



# MODELOS LINEALIZABLES

REGRESIÓN LINEAL SIMPLE

# Modelos no lineales

A veces la premisa de una relación lineal entre  $x$  y el regresor es inadecuada, esto se podría revisar mediante un diagrama de dispersión o mediante una prueba de ajuste (la cual mostraría la falta de dicho ajuste). En algunos casos, una función no lineal se puede linealizar mediante una transformación adecuada. A éstos modelos no lineales se les llama intrínsecamente lineales o transformablemente lineales.

MODELO	FUNCIÓN LINEALIZABLE	TRANSFORMACIÓN	MODELO LINEAL ASOCIADO
potencia	$y = \beta_0 x^{\beta_1}$	$y^* = \ln y \quad x^* = \ln x$	$y^* = \ln \beta_0 + \beta_1 x^*$
exponencial	$y = \beta_0 e^{\beta_1 x}$	$y^* = \ln y$	$y^* = \ln \beta_0 + \beta_1 x$
logaritmo	$y = \beta_0 + \beta_1 \ln x$	$x^* = \ln x$	$y = \beta_0 + \beta_1 x^*$
recíproco	$y = \frac{x}{\beta_0 x - \beta_1}$	$y^* = 1/y \quad x^* = 1/x$	$y^* = \beta_0 - \beta_1 x^*$

# Sobre la transformación

Para trabajar los modelos linealizables debes considerar algunos puntos:

- ▶ Los modelos anteriormente descritos resultan de aplicar  $\ln$  a el modelo original, las  $x^*$  y  $y^*$  son las variables transformadas
- ▶ El modelo lineal asociado, así como su estimación de parámetros, me permite encontrar (mediante un despeje) los valores de los estimadores de la ecuación del modelo original
- ▶ La prueba de significancia se realiza considerando el modelo lineal asociado

# Uso del modelo lineal asociado

Consideremos el **modelo potencia**

$$y = \beta_0 x^{\beta_1}$$

Aplicando ln y sus propiedades

$$\ln y = \ln \beta_0 + \beta_1 \ln x$$

Utilizando \* para indicar la transformación

$$y^* = \beta_0^* + \beta_1 x^*$$

Hasta aquí se tiene que analizar una regresión con  $x^*$  vs  $y^*$  donde éstas son ln de las variables  $x$  y  $y$

Para “regresar” al modelo potencia hacemos uso de la igualdad

$$\beta_0^* = \ln \beta_0$$

Aplicando exponencial en ambos lados de la igualdad

$$e^{\beta_0^*} = \beta_0$$

Así este es el valor de  $\beta_0$  que sustituimos en el modelo potencia

$$y = \beta_0 x^{\beta_1}$$

# Uso del modelo lineal asociado

Consideremos el **modelo exponencial**

$$y = \beta_0 e^{\beta_1 x}$$

Aplicando  $\ln$  y sus propiedades

$$\ln y = \ln \beta_0 + \beta_1 x$$

Utilizando  $*$  para indicar la transformación

$$y^* = \beta_0^* + \beta_1 x$$

Hasta aquí se tiene que analizar una regresión con  $x$  vs  $y^*$  donde  $y^*$  es  $\ln$  de la variable  $y$

Para “regresar” al modelo exponencial hacemos uso de la igualdad

$$\beta_0^* = \ln \beta_0$$

Aplicando exponencial en ambos lados de la igualdad

$$e^{\beta_0^*} = \beta_0$$

Así este es el valor de  $\beta_0$  que sustituimos en el modelo exponencial

$$y = \beta_0 e^{\beta_1 x}$$

# Uso del modelo lineal asociado

Consideremos el **modelo logaritmo**

$$y = \beta_0 + \beta_1 \ln x$$

Utilizando \* para indicar la transformación

$$y = \beta_0 + \beta_1 x^*$$

Hasta aquí se tiene que analizar una regresión con  $x^*$  vs  $y$  donde  $x^*$  es  $\ln$  de la variable  $x$

Para “regresar” al modelo logaritmo hacemos uso de la igualdad  $x^* = \ln x$

# Uso del modelo lineal asociado

Consideremos el **modelo recíproco**

Utilizando el recíproco

Distribuyendo el denominador

Utilizando \* para indicar la transformación

Hasta aquí se tiene que analizar una regresión con  $x^*$  vs  $y^*$  donde  $x^*$  y  $y^*$  son el recíproco de las variables  $x$  y  $y$

Para “regresar” al modelo recíproco hacemos uso de las estimaciones de los Bs en el modelo recíproco

$$y = \frac{x}{\beta_0 x - \beta_1}$$

$$\frac{1}{y} = \frac{\beta_0 x - \beta_1}{x}$$

$$\frac{1}{y} = \beta_0 - \frac{\beta_1}{x}$$

$$y^* = \beta_0 - \beta_1 x^*$$

# ¿Cómo se realiza el reporte y comparación?

A modo de resumen del análisis de los modelos linealizables además de la grafica que incluya los modelos revisados en el mismo plano se llenara la siguiente tabla, utilizando como primer filtro la significancia del modelo, seguida del ajuste.

MODELO	ecuación estimada	Modelo lineal asociado	Prueba de significancia			R2
			hipótesis	p valor	conclusión	
lineal						
potencia						
exponencial						
logaritmo						
recíproco						



# Detección de datos atípicos

Ya que la varianza aproximada de los residuales es estimada con MSE, un escalamiento lógico de los residuales sería el de los residuales estandarizados, denotados por  $d_i$  tenemos  $d_i = \frac{e_i}{\sqrt{MSE}}$

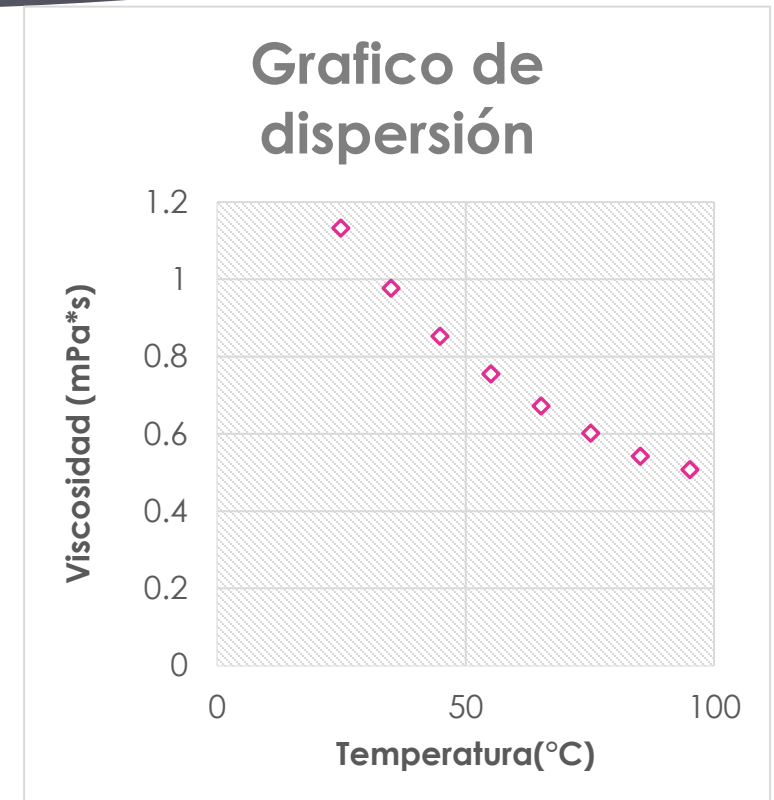
Si los residuales se estandarizan, valores  $|d_i|$  mayores a 3 indicaría la presencia de un valor atípico potencial

Cabe mencionar que, en algunos paquetes estadísticos, a partir de valores mayores a 2 ya se detectan como datos atípicos (por lo que este será nuestro criterio a considerar)

# Ejemplo en excel

Byers y Williams ("Viscosities of Binary and Ternary Mixtures of Polyaromatic Hydrocarbons", Journal of Chemical and Engineering Data, 32, 349-354, 1987), estudiaron el impacto de la temperatura sobre la viscosidad de las mezclas de tolueno y tetralina. A continuación los datos para mezclas con fracción molar de tolueno igual a 0.4

Temperatura(°C)	Viscosidad (mPa*s)
24.9	1.133
35	0.9772
44.9	0.8532
55.1	0.755
65.2	0.6723
75.2	0.6021
85.2	0.542
95.2	0.5074



# Ejemplo en excel

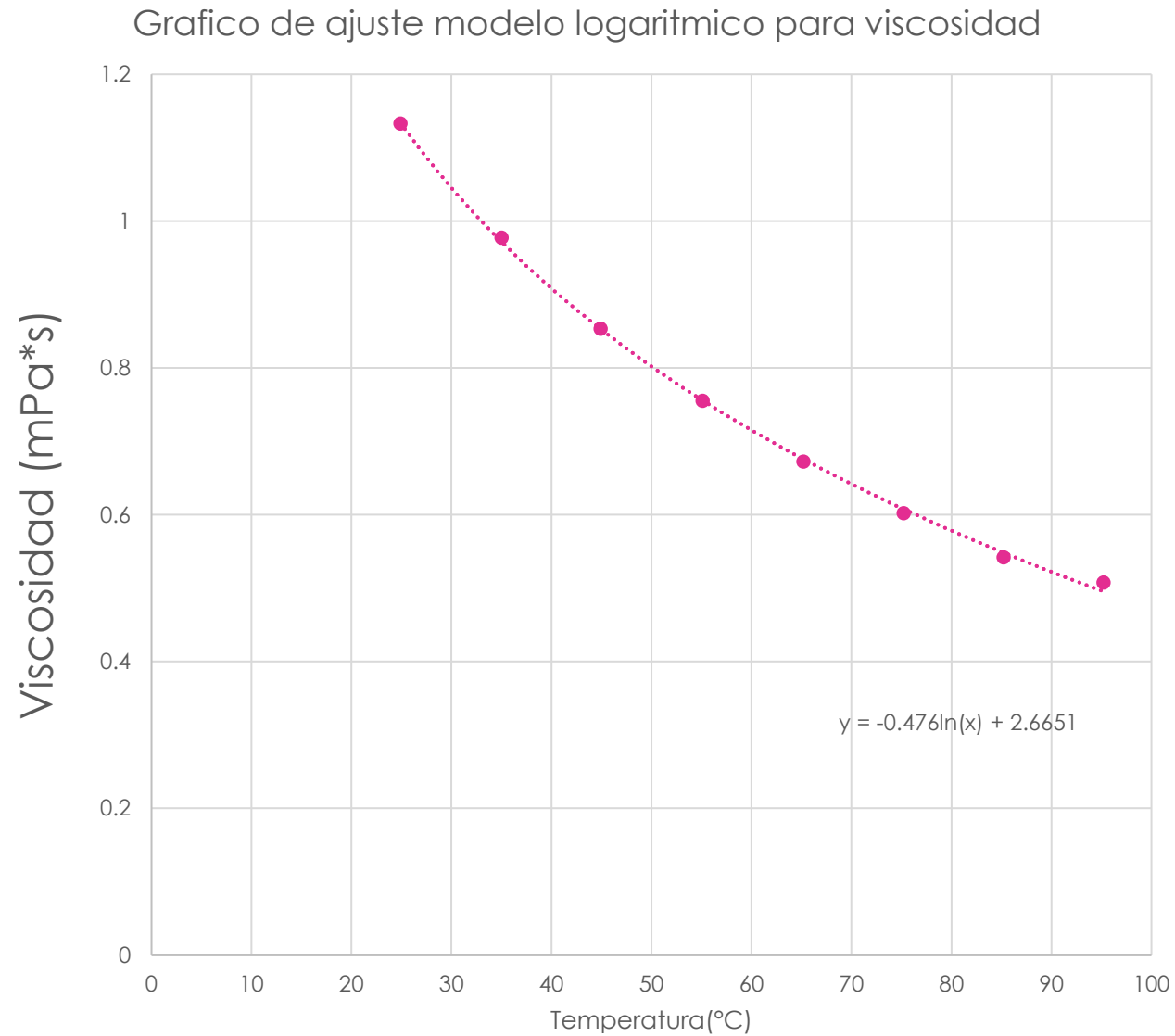
1. Debemos completar la tabla con las columnas correspondientes a las transformaciones
2. Para llenar la tabla de modelos linealizables se deberá correr una regresión por renglón: para el modelo potencia  $x^*$  vs  $y^*$ , para el modelo exponencial  $x$  vs  $y^*$ , para el modelo logaritmo  $x^*$  vs  $y$ , para el modelo recíproco  $1/x$  vs  $1/y$
3. Llenar la tabla (la siguiente diapositiva muestra la tabla completa y a continuación mostrare la prueba de significancia, ajuste y ecuación del modelo asociado para cada modelo)

x	y	$x^*$	$y^*$	$1/x$	$1/y$
Temperatura(°C)	Viscosidad (mPa*s)	lnx	lny		
24.9	1.133	3.2148678	0.12486898	0.04016064	0.88261253
35	0.9772	3.55534806	-0.02306394	0.02857143	1.02333197
44.9	0.8532	3.80443779	-0.15876129	0.02227171	1.17205813
55.1	0.755	4.00914972	-0.28103753	0.01814882	1.32450331
65.2	0.6723	4.17745947	-0.39705061	0.01533742	1.48743121
75.2	0.6021	4.32015123	-0.50733173	0.01329787	1.66085368
85.2	0.542	4.44500143	-0.61248928	0.01173709	1.84501845
95.2	0.5074	4.55597994	-0.67845563	0.0105042	1.97083169

# Tabla de modelos linealizables

MODELO	ecuación estimada	Modelo lineal asociado	Prueba de significancia			R2
			hipótesis	p valor	conclusión	
lineal	Viscosidad = $1.28 - 0.00876Temp$	*****	H0:regresión no significativa H1: regresión significativa	0.000	La regresión es significativa, las variables se relacionan de forma lineal al 99% de confianza	95.4%
potencia	Viscosidad= $8.499Temp^{-.614}$	$y^* = 2.14 - 0.614 x^*$	H0:regresión no significativa H1: regresión significativa	0.000	La regresión es significativa, las variables se relacionan de forma lineal al 99% de confianza	98.6%
exponencial	Viscosidad= $1.4564e^{-.0115Temp}$	$y^* = 0.376 - 0.0115Temp$	H0:regresión no significativa H1: regresión significativa	0.000	La regresión es significativa, las variables se relacionan de forma lineal al 99% de confianza	99.0%
logaritmo	Viscosidad= $2.6651 - 0.476\ln Temp$	Viscosidad = $2.6651 - 0.476x^*$	H0:regresión no significativa H1: regresión significativa	0.000	La regresión es significativa, las variables se relacionan de forma lineal al 99% de confianza	99.9%
recíproco	Viscosidad= $\frac{Temp}{2.13Temp-35.4}$	$(1/y) = 2.13 - 35.4 (1/x)$	H0:regresión no significativa H1: regresión significativa	0.001	La regresión es significativa, las variables se relacionan de forma lineal al 99% de confianza	81.7%

# Modelo logarítmico



# Residuales modelo logaritmo

Los residuales se calculan bajo el modelo elegido, según la tabla anterior; bajo el criterio de selección sobre los modelos significativos seguido del modelo con mejor ajuste

La  $y$  estimada resulta de la sustitución de los valores de  $x$  en el modelo elegido

Los residuales resultan de las diferencias de la  $y$  observada (en la muestra) y la  $y$  resultante del modelo elegido

El residual estandarizado resulta del cociente del residual ordinario entre la raíz de la varianza estimada (la varianza se estima con MSE)

Del valor de los residuales estandarizados se tiene que ninguno cumple con  $|di| > 2$  por lo que no hay datos atípicos

$y$ estimada	residual	residual estandarizado
1.13482293	-0.00182293	-0.288229837
0.97275432	0.00444568	0.702923296
0.85418761	-0.00098761	-0.156154805
0.75674474	-0.00174474	-0.275866843
0.67662929	-0.00432929	-0.684521293
0.60870801	-0.00660801	-1.044818763
0.54927932	-0.00727932	-1.150961155
0.49645355	0.01094645	1.730786078