Dispersión de los Datos



Como podemos observar el modelo no tiene un comportamiento lineal lo cual es de esperarse ya que la presión tiende a aumentar bastante conforme la temperatura aumenta por lo que lo ideal seria aplicarle una transformación al modelo.



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1956.258	363.807	-5.377	0.000446
Temperatura	6.686	1.121	5.964	0.000212

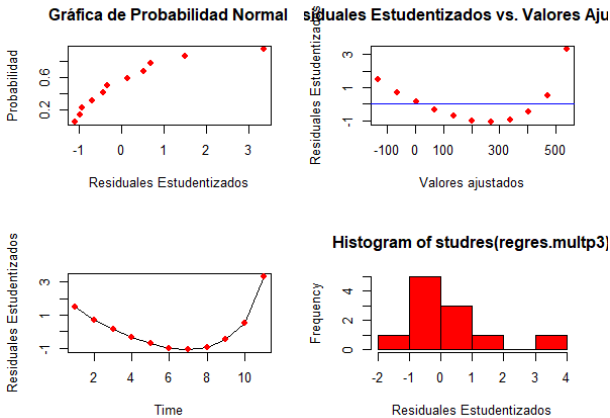
---

Residual standard error: 117.6 on 9 degrees of freedom

Multiple R-squared: 0.7981, Adjusted R-squared: 0.7756

F-statistic: 35.57 on 1 and 9 DF, p-value: 0.0002117





**Figura:** Análisis de Varianza





- b) Para ajustar el modelo consideraremos a  $y' = \log(y)$ , ya que la presión tiende a seguir un crecimiento exponencial, de este supuesto tenemos que la diagrama de dispersión es:

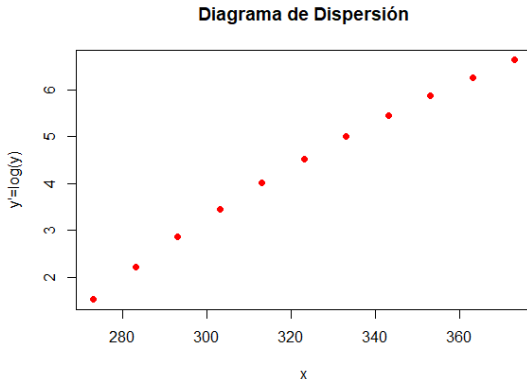


Figura: Diagrama de Dispersión



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-12.016241	0.543096	-22.12	3.72e-09
Temp	0.050666	0.001673	30.28	2.29e-10

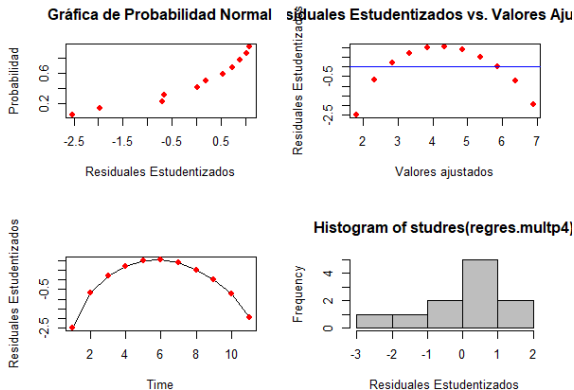
---

Residual standard error: 0.1755 on 9 degrees of freedom

Multiple R-squared: 0.9903,      Adjusted R-squared: 0.9892

F-statistic: 916.7 on 1 and 9 DF,    p-value: 2.288e-10





**Figura:** Análisis de la varianza



- La prueba Shapiro-Wilk arroja un  $p$ -valor de 0.06254 por lo que a un nivel de significancia de 5 % la hipótesis nula de que los residuales estudentizados distribuyen normal no se rechaza.
- La prueba Durbin-Watson arroja  $D = 0.3852354$ , con  $D_L = 1.08$  y  $D_U = 1.36$  a un nivel de significancia de 5 % entonces  $D < D_L$  por lo que se rechaza la hipótesis nula por lo que la correlación entre los residuales es distinta de cero.



Como la varianza no llega a ser constante con esta transformación se decidió aplicar una transformación a la regresora por el método de Box-Tidwell que nos da  $\alpha \approx 14$ , el modelo arrojó los siguientes datos:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-7.812e+00	1.503e+00	-5.197	0.000566
Temp	7.665e-34	3.634e-36	210.917	< 2e-16

Residual standard error: 3.721 on 9 degrees of freedom

Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998

F-statistic: 4.449e+04 on 1 and 9 DF,

p-value: < 2.2e-16



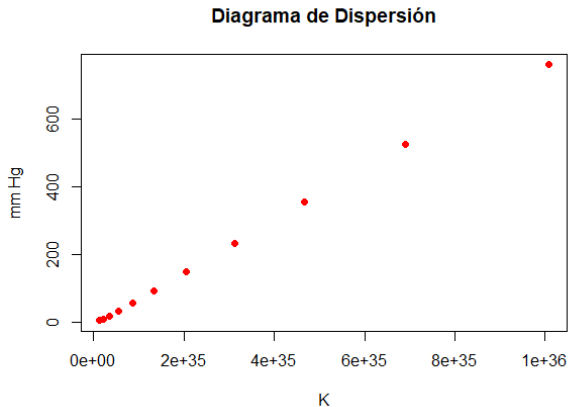
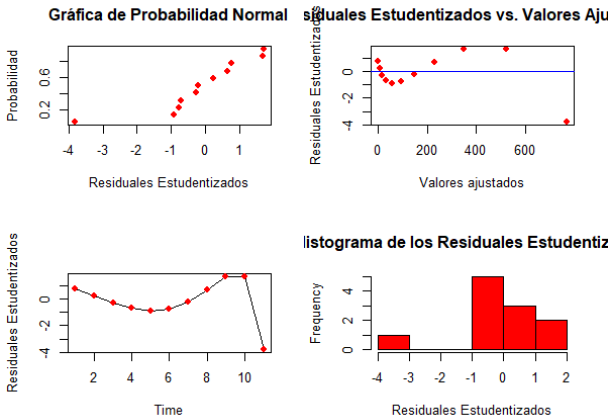


Figura: Transformación a la Regresora con  $x' = x^{14}$





**Figura:** Análisis de Varianza

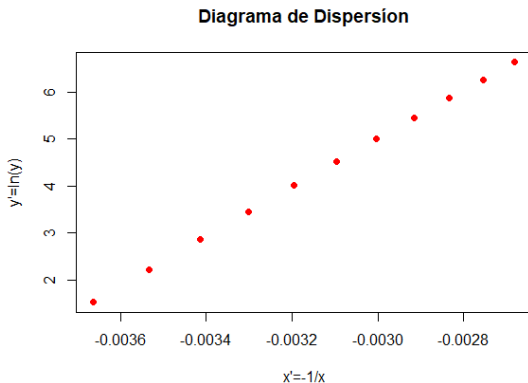


- La prueba Shapiro-Wilk arroja un  $p$ -valor de 0.1287 por lo que aun nivel de significancia de 5 % la hipótesis nula de que los residuales estudentizados distribuyen normal no se rechaza.
- La prueba Durbin-Watson arroja  $D = 1.164707$ , con  $D_L = 1.08$  y  $D_U = 1.36$  a un nivel de significancia de 5 % entonces  $D_L < D < D_U$  por lo que la prueba es inconclusa.





c) Tomando en consideración de que  $\ln(p_\nu) \propto -\frac{1}{T}$  ajustaremos el modelo con  $y' = \ln(y)$  y además  $x' = -1/x$  tenemos que el diagrama de dispersión se ajusta muy bien a un modelo lineal:



**Figura:** Diagrama de Dispersión



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.061e+01	6.325e-02	325.8	<2e-16
Tempe	5.201e+03	2.014e+01	258.3	<2e-16

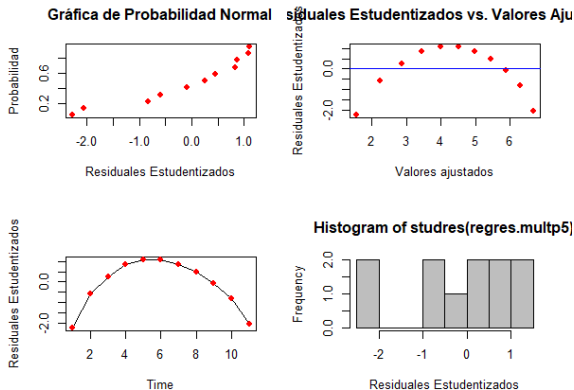
---

Residual standard error: 0.02067 on 9 degrees of freedom

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999

F-statistic: 6.672e+04 on 1 and 9 DF,  
p-value: < 2.2e-16





**Figura:** Análisis de la Varianza



- La prueba Shapiro-Wilk arroja un  $p$ -valor de 0.09603 por lo que aun nivel de significancia de 5 % la hipótesis nula de que los residuales estudentizados distribuyen normal no se rechaza.
- La prueba Durbin-Watson arroja  $D = 0.372751$ , con  $D_L = 1.08$  y  $D_U = 1.36$  a un nivel de significancia de 5 % entonces  $D < D_L$  por lo que se rechaza la hipótesis nula por lo que la correlación entre los residuales es distinta de cero.

