R-Assignment-8

Include the R-code in the following questions.

1. The data given in Table8-1 shows the average public teacher pay and spending on public schools per pupil in 1985 for 50 states and the District of Columbia. The data is also provided in an excel-file called Data8-1. Read in the data from Data8-1 using the **read.delim** command in R. (See examples in lecture notes from 4.4 or 4.5).

Source: National Education Association, as reported by the Albuquerque Tribune, 11/7/86 at DASL at http://lib.stat.cmu.edu/DASL/Datafiles/teacherpaydat.html.

(A) Plot y= the average public teacher pay against x= spending on public school, using plot(x,y). Label your x-axis as Spend and the y-axis as Pay using the commands xlab="Spend" and ylab="Pay".

(B) Use R to calculate the regression line when the spending on public school is the explanatory variable, x, and the average public teacher pay is the response variable, y. The command in R is $lm(y \sim x)$
Add the regression line to the plot in part (A) using the command $abline(lm(y \sim x))$.

(C) Use the **summary** command in R, that is type in $summary(lm(y \sim x))$. Test the hypothesis

$$H_0 : \beta = 0 \quad \text{against} \quad H_1 : \beta > 0.$$

What is your conclusion. Is there a linear relationship between y and x?

(D) What is the correlation coefficient?

(E) What fraction of the variation in average public teacher pay can be explained by the spending on public school, using the model found in part (B)?

(F) Find the predicted average public teacher pay when the spending on public school is $3500.

(G) Use the R-command $rstandard(lm(y \sim x))$ to standardize the residuals. Draw a plot of the residuals against x, using the command, $plot(x, rstandard(lm(y \sim x)))$. Label your x-axis as Spend and the y-axis as Residuals. Are there any potential influential observations or leverage points?

(H) Is the linear regression model above a reasonable model to use to predict the average public teacher pay?

(I) Use R to find the separate regression lines for the three regions of the country? Does the slope change? Check if the linear regression models obtained are reasonable models by looking at the summary statistics and analyzing the residual plots.