Problem 1

Consider the following dataset:

Y1	X1	Y2	X2	Y3	X3	Y4	X4
8.04	10	9.14	10	7.46	10	6.58	8
6.95	8	8.14	8	6.77	8	5.76	8
7.58	13	8.74	13	12.74	13	7.71	8
8.81	9	8.77	9	7.11	9	8.84	8
8.33	11	9.26	11	7.81	11	8.47	8
9.96	14	8.10	14	8.84	14	7.04	8
7.24	6	6.13	6	6.08	6	5.25	8
4.26	4	3.10	4	5.39	4	12.50	19
10.84	12	9.13	12	8.15	12	5.56	8
4.82	7	7.26	7	6.42	7	7.91	8
5.68	5	4.74	5	5.73	5	6.89	8

For the above dataset:

a. Compute correlation between Yi and Xi where i = 1, 2,3,4. Comment on it.
b. Fit the simple linear regression for each pair of Yi and Xi (i = 1, 2,3,4) and find the following:
   i. $\widehat{\beta_0}\ and\ \widehat{\beta_1}$
   ii. R^2
   iii. $t - test$
c. Comment on the results which you get.

Problem 2

In order to investigate the feasibility of starting a Sunday edition for a large metropolitan newspaper, information was obtained from a sample of 34 newspapers concerning their daily and Sunday circulations (in thousands) (Source: Gale Directory of Publications, 1994)

(a) Construct a scatter plot of Sunday circulation versus daily circulation. Does the plot suggest a linear relationship between Daily and Sunday circulation? Do you think this is a plausible relationship?

(b) Fit a regression line predicting Sunday circulation from Daily circulation.

(c) Obtain the 95% confidence intervals for β_0 and β_1.

(d) Is there a significant relationship between Sunday circulation and Daily circulation? Justify your answer by a statistical test. Indicate what hypothesis you are testing and your conclusion.

(e) What proportion of the variability in Sunday circulation is accounted for by Daily circulation?

(f) Provide an interval estimate (based on 95% level) for the **true average** Sunday circulation of newspapers with Daily circulation of 500,000.

(g) The particular newspaper that is considering a Sunday edition has a Daily circulation of 500,000. **Provide an interval estimate (based on 95% level) for the predicted Sunday circulation of this paper.** How does this interval differ from that given in (f)?

(h) Another newspaper being considered as a candidate for a Sunday edition has a Daily circulation of 2,000,000. Provide an interval estimate for the predicted Sunday circulation for this paper? How does this interval compare with the one given in (g)? Do you think it is likely to be accurate?

Newspaper	Daily	Sunday
1	391.952	48 8.506
2	516.981	798.298
3	355.628	235.084
4	238.555	299.451
5	537.78	559.093
6	733.775	1133.249
7	198.832	348.744
8	252.624	417.779
9	206.204	344.522
10	231.177	323.084
11	449.755	620.752
12	288.571	423.305
13	185.736	202.614
14	1164.388	1531.527

15	444.581	553.479
16	412.871	685.975
17	272.28	324.241
18	781.796	983.24
19	1209.225	1762.015
20	825.512	960.308
21	223.748	284.611
22	354.843	407.76
23	515.523	982.663
24	220.465	557
25	337.672	440.923
26	197.12	268.06
27	133.239	262.048
28	374.009	432.502
29	273.844	338.355
30	570.364	704.322
31	391.286	585.681
32	201.86	267.781
33	321.626	408.343
34	838.902	1165.567

Problem 3

Consider the simple linear regression model $y = 50 + 10\,x + \varepsilon$ where ε is NID $(0, 16)$. Suppose that $n = 20$ pairs of observations are used to fit this model. Generate 500 samples of 20 observations, drawing one observation for each level of $x = 1, 1.5, 2, \ldots,$ 10 for each sample.

a. For each sample computes the least - squares estimates of the slope and intercept. Construct histograms of the sample values of $\widehat{\beta_0}$ and $\widehat{\beta_1}$. Discuss the shape of these histograms.

b. For each sample, compute an estimate of $E(y \mid x = 5)$. Construct a histogram of the estimates you obtained. Discuss the shape of the histogram.

c. For each sample, compute a 95% CI on the slope. How many of these intervals contain the true value $\beta_1 = 10$? Is this what you would expect?

d. For each estimate of $E(y \mid x = 5)$ in part b, compute the 95% CI. How many of these intervals contain the true value of $E(y \mid x = 5) = 100$? Is this what you would expect?