Example of simple linear regression in matrix form

An auto part is manufactured by a company once a month in lots that vary in size as demand fluctuates. The data below represent observations on lot size (y), and number of man-hours of labor (x) for 10 recent production runs. Suppose that you need to fit the simple regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where $E(\epsilon_i) = 0$, $E(\epsilon_i \epsilon_j) = 0$ for $i \neq j$, and $\text{var}(\epsilon_i) = \sigma^2$. In vector form the data are:

$$\mathbf{Y} = \begin{pmatrix} 73 \\ 50 \\ 128 \\ 170 \\ 87 \\ 108 \\ 135 \\ 69 \\ 148 \\ 132 \end{pmatrix} \text{ and } \mathbf{X} = \begin{pmatrix} 1 & 30 \\ 1 & 20 \\ 1 & 60 \\ 1 & 80 \\ 1 & 40 \\ 1 & 50 \\ 1 & 60 \\ 1 & 30 \\ 1 & 70 \\ 1 & 60 \end{pmatrix}.$$

It turns out that

$$\boldsymbol{(X'X)^{-1}} = \begin{pmatrix} 0.83529412 & -0.01470588 \\ -0.01470588 & 0.00029412 \end{pmatrix}, \ \boldsymbol{X'Y} = \begin{pmatrix} 1100 \\ 61800 \end{pmatrix}, \text{ and } \mathbf{Y'Y} = \mathbf{134660}.$$

a. Find the least squares estimator of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$.

b. Find the variance-covariance matrix of the previous estimator.

c. Compute the estimate s_e^2 of σ^2.

d. Using your answers to parts (b) and (c) find the variances of $\hat{\beta}_0, \hat{\beta}_1$.

e. Find the fitted value $\hat{y}_1$ and its variance.

f. What is the variance of the first residual ($\text{var}(e_1)$)?

Answers:

a. $$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{pmatrix} 0.83529412 & -0.01470588 \\ -0.01470588 & 0.00029412 \end{pmatrix} \begin{pmatrix} 1100 \\ 61800 \end{pmatrix} = \begin{pmatrix} 10.0 \\ 2.0 \end{pmatrix}.$$

Therefore $\hat{\beta}_0 = 10.0$, and $\hat{\beta}_1 = 2.0$.

b. The variance-covariance matrix $\hat{\beta}$ is given by:

$$\text{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 \begin{pmatrix} 0.83529412 & -0.01470588 \\ -0.01470588 & 0.00029412 \end{pmatrix}.$$

Of course σ^2 is unknown and needs to be estimated (next question).

c. We estimate σ^2 with $s_e^2 = \frac{SSE}{n-k-1}$. The error sum of squares in matrix form is given by:

$$SSE = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} = 134660 - \begin{pmatrix} 10.0 & 2.0 \end{pmatrix} \begin{pmatrix} 1100 \\ 61800 \end{pmatrix} = 134660 - 134600 = 60.$$

Therefore $s_e^2 = \frac{60}{8} = 7.5$.

d. Now, using part (b) and (c) we can find the variance of $\hat{\beta}_0$, and $\hat{\beta}_1$ as follows:

$$\text{var}(\hat{\beta}_0) = 7.5(0.83529412) = 6.2647.$$
$$\text{var}(\hat{\beta}_1) = 7.5(0.00029412) = 0.0022.$$

e. The first fitted value $\hat{y}_1$ is:

$$\hat{y}_1 = 10.0 + 2.0(30) = 70.$$

To find the variance of the fitted values we need to compute the variance-covariance matrix of $\hat{\mathbf{Y}}$.

$$\text{cov}(\hat{\mathbf{Y}}) = \sigma^2\mathbf{H}, \text{ where } \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \text{ is the hat matrix.}$$

The hat matrix is an $n \times n$ matrix, here 10×10. We are not going to compute all the 100 elements of this matrix. We only need the element of the hat matrix that corresponds to the first row and first column, that is h_{11}. We need this because the variance of the first fitted value is $\text{var}(\hat{y}_1) = \sigma^2 h_{11}$. The value of h_{11} is computed as follows:

$$h_{11} = \mathbf{x}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_1 = \begin{pmatrix} 1 & 30 \end{pmatrix} \begin{pmatrix} 0.83529412 & -0.01470588 \\ -0.01470588 & 0.00029412 \end{pmatrix} \begin{pmatrix} 1 \\ 30 \end{pmatrix} = 0.218.$$

Therefore $\text{var}(\hat{y}_1) = \sigma^2 h_{11} = \sigma^2(0.218)$. It is estimated after we replace σ^2 by s_e^2 as: $\text{var}(\hat{y}_1) = 7.5(0.218) = 1.635$.

f. The variance-covariance matrix of the residuals is given by:

$$\text{cov}(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H}).$$

Therefore the variance of the first residual is:

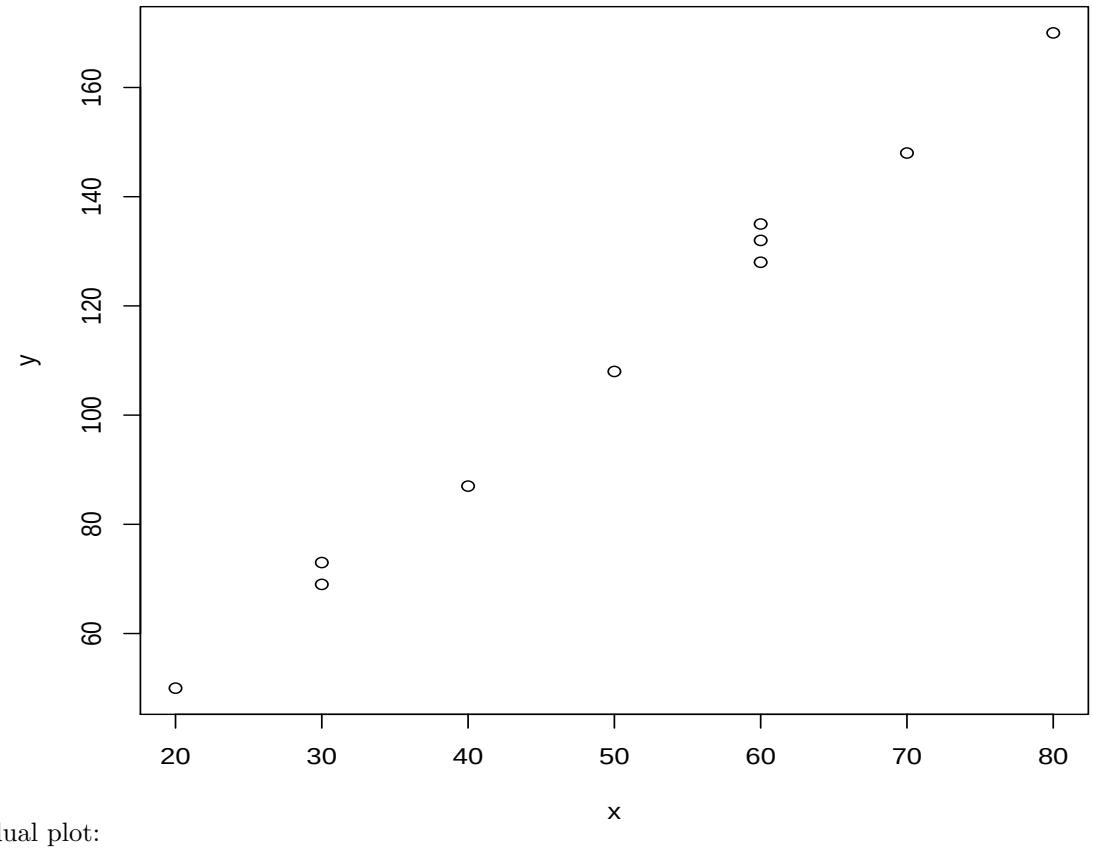
$$\text{var}(e_1) = \sigma^2(1 - h_{11}) = \sigma^2(1 - 0.218) = \sigma^2(0.782).$$

It is estimated as $\text{var}(e_1) = 7.5(0.782) = 5.865$.

Some simple R commands:

```
> y <- c(73,50,128,170,87,108,135,69,148,132)

> x <- c(30,20,60,80,40,50,60,30,70,60)

> ones <- rep(1,10)

> X <- as.matrix(cbind(ones,x))

> X
      ones  x
 [1,]    1 30
 [2,]    1 20
 [3,]    1 60
 [4,]    1 80
 [5,]    1 40
 [6,]    1 50
 [7,]    1 60
 [8,]    1 30
 [9,]    1 70
[10,]    1 60

> class(X)
[1] "matrix"

> t(X)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
ones    1    1    1    1    1    1    1    1    1     1
x      30   20   60   80   40   50   60   30   70    60

> t(X) %*% X
     ones     x
ones   10   500
x     500 28400

> solve(t(X) %*% X)
            ones             x
ones  0.83529412 -0.0147058824
x    -0.01470588  0.0002941176


> beta_hat <- solve(t(X) %*% X) %*% t(X) %*% y
> beta_hat
     [,1]
ones   10
x       2

> H <- X %*% solve(t(X) %*% X) %*% t(X)
> round(H, digits=3)
        [,1]   [,2]  [,3]   [,4]  [,5] [,6]  [,7]   [,8]   [,9] [,10]
 [1,]  0.218  0.276 0.041 -0.076 0.159  0.1 0.041  0.218 -0.018 0.041
 [2,]  0.276  0.365 0.012 -0.165 0.188  0.1 0.012  0.276 -0.076 0.012
 [3,]  0.041  0.012 0.129  0.188 0.071  0.1 0.129  0.041  0.159 0.129
 [4,] -0.076 -0.165 0.188  0.365 0.012  0.1 0.188 -0.076  0.276 0.188
 [5,]  0.159  0.188 0.071  0.012 0.129  0.1 0.071  0.159  0.041 0.071
 [6,]  0.100  0.100 0.100  0.100 0.100  0.1 0.100  0.100  0.100 0.100
 [7,]  0.041  0.012 0.129  0.188 0.071  0.1 0.129  0.041  0.159 0.129
 [8,]  0.218  0.276 0.041 -0.076 0.159  0.1 0.041  0.218 -0.018 0.041
 [9,] -0.018 -0.076 0.159  0.276 0.041  0.1 0.159 -0.018  0.218 0.159
[10,]  0.041  0.012 0.129  0.188 0.071  0.1 0.129  0.041  0.159 0.129
```

Scatterplot of y on x:



Residual plot: