

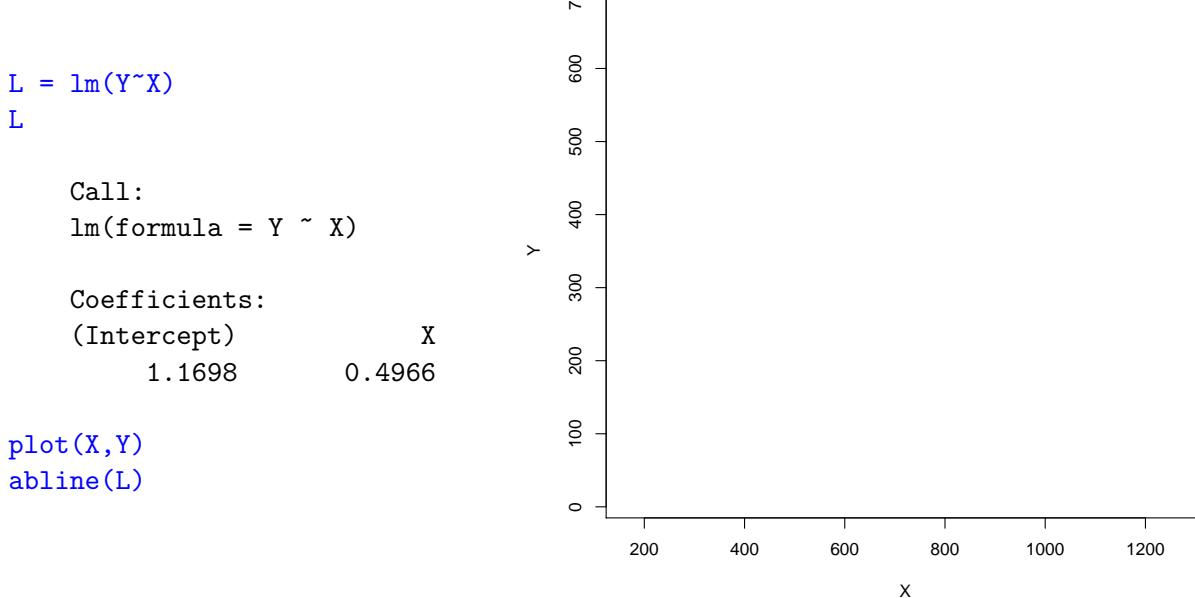
La esteatosis (acumulación anormal de grasa en las células) se da en más de la mitad de los pacientes con infección crónica del virus de la hepatitis C (VHC). Hickman *et al.* (2002)¹ conjeturaron que una reducción de peso en estos pacientes produciría una disminución en el grado de la esteatosis. Para comprobarlo, se sometió a 10 sujetos con VHC crónico a un programa de reducción de peso de tres meses de duración. En cada sujeto se biopsió el hígado antes y después del tratamiento. Entre otros marcadores, en cada biopsia se midió la expresión (en células/mm²) de alfa actina de músculo liso (ACTA-2) en el tracto portal, obteniéndose los siguientes resultados:

	ACTA-2									
Antes	347	412	488	219	168	1273	410	319	209	225
Después	60	363	12	156	199	691	123	239	141	49

- a) En la consola de R escribimos

```
Datos = read.table("ACTA2portal.txt", header=TRUE)
X = Datos$Antes
Y = Datos$Despues
```

Explica qué se hace con el siguiente código de R y lo que se obtiene. Dibuja (aproximadamente) el gráfico resultante en el recuadro.



- b) Obtén el coeficiente de correlación entre la expresión de ACTA-2 antes y después de la dieta a partir de la siguiente información:

```
var(X)
[1] 103632
var(Y)
[1] 39794.9
cov(X,Y)
[1] 51467.22
```

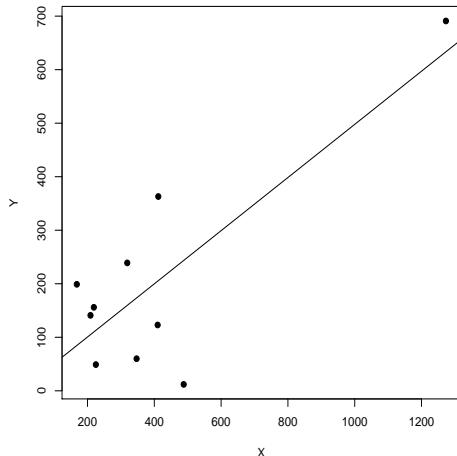
Interpreta el valor obtenido de la correlación ayudándote del gráfico que hayas dibujado en (a).

- c) Sabiendo que $\sum_{i=1}^{10} x_i = 4070$, calcular $\sum_{i=1}^{10} x_i^2$ y $\sum_{i=1}^{10} x_i y_i$ a partir de los resúmenes de los datos dados en (a) y (b).

¹I J Hickman, A D Clouston, G A Macdonald, D M Purdie, J B Prins, S Ash, J R Jonsson, E E Powell (2002). Effect of weight reduction on liver histology and biochemistry in patients with chronic hepatitis C. *Gut*, 51, 89–94.

Solución:

- a) L es un objeto de R que contiene la recta de regresión de Y (nivel de ACTA-2 después del programa) sobre X (nivel de ACTA-2 antes del programa), resultando $\hat{y} = 1.1698 + 0.4966 x$. Con `plot(X, Y)` se dibuja el diagrama de dispersión de Y en función de X y con `abline(L)` se dibuja la recta de regresión.



- b) $r = 51467.22/\sqrt{103632 \cdot 39794.9} = 0.8014$. En este caso, la correlación toma un valor artificialmente alto, pues el diagrama de dispersión no muestra tal relación lineal entre x e y . Esto es debido al dato atípico correspondiente al valor máximo de X . De hecho, si retiramos ese punto, la correlación pasa a ser -0.063.

c)

$$\bar{y} = a + b\bar{x} = 1.1698 + 0.4966 \cdot 407 = 203.3$$

$$\sum_{i=1}^{10} x_i^2 = (10 - 1)s_x^2 + 10\bar{x}^2 = 9 \cdot 103632 + 10 \cdot 407^2 = 2589178$$

$$\sum_{i=1}^{10} x_i y_i = (10 - 1)\text{cov}_{xy} + 10\bar{x}\bar{y} = 9 \cdot 51467.22 + 10 \cdot 407 \cdot 203.3 = 2589178$$