Statistics 108, Homework 4

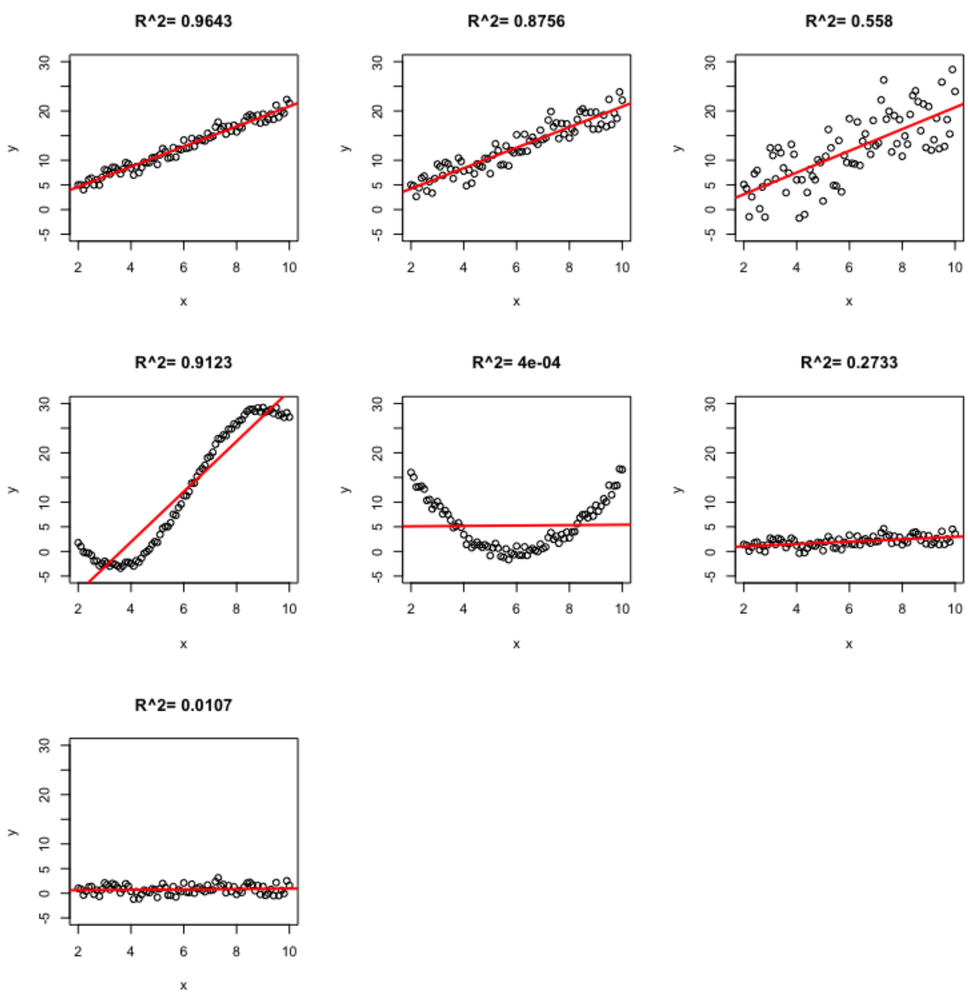
Due: November 2nd, 2018, In Class (turn in paper form)

*You need to show the steps to get the full credits.

This homework is to transit from simple linear regression to multiple linear regression. Total: 90 points.

1. (40 points) *Understanding the coefficient of determination, R^2.*

   (a) *Visualization.* For each of the seven datasets under Canvas/Files/Datasets (data1.txt – data7.txt), plot the scatter plot, add the fitted linear regression line, and compute the R^2. What do you find? (Plot each dataset on the same y-scale to compare across the datasets.)



We can see not all datasets can be fitted by a linear regression model well, and the scales of response are different. As for data1 and data2, a simple linear regression model is good enough based on their high R^2. We can also find that R^2 is not always a good criterion for goodness of fit. As for data4, the response and predictor is not linearly associated, but the R^2 is still decent.

(b) *True/False.* For each of the following statement, say whether it is true or false and explain why.

(i) A large R^2 always means that the fitted linear regression line is a good fit of the data.
False. Please see the fourth figure above. R^2 is high but the best fit of the relationship between X and Y is clearly nonlinear. The linear fit will underestimate at certain points in the domain and overestimate at others.

(ii) A small R^2 always means that the predictor and the response are not related.
False. Nonlinear relationships can have low R^2 values. Please see the fifth figure above.

(iii) If all observations Y_i fall on on straight line and the line is not horizontal, then $R^2 = 1$.
True. If the points fall on a line, then $\hat{Y}_i = Y_i$, and so then

$$\text{SSR} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \text{SSTO}$$

and SSR = SSTO $\neq 0$ because the line is not horizontal. Consequently,

$$\text{R}^2 = \frac{\text{SSR}}{\text{SSTO}} = 1$$

2. (40 points) *Simple linear regression in matrix form.* Here, we analyze the data in Homework 1, problem 2, again, but using the matrix form.

(a) What are the dimensions of the response vector $\mathbf{Y}$ and the design matrix $\mathbf{X}$. Write down the first five rows of them.

Solution:

$\mathbf{Y}$ is a $n \times 1$ colunum vector, and $\mathbf{X}$ is a $n \times 2$ matrix. Here $n = 150$.

The first 5 rows of $\mathbf{Y}$ is

$$\begin{bmatrix} 83.0 \\ 66.5 \\ 73.7 \\ 74.8 \\ 78.0 \end{bmatrix}$$

And the first 5 rows of $\mathbf{X}$ is

$$\begin{bmatrix} 1 & 180.1 \\ 1 & 176.1 \\ 1 & 178.5 \\ 1 & 183.9 \\ 1 & 171.7 \end{bmatrix}$$

(b) Calculate the following two quantities: $\mathbf{X}^T\mathbf{X}, \mathbf{X}^T\mathbf{Y}$.
Solution:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 150 & 26912.3 \\ 26912.3 & 4832921.6 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 11442.7 \\ 2056544.0 \end{bmatrix}$$

(c) Calculate the least squares estimators by

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}).$$

Compare the results here with those from that in Homework 1. Are they the same?
Solution:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}) = \begin{bmatrix} -67.0169636 \\ 0.7987145 \end{bmatrix}$$

This results are the same as those in Homework 1. The reason is presented below. Let $\mathbf{1}_n = (1, ..., 1)^T$ be a $n \times 1$ vector, and $\mathbf{X}_1 = (X_1, ..., X_n)^T$ is a $n \times 1$ vector. Then

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_n & \mathbf{X}_1 \end{bmatrix}$$

So

$$\begin{aligned} \mathbf{X}^T\mathbf{X} &= \begin{bmatrix} \mathbf{1}_n^T \\ \mathbf{X}_1^T \end{bmatrix} \begin{bmatrix} \mathbf{1}_n & \mathbf{X}_1 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{1}_n^T\mathbf{1}_n & \mathbf{1}_n^T\mathbf{X}_1 \\ \mathbf{X}_1^T\mathbf{1}_n & \mathbf{X}_1^T\mathbf{X}_1 \end{bmatrix} \\ &= \begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{bmatrix} \end{aligned}$$

By Problem 3,

$$\begin{aligned} (\mathbf{X}^T\mathbf{X})^{-1} &= \frac{1}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix} \\ &= \frac{1}{n\sum_{i=1}^n (X_i - \bar{X})^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix} \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{X}^T\mathbf{Y} &= \begin{bmatrix} \mathbf{1}_n^T \\ \mathbf{X}_1^T \end{bmatrix} \mathbf{Y} \\ &= \begin{bmatrix} \mathbf{1}_n^T\mathbf{Y} \\ \mathbf{X}_1^T\mathbf{Y} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_iY_i \end{bmatrix} \end{aligned}$$

Therefore,

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}) \\
&= \frac{1}{n\sum_{i=1}^n (X_i-\bar{X})^2}\begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix}\begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_iY_i \end{bmatrix} \\
&= \frac{1}{n\sum_{i=1}^n (X_i-\bar{X})^2}\begin{bmatrix} \sum_{i=1}^n X_i^2 \sum_{i=1}^n Y_i - \sum_{i=1}^n X_iY_i \sum_{i=1}^n X_i \\ n\sum_{i=1}^n X_iY_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i \end{bmatrix} \\
&= \frac{1}{nS_{XX}}\begin{bmatrix} n(\bar{Y}\sum_{i=1}^n X_i^2 - \bar{X}\sum_{i=1}^n X_iY_i) \\ n(\sum_{i=1}^n X_iY_i - n\bar{X}\bar{Y}) \end{bmatrix} \\
&= \frac{1}{nS_{XX}}\begin{bmatrix} n[(\bar{Y}\sum_{i=1}^n X_i^2 - n\bar{X}^2\bar{Y}) - (\bar{X}\sum_{i=1}^n X_iY_i - n\bar{X}^2\bar{Y})] \\ n(\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})) \end{bmatrix} \\
&= \frac{1}{nS_{XX}}\begin{bmatrix} n[(\bar{Y}\sum_{i=1}^n X_i^2 - n\bar{X}^2\bar{Y}) - (\bar{X}\sum_{i=1}^n X_iY_i - n\bar{X}^2\bar{Y})] \\ n(\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})) \end{bmatrix} \\
&= \frac{1}{nS_{XX}}\begin{bmatrix} n[(\bar{Y}S_{XX}) - (\bar{X}\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y}))] \\ n(\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})) \end{bmatrix} \\
&= \begin{bmatrix} \bar{Y} - \frac{\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})}{S_{XX}}\bar{X} \\ \frac{\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})}{S_{XX}} \end{bmatrix}
\end{aligned}
$$

which is exactly the least square estimator for simple linear regression.

(d) Give an estimate of the variance of $\hat{\boldsymbol{\beta}}$. Based on this, what are the standard error of $\hat{\beta}_0$, the standard error of $\hat{\beta}_1$, and the estimate of $Cov(\hat{\beta}_0, \hat{\beta}_1)$? Compare these with Homework 2, problem 2, what do you find?

Solution:

$$
\begin{aligned}
Var(\hat{\boldsymbol{\beta}}) &= Var((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Var(Y)\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \sigma^2\mathbf{I}_n\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}
\end{aligned}
$$

As

$$
\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 168.594203 & -0.93882297 \\ -0.93882297 & 0.00523268 \end{bmatrix}
$$

We have

$$
\begin{aligned}
se(\hat{\beta}_0) &= \sqrt{168.594203} \\
&= 12.98438
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
se(\hat{\beta}_1) &= \sqrt{0.00523268} \\
&= 0.07233726
\end{aligned}
$$

And the estimate of $Cov(\hat{\beta}_0, \hat{\beta}_1)$ is -0.93882297.

The standard errors calculated from the matrix form are the same as those from the ordinary form. Because

$$
\begin{aligned}
Var(\hat{\boldsymbol{\beta}}) &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \frac{\sigma^2}{n\sum_{i=1}^n (X_i - \bar{X})^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} \frac{\sum_{i=1}^n X_i^2}{nS_{XX}} & -\frac{\bar{X}}{S_{XX}} \\ -\frac{\bar{X}}{S_{XX}} & \frac{1}{S_{XX}} \end{bmatrix}
\end{aligned}
$$

Then

$$
\begin{aligned}
Var(\hat{\beta}_0) &= \sigma^2 \frac{\sum_{i=1}^n X_i^2}{nS_{XX}} \\
&= \sigma^2 \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2 + n\bar{X}^2}{nS_{XX}} \\
&= \sigma^2 (\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}})
\end{aligned}
$$

and

$$
Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{XX}}
$$

The R code for Problem 2 is below:

```r
data_wh = read.table("weight_full.txt", header=TRUE)

#(a)
n = dim(data_wh)[1]

Y = data_wh$weight
X = cbind(rep(1,n),data_wh$height)
Y[1:5]
X[1:5 ,]

#(b)
t(X) %*% X
```

```
t(X) %*% Y

#(c)
solve(t(X) %*% X) %*% t(X) %*% Y
beta_0_hat = solve(t(X) %*% X) %*% t(X) %*% Y[1,1]
beta_1_hat = solve(t(X) %*% X) %*% t(X) %*% Y[2,1]

#(d)
MSE = sum((Y - X %*% (solve(t(X) %*% X)) %*% t(X) %*% Y)^2)/(n-2)
MSE*solve(t(X) %*% X)

se_beta_0_hat = sqrt(MSE*solve(t(X) %*% X)[1,1])
se_beta_1_hat = sqrt(MSE*solve(t(X) %*% X)[2,2])
```

3. (10 points) *Rigorous derivations.* Suppose

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and $ad - bc \neq 0$. Let

$$\mathbf{B} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Show that

$$\mathbf{BA} = \mathbf{AB} = \mathbf{I}_2.$$

Therefore, by definition $\mathbf{B}$ is the inverse of $\mathbf{A}$, i.e., $\mathbf{B} = \mathbf{A}^{-1}$.

Solution:

$$\begin{aligned}
\mathbf{BA} &= \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} \\
&= \frac{1}{ad-bc}\begin{bmatrix} da-bc & db-bd \\ -ca+ac & -cb+ad \end{bmatrix} \\
&= \frac{1}{ad-bc}\begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \mathbf{I}_2
\end{aligned}$$

$$
\begin{aligned}
\mathbf{AB} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \\
&= \frac{1}{ad-bc} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \\
&= \frac{1}{ad-bc} \begin{bmatrix} ad-bc & a \times (-b) + ba \\ cd + d \times (-c) & c \times (-b) + da \end{bmatrix} \\
&= \frac{1}{ad-bc} \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} \\
&= \mathbf{I}_2
\end{aligned}
$$