

# Pauta Auxiliar 3

## Regresión lineal

19 de junio de 2019

**Profesora:** Carla Hermann A.

**Auxiliares:** Isabel Henríquez, Camilo Ramírez, Iván Rosas

**Pregunta 1** En esta pregunta se entregan expresiones a las que se les desea hacer regresión lineal, para esto se indica cual es la VARIABLE INDEPENDIENTE (denotada v.i.) y la VARIABLE DEPENDIENTE (denotada v.d.), mientras que los demás términos son constantes. Se pide modificar esta expresión para obtener una relación explícita a la que se le pueda hacer regresión lineal, identificando  $a$ ,  $b$  y las variables  $x$  e  $y$ .

- La variable independiente, corresponde a aquella que se puede controlar a voluntad en un experimento y se conoce con alta certeza. Para efectos de una regresión esta variable corresponde a  $x^*$  y se desea graficar esta, o una función de esta, en el eje  $x$ .

EJEMPLO: Al medir las frecuencias de resonancia para distintos modos de una barra, el número del modo  $n$  corresponde a aquello que se conoce de antemano y va variando, por ende es la variable independiente.

- La variable dependiente, corresponde a aquella que se ve modificada en función de  $x^*$ , notar que los términos constantes no se ven modificados al variar  $x^*$ . Para efectos de una regresión esta corresponde a  $y^*$  y se desea graficar esta, o una función de esta, en el eje  $y$ .

EJEMPLO: Para el mismo experimento del ejemplo anterior, la frecuencia corresponde a la variable dependiente, pues esta es la que varía a medida que  $n$  lo hace.

Para resolver esta problemática en general, se resuelven los siguientes pasos:

- (1) En caso que las v.i. y v.d. no estén explícitas, se requiere inferir del problema y de la expresión que lo modela cual variable corresponde a v.i. y v.d. respectivamente.
- (2) Identificar que tipo de relación constituye la expresión que modela el problema (lineal, semilogarítmica en  $x$ , semilogarítmica en  $y$  o ley de potencia).
- (3) Aplicar las funciones necesarias, como  $\ln(\cdot)$  y ordenar la expresión para que quede de la forma  $y = ax + b$ , identificando  $a$ ,  $b$ ,  $x$  e  $y$ .

**P1 (a)** Se tiene la expresión:

$$q + 1 = 2^{-n}$$

Se nos dice que  $n$  es la v.i. y que  $q$  es la v.d.

Una forma de ver esto es que  $n = x^*$  y  $q = y^*$  para visualizar bien la regresión, notemos que  $n$  es parte del exponente de una constante, esto corresponde a un caso semilogarítmico en  $y^*$ . De esta forma aplicamos logaritmo a ambos lados, quedando:

$$\ln(q + 1) = -\ln(2) \cdot n$$

(¿El motivo de esto? Al aplicar logaritmo estamos “bajando” la variable independiente ( $n$ ) del exponente, ya que nos molesta que esté ahí) Ahora ya podemos ver que la expresión está linealizada. Ya que en  $\ln(q + 1)$  no

se puede separar la  $q$  del 1, se considera la expresión completa  $\ln(q + 1)$  como  $y$ . Dejando explícita la forma  $y = ax + b$  queda:

$$\ln(q + 1) = (-\ln(2)) \cdot n + 0$$

Así se identifica que:

$$x = n \quad y = \ln(q + 1) \quad a = -\ln(2) \quad b = 0$$

**P1 (b)** Se tiene la expresión:

$$\epsilon E = \ln(f)B + \alpha^\beta + c$$

Se nos dice que  $B$  es v.i. y que  $E$  es v.d.

Al igual que en la parte anterior, se puede considerar  $B = x^*$  y  $E = y^*$ , notemos que tanto  $B$  como  $E$  no son parte de ningún exponente, logaritmo, potencia u otra función, mas aún, la expresión *ya está linealizada*, lo único que le falta para que quede de la forma  $y = ax + b$  es aislar la variable dependiente (*Esto para que la variable (y) quede "sola"*), por ende se procede a dividir por  $\epsilon$  quedando:

$$E = \frac{\ln(f)B}{\epsilon} + \frac{\alpha^\beta}{\epsilon} + \frac{c}{\epsilon}$$

Lo que corresponde a la expresión de la forma  $y = ax + b$ , haciendo explícito esto, se tiene:

$$E = \left( \frac{\ln(f)}{\epsilon} \right) \cdot B + \left( \frac{\alpha^\beta + c}{\epsilon} \right)$$

Así se identifica que:

$$x = B \quad y = E \quad a = \frac{\ln(f)}{\epsilon} \quad b = \frac{\alpha^\beta + c}{\epsilon}$$

**P1 (c)** Se tiene la expresión:

$$\log_{16}(\beta^2 X^3) = Y$$

Según lo presentado en el enunciado, se considera  $X = x^*$  e  $Y = y^*$ , para este caso, notar que  $X$  es parte de un logaritmo, por lo que esta expresión corresponde a un caso semilogarítmico en  $x^*$ , por lo que solo basta expandir la expresión de una forma adecuada, usando las propiedades de los logaritmos se llega a que:

$$\log_{16}(\beta^2 X^3) = Y \Rightarrow Y = 3 \log_{16}(X) + 2 \log_{16}(\beta) = Y \Rightarrow Y = \frac{3 \ln(X)}{\ln(16)} + \frac{2 \ln(\beta)}{\ln(16)}$$

Dejando explícita la forma  $y = ax + b$  queda:

$$Y = \left( \frac{3}{\ln(16)} \right) \cdot \ln(X) + \left( \frac{2 \ln(\beta)}{\ln(16)} \right)$$

Con esto:

$$x = \ln(X) \quad y = Y \quad a = \frac{3}{\ln(16)} \quad b = \frac{2 \ln(\beta)}{\ln(16)}$$

**P1 (d)** Se tiene la expresión:

$$\frac{(n-1)f}{n} = \frac{m}{\sqrt{z}}$$

En este caso  $z = x^*$  y  $f = y^*$ , se comienza por dejar aislado a un lado de la expresión el valor de  $y^*$  y a dejar explícita la potencia de la que es parte  $z$ , quedando:

$$f = \frac{nm}{n-1} z^{-\frac{1}{2}}$$

Ahora, notando que  $z$  es la base de una potencia, esta expresión corresponde a una ley de potencia, así, le aplicamos  $\ln(\cdot)$ , quedando:

$$\ln(f) = -\frac{1}{2} \ln(z) + \ln\left(\frac{nm}{n-1}\right)$$

Haciendo explícita la forma  $y = ax + b$ :

$$\ln(f) = \left(-\frac{1}{2}\right) \cdot \ln(z) + \left(\ln\left(\frac{nm}{n-1}\right)\right)$$

Con esto:

$$x = \ln(z) \quad y = \ln(f) \quad a = -\frac{1}{2} \quad b = \ln\left(\frac{nm}{n-1}\right)$$

**Pregunta 2** En esta pregunta se da la relación  $\eta = F/P$  y se pide verificar si  $\eta$  supera el umbral de 100lm/W, se menciona que  $P$  ya es conocido (vale 10.5 W) por lo que solo falta conocer el valor de  $F$ . Para resolver el problema se plantea la vía manual, como haciendo uso de Matlab, los párrafos que tengan relación con la solución en Matlab comenzarán con [MATLAB], puede ignorar estos párrafos si desea.

Para obtener el valor de  $F$  se diseña un experimento tal, que es posible obtenerlo haciendo una regresión lineal de ciertas medidas realizadas, la expresión que modela el experimento corresponde a:

$$E = \frac{F}{3\pi d^2}$$

Y los datos obtenidos son los siguientes:

Distancia $d$ (m)	0,1	0,5	0,7	1,5	5
Iluminancia $E$ (lx)	13 300	520	275	62	5,36

Teniendo estos dos valores, el primer paso corresponde a modificar la expresión que modela el experimento tal que sea posible realizarle una regresión, se comienza identificando las variables independientes y dependiente.

Analizando término a término:

- $3\pi$  es un valor constante conocido.
- $F$  si bien es desconocido, es constante (es único para la ampollita) y es precisamente el valor que estamos buscando.
- $d$  es un valor que podemos controlar y modificar a voluntad, por ende corresponde a la **variable independiente**, se puede decir que  $d = x^*$ .
- $E$  también es variable, pero esta no la podemos controlar directamente, y varía en función que  $d$  lo hace, por ende esta es la **variable dependiente**, se puede decir que  $E = y^*$ .

[MATLAB] Teniendo claros v.i. y v.d., se procede a anotar estos en un vector para las futuras operaciones a realizar, es importante anotarlos en orden, ya que así el  $i$ -ésimo elemento de uno se correlaciona con el  $i$ -ésimo del otro:

```
d = [0.1, 0.5, 0.7, 1.5, 5]; % Distancia a la ampollita (m)
E = [13300, 520, 275, 62, 5.36]; % Iluminancia medida (lx)
```

Ahora, sabiendo cual es la variable independiente y dependiente, se debe modificar la relación para que quede linealizada, notar que la v.i. es la base de una potencia, por lo que aplican los pasos de linealización de una ley de potencia, haciendo esto:

$$E = \frac{F}{3\pi d^2} = \frac{F}{3\pi} d^{-2} \Rightarrow \ln(E) = -2 \ln(d) + \ln\left(\frac{F}{3\pi}\right)$$

En donde queda la expresión en forma de recta si se toman:

$$x = \ln(d) \quad y = \ln(E) \quad a = -2 \quad b = \ln\left(\frac{F}{3\pi}\right)$$

Sabiendo a que corresponden  $x$  e  $y$ , se tabulan sus valores:

$x = \ln(d)$	-2,30	-0,69	-0,36	0,41	1,61
$y = \ln(E)$	9,50	6,25	5,62	4,13	1,68

[MATLAB] Usar software ahorra mucho cálculo manual, para definir los valores de  $x$  e  $y$  en Matlab según lo determinado se definen los siguientes nuevos vectores.

```
x = log(d);
y = log(E);
```

[MATLAB] Donde  $x$  e  $y$  corresponden a vectores en donde en su posición  $i$ -ésima contienen el valor del logaritmo natural del  $i$ -ésimo valor del vector sobre el que se aplicó este.

Teniendo claros los valores de  $x$  e  $y$  es posible calcular tanto  $a$  como  $b$ , para hacer esto se observan que las fórmulas son:

$$a = \frac{\frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i} - \frac{\sum_{i=1}^N y_i}{N}}{\frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N x_i} - \frac{\sum_{i=1}^N x_i}{N}} = \frac{\frac{S_{xy}}{S_x} - \frac{S_y}{N}}{\frac{S_{x^2}}{S_x} - \frac{S_x}{N}}$$

$$b = \frac{\sum_{i=1}^N y_i}{N} - a \cdot \frac{\sum_{i=1}^N x_i}{N} = \frac{S_y}{N} - a \frac{S_x}{N}$$

En donde  $S_x$ ,  $S_y$ ,  $S_{xy}$  y  $S_{x^2}$  se han definido para simplificar la notación, para continuar, hay que hacer los cálculos de  $x_i^2$  y  $x_i y_i$ , estos se dejan tabulados:

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$
$x_i^2$	5,30	0,48	0,13	0,16	2,59
$x_i y_i$	-21,86	-4,33	-2,00	1,67	2,70

[MATLAB] Estos datos pueden ser fácilmente calculados, para esto se definen los nuevos vectores:

```
x2 = x.^2;
xy = x.*y;
```

Teniendo todos los valores de  $x_i$ ,  $y_i$ ,  $x_i y_i$  y  $x_i^2$  ahora obtener es posible obtener  $S_x$ ,  $S_y$ ,  $S_{x^2}$  y  $S_{xy}$ , estos valores se dejan igualmente tabulados.

$S_x$	$S_y$	$S_{x^2}$	$S_{xy}$
-1.34	27.17	8.66	-23.83

[MATLAB] Usando las siguientes líneas de código se pueden calcular estas sumas, notar que además, las expresiones de  $a$  y  $b$  contienen un  $N$ , que representa el total de datos obtenidos, esto también se tiene que dejar expreso en el código:

```
N = 5;
Sx = sum(x);
Sy = sum(y);
Sx2 = sum(x2);
Sxy = sum(xy);
```

Con todos los valores necesarios listos, se calculan los coeficientes  $a$  y  $b$  se llega a que:

$$a = \frac{\frac{-23.83}{-1.34} - \frac{27.17}{5}}{\frac{8.66}{-1.34} - \frac{-1.34}{5}} = -1,99 \quad \wedge \quad b = \frac{27.17}{5} - (-1.99) \cdot \frac{-1.34}{5} = 4.90$$

[MATLAB] Para calcular estas expresiones, se operan los números tal y como están en la fórmula, notar que no se tiene que poner el caracter `.` antes de las operaciones, ya que ninguno de los elementos a operar es un vector.

```
a = (Sxy/Sx - Sy/N)/(Sx2/Sx - Sx/N);
b = Sy/N - a*Sx/N;
```

Teniendo  $a$  y  $b$  solo queda hacer el cálculo del coeficiente de regresión lineal, para esto se consideran las expresiones para  $\chi^2$  y  $R^2$ :

$$\chi^2 = \sum_{i=1}^N (ax_i + b - y_i)^2$$

$$R^2 = 1 - \frac{\chi^2}{\sum_{i=1}^N (y_i - \langle y \rangle)^2} \quad \text{con} \quad \langle y \rangle = \frac{1}{N} \sum_{i=1}^N y_i$$

Se comienza por calcular  $\chi^2$ , de esta forma es necesario obtener cada uno de los términos  $(ax_i + b - y_i)$ , estos están tabulados a continuación:

$ax_1 + b - y_1$	$ax_2 + b - y_2$	$ax_3 + b - y_3$	$ax_4 + b - y_4$	$ax_5 + b - y_5$
-0.0113	0.0277	-0.0049	-0.0319	0.0204

[MATLAB] El cálculo de estos términos se puede almacenar en un vector que se llame `Chi_t` (Chi-términos):

```
Chi_t = a*x+b-y;
```

Ahora, calculando los término de la suma presente en la expresión de  $R^2$ , calculando que  $\langle y \rangle = 5.434$ , se tiene:

$y_1 - \langle y \rangle$	$y_2 - \langle y \rangle$	$y_3 - \langle y \rangle$	$y_4 - \langle y \rangle$	$y_5 - \langle y \rangle$
4.062	0.820	0.183	-1.307	-3.755

[MATLAB] Para estos valores se definirán `Pro_y` (Promedio de  $y$ ), `Var_y` (Variación de  $y$  respecto a  $\langle y \rangle$ ), como:

```
Pro_y = mean(y);
Var_y = y-Pro_y;
```

Con todos los términos calculados y aplicando las fórmulas, se llega a que que  $\chi^2 = 0.0023$  y  $R^2 = 0.9999$  por lo que se da a entender que la regresión entrega un resultado altamente preciso.

[MATLAB] Para calcular  $\chi^2$  y  $R^2$  se programa:

```
Chi2 = sum(Chi_t.^2);
R2 = 1 - Chi2/sum(Var_y.^2);
```

Habiendo realizado la regresión lineal y dejado explícitos todos los valores necesarios, solo basta recuperar  $F$  de la expresión que se definió como  $b$ , en efecto se tiene que:

$$b = \ln\left(\frac{F}{3\pi}\right) \Rightarrow e^b = \frac{F}{3\pi} \Rightarrow F = 3\pi e^b \approx 1265$$

Recordando su unidad de medida, se tiene que  $F = 1265 \text{ lm}$ , así finalmente se extrae el rendimiento luminoso que corresponde a:

$$\eta = \frac{F}{P} = \frac{1265 \text{ lm}}{10.5 \text{ W}} = 120.5 \frac{\text{lm}}{\text{W}}$$

Por lo que la ampolleta sí cumple los estándares solicitados.

[MATLAB] Si bien el cálculo final es sencillo, este también pudo haber sido programado como:

```
P = 10.5;           % Potencia consumida (W)
F = 3*pi*exp(b);   % Flujo luminoso (lm)
eta = F/P;         % Rendimiento luminoso (lm/W)
```

Para corroborar que la regresión esté bien hecha, basta usar la función `polyfit` en primer orden (ya que estamos aproximando por una recta), la cual devolverá un vector de dos números, el primero corresponde a  $a$  y el segundo a  $b$ , esta simple línea, resume todo el desarrollo anterior:

```
Poly = polyfit(x,y,1);
```

En la página siguiente se deja un código ejecutable que permite la resolución del problema. **(¡Lo hace bastante más sencillo!).**



```
1 % Resolución pregunta 2 del aux. 3 de FI2003-6 Otoño 2019
2
3 % Vectores de datos obtenidos experimentalmente
4 d = [0.1, 0.5, 0.7, 1.5, 5];
5 E = [13300, 520, 275, 62, 5.36];
6
7 % Valores transformados útiles para regresión lineal
8 x = log(d);
9 y = log(E);
10 x2 = x.^2;
11 xy = x.*y;
12
13 % Cálculo de sumatorias
14 N = 5;
15 Sx = sum(x);
16 Sy = sum(y);
17 Sx2 = sum(x2);
18 Sxy = sum(xy);
19
20 % Cálculo de coeficientes a y b
21 a = (Sxy/Sx - Sy/N)/(Sx2/Sx - Sx/N);
22 b = Sy/N - a*Sx/N;
23
24 % Cálculo de términos útiles para obtener Chi cuadrado y R cuadrado
25 Chi_t = a*x+b-y;
26 Pro_y = mean(y);
27 Var_y = y-Pro_y;
28
29 % Cálculo de Chi cuadrado y R cuadrado
30 Chi2 = sum(Chi_t.^2);
31 R2 = 1 - Chi2/sum(Var_y.^2);
32
33 % Obtención de parámetros del problema
34 P = 10.5; % Potencia consumida (W)
35 F = 3*pi*exp(b); % Flujo luminoso (lm)
36 eta = F/P; % Rendimiento luminoso (lm/W)
37
38 % Verificación final de que la regresión lineal fue realizada correctamente
39 Poly = polyfit(x,y,1);
```