

## Resumen de comandos sobre REGRESIÓN LINEAL con wxMaxima

**Introducción de datos:** deben expresarse como una *matriz de 2 columnas*, de la forma

$$z = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}$$

esto puede hacerse de varios modos

Modo 1: introduciendo los datos por parejas

```
z : matrix([x1, y1], [x2, y2], ..., [xn, yn])
```

Modo 2 (recomendado): introduciendo separadamente  $x$  e  $y$ , y creando después la matriz  $z$

```
x : [x1, x2, ..., xn]; y : [y1, y2, ..., yn]
z : transpose(matrix(x, y))
```

**Representación gráfica:** para el diagrama de puntos

```
draw2d(points(z))
```

Algunas opciones `draw2d(point_type = 7, xrange = [x0, x1], yrange = ...)`

**Recta de regresión:** cargando el paquete `stats`, se calcula directamente con

```
simple_linear_regression(z)
```

que devuelve la ecuación de la recta  $Y = a + bX$ , el coeficiente de correlación  $r$ , y alguna información adicional (ver el ejemplo).

**Gráfica conjunta de diagrama de puntos y recta de regresión:**

```
wxdraw2d(points(z), explicit(a + bX, X, p, q))
```

donde ' $a + bX$ ' es la ecuación de la recta, y  $[p, q]$  un rango **adecuado** en el eje  $x$ .

**Media y desviación típica:** se pueden aplicar a la matriz  $z$

```
mean(z) devuelve [x̄, ȳ]
std(z) devuelve [σx, σy]
```

**Covarianza y correlación**

```
cov(z) devuelve  $\begin{bmatrix} \sigma_x^2 & \text{COV}_{x,y} \\ \text{COV}_{x,y} & \sigma_y^2 \end{bmatrix}$ ; cor(z) devuelve  $\begin{bmatrix} 1 & r_{x,y} \\ r_{x,y} & 1 \end{bmatrix}$ 
```

**NOTAS:**

1.- `wxdraw2d` permite añadir algunos adornos, por ejemplo

```
wxdraw2d(explicit(3+2*X,X,-2,2),title="ejerc 2",xlabel="peso",ylabel="altura")
```

Ver más sobre decoración en <http://riotorto.users.sourceforge.net/gnuplot/func2d>

**Ejemplo:** En el experimento de Galton se estudia cómo afecta el tamaño de una semilla en el tamaño de sus descendientes. Se tienen los datos (diámetro de la semilla en mm)

tamaño padres	15	16	17	18	19	20	21
tamaño hijos	15,4	15,7	16	16,3	16,6	17	17,3

Calcular las medias,  $dt$  y covarianza de las variables. Calcular la recta de regresión, y dibujar la gráfica correspondiente.

**Solución:** comenzamos cargando el paquete *stats* (si fuera necesario)

```
(%i1) load(stats); fpprintprec:5;
(%i2) x:[15,16,17,18,19,20,21]; y:[15.4,15.7,16,16.3,16.6,17,17.3]
(%i3) z:transpose(matrix(x,y))
(%i4) mean(z); std(z)
      [18,16.329] [2, 0.636]
```

Es decir,  $\bar{x} = 18$ ,  $\sigma_x = 2$ ,  $\bar{y} = 16,329$ ,  $\sigma_y = 0,636$

```
(%i5) cov(z); cor(z)
      [ 4 1,2714 ] [ 1 0,999 ]
      [1,2714 0,405] [0,999 1 ]
```

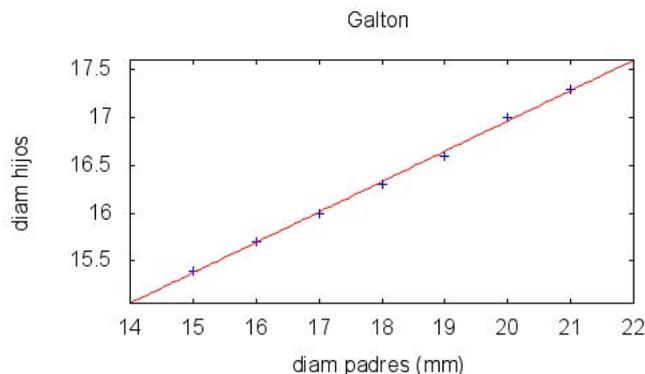
Por tanto concluimos  $cov_{x,y} = 1,2714$  y  $r = 0,999$  (muy buen ajuste).

```
(%i6) simple_linear_regression(z)
```

```
model = 0,318x + 10,607 correlation = 0,999
v_est = 0,00107 b_conf_int = [0,302,0,334]
hypoth H0 : b = 0, H1 : b ≠ 0 statistic=51.384
distribution = [t-Student, 5] p_val = 5,27711 · 10-8
```

Es decir, la recta de regresión es  $Y = 0,318X + 10,607$  y  $r = 0,999$  (el resto de información no la usamos en este curso, pero la explico abajo)

```
(%i7) wxdraw2d(points(z), color=red, explicit(0.318*X+10.607, X, 14, 22),
title="exper Galton", xlabel="diam padres (mm)", ylabel="diam hijos")
```



**NOTA:** los otros comandos que devuelve *simple\_linear\_regression* son

- $v_{\text{est}} = \frac{1}{n-2} \sum_{i=1}^n (y_i - a - bx_i)^2 = \text{ECM}_{n-2}$ , es un estimador insesgado del ECM
- $b_{\text{conf int}}$  = intervalo de confianza (al 95%) para la pendiente  $b$
- *hypoth H0...* test de hipótesis para  $b = 0$ ; cuando  $p_{\text{val}} < 0,001$  se rechaza la hipótesis  $b = 0$ , y se concluye con probabilidad alta que  $b \neq 0$  (y por tanto, que existe dependencia lineal entre  $x$  e  $y$ ).