

Tarea 1

Karen Delgado Curiel 142252, Bernardo Mondragón Brozon 143743, Juan Casas Alatríste Rion 141913, David Almog Salama Finkelstein 136731

08/02/2018

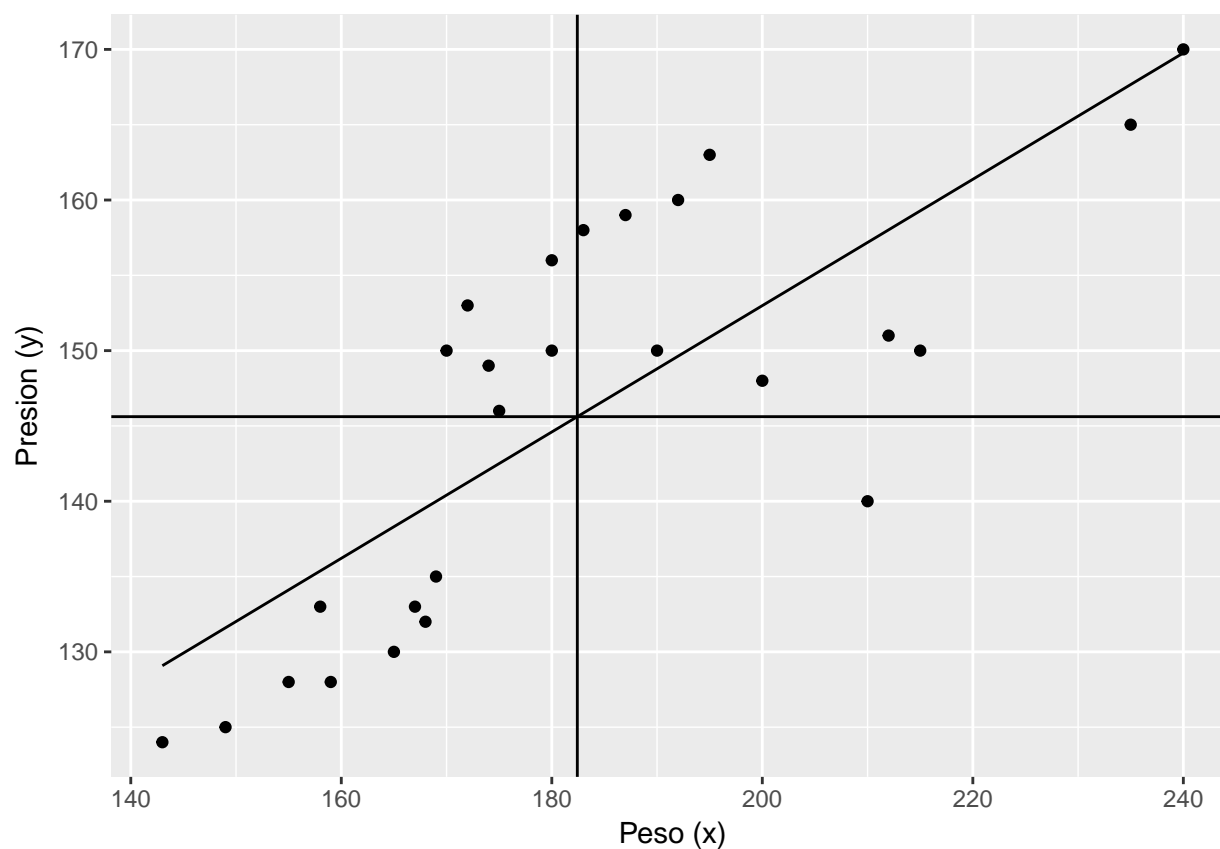
Los datos

Se cuenta con el peso, en libras, y la presión sistólica de 26 personas. En este análisis, la presión será la variable de respuesta y el peso será la variable regresora o predictora. A continuación se muestra el aspecto que tiene la información de las 26 personas.

```
##  peso presion
## 1  165    130
## 2  167    133
## 3  180    150
## 4  155    128
## 5  212    151
## 6  175    146
```

Regresión lineal simple

A continuación se muestra un diagrama de dispersión y la recta que mejor se ajusta a los datos.



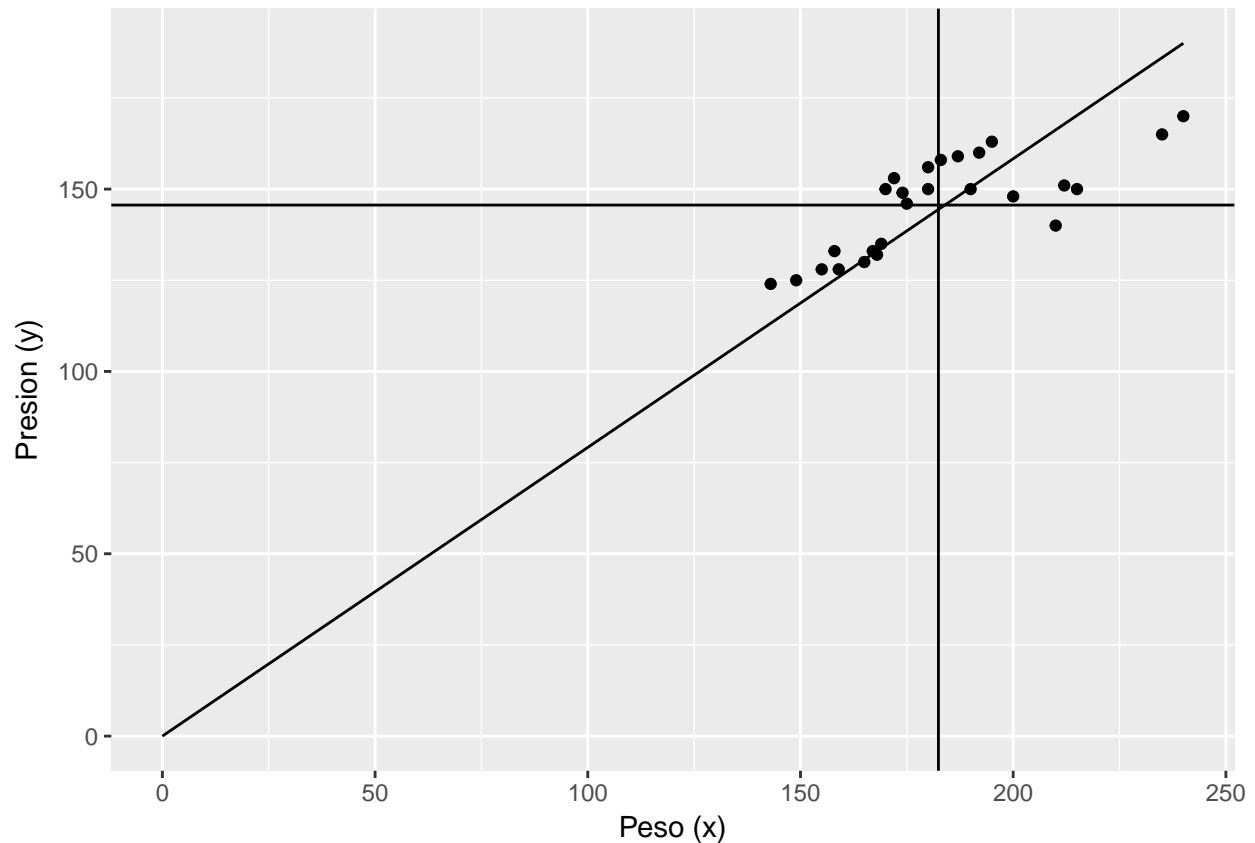
La recta de regresión está dada por

$$y = 0.4194152x + 69.1043728$$

y el punto en donde la recta horizontal intersecta con la recta vertical es el centroide (\bar{x}, \bar{y}) .

Regresión lineal por el origen

A continuación se muestra un diagrama de dispersión y la recta de regresión lineal por el origen que mejor se ajusta a los datos.



La recta de regresión lineal por el origen está dada por la siguiente ecuación:

$$y = 0.7916415x.$$

Note que en el modelo de regresión por el origen la recta de regresión no pasa por el centroide (\bar{x}, \bar{y}) . En el modelo de regresión lineal simple se calculó la recta que mejor se ajusta a los datos, sin embargo, tiene sentido pensar en un modelo de regresión lineal por el origen, pues es imposible que una persona cuyo peso es cero tenga una presión sistólica positiva e igual 69.1043728. Por otro lado, el modelo de regresión lineal simple es mejor porque produce estimadores insesgados.

Intervalos de predicción

El intervalo de predicción para una persona que 122lb es el siguiente:

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm \tau_{n-2, \frac{\alpha}{2}} \sqrt{\frac{\sum_{i=1}^n (y_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i])^2}{n-2} \left(1 + \frac{1}{n} + \frac{x - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)} = (102.0500556, 138.4959994)$$

Problema teórico

Demuestre que en regresión múltiple los valores ajustados y los intervalos de predicción no se alteran si la matriz de diseño se post-multiplica por una matriz S no singular. Diga como sería esta matriz en el caso de regresión simple en donde el regresor X está medido en pulgadas y se quiere que el resultado de la postmultiplicación sea que X se mide ahora en cm.

Demostración

Sea X la matriz de diseño y sea S una matriz no singular. Los valores ajustados del modelo $Y = X * S * \beta$ están dados por

$$\begin{aligned}\hat{Y}_0 &= (XS)((XS)^T(XS)^{-1})(XS)^T Y \\ &= XS(S^T X^T XS)^{-1} S^T XY \\ &= X(X^T X)^{-1} XY \\ &= HY\end{aligned}$$

Por lo tanto los valores ajustados siguen siendo los mismos. Del mismo modo se tiene que el intervalo

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm \tau_{n-2, \frac{\alpha}{2}} \sqrt{MS_{res}(1 + Sx_0'((SX)'SX)^{-1}Sx_0)}$$

coincide con el intervalo

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm \tau_{n-2, \frac{\alpha}{2}} \sqrt{MS_{res}(1 + x_0'(X'X)^{-1}x_0)}.$$

Si se quiere que la variable regresora X ahora esté medida en centímetros (cm) en lugar de pulgadas ($''$), entonces la matriz no singular S tendría el siguiente aspecto:

$$S = \begin{bmatrix} 1 & 0 \\ 0 & 2.54 \end{bmatrix} \cdot \blacksquare$$