Show your work to receive full credit.

Chapters covered : 3, 12

Problem 1

Let x be students' verbal SAT score and Y be the math SAT score. From a large sample, we found that $\bar{x} = 480$, $\bar{y} = 500$, $s_x = 80$, $s_y = 120$, and correlation between x and y is $r = 0.6$.

a) Find the slope (b) of the regression line.

b) Find the y-intercept (a) of the regression line.

c) Based on your answers from parts a) and b), write down the equation of least squares regression (or prediction equation).

d) Find r^2 and interpret.

Problem 2: Multiple choice

The slope (b) of the regression equation and the correlation (r) are similar in the sense that:

(A) they do not depend on the unit of measurement.

(B) they both must fall between -1 and 1.

(C) they both remain the same when the role of x and y are switched.

(D) they both have the same sign.

Problem 3

Data set `crime2055` contains 51 observations (50 states and Washington D.C.) and two variables:

- HS: percentage of adult residents with a high school diploma in a given state

- PO: Percentage below the poverty level

Use the following R command and output to answer questions.

```
> summary(lm(PO~HS))
```

```
Call:
lm(formula = PO ~ HS)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.5494 -1.5402 -0.3753  0.9836  8.9562
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 61.80106    7.65250   8.076 1.46e-10 ***
HS          -0.60483    0.08843  -6.840 1.18e-08 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.261 on 49 degrees of freedom
Multiple R-squared:  0.4884,Adjusted R-squared:  0.478
F-statistic: 46.78 on 1 and 49 DF,  p-value: 1.176e-08
```

Let $\mu_Y = \alpha + \beta x$ be the population regression equation.

a) Write down the estimate regression equation.

b) Interpret both y-intercept and slope in context. Comment whether each interpretation is realistic or not. Scatter plot of x and y are available at the end of this assignment for your information.

c) Identify r^2 and r. No interpretation required.

d) Construct a 95% confidence interval for β and interpret the result.

e) Conduct a five step hypothesis test on whether the true slope β is significantly different from 0. Use $\alpha = 0.05$. Use the following three diagnostic plots to check assumptions. Interpret the conclusion in the context of the problem.
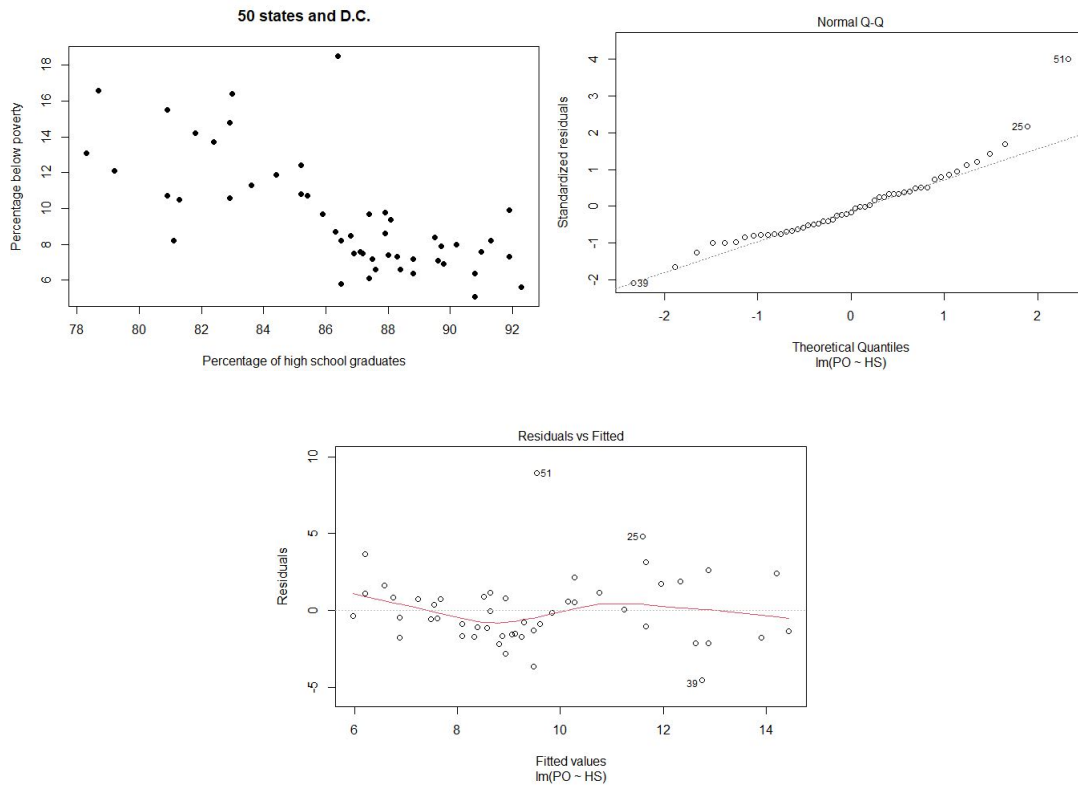
Problem 4

The scatterplots below show the relationship between height, diameter, and volume of timber in 31 felled black cherry trees. The diameter of the tree is measured 4.5 feet above the ground.
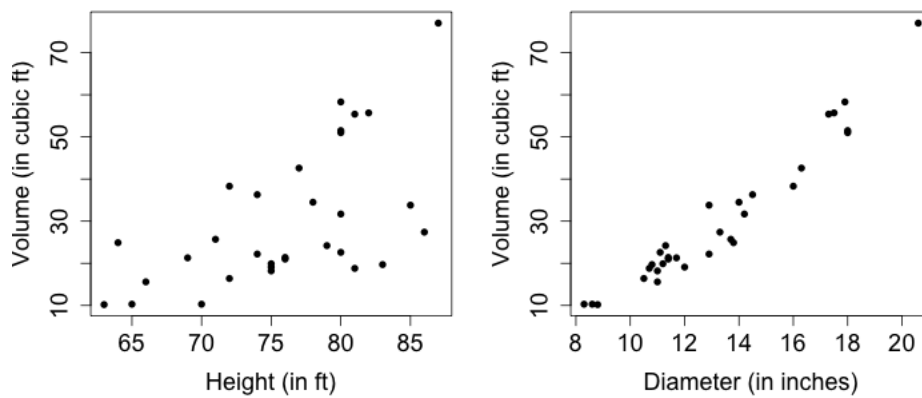
a) Describe the relationship between the volume and height of these trees (in terms of the overall pattern, strength, and direction).

b) Describe the relationship between the volume and diameter of these trees (in terms of the overall pattern, strength, and direction).

c) Suppose you have height and diameter measurements for another black cherry tree. Which of these variables would you prefer to use to predict the volume of timber in this tree using a simple linear regression model? Explain your reasoning.

Problem 5

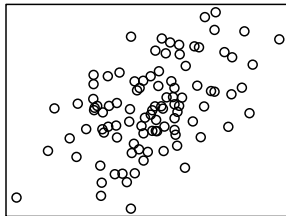Match the calculated correlations to the corresponding scatterplot (**no explanations needed**).
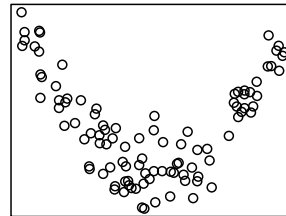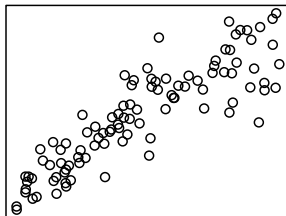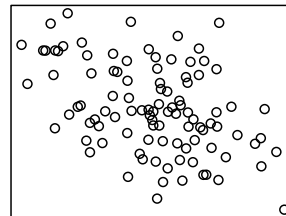
Figure 1

Figure 2

Figure 3

Figure 4

a) $r = -0.43$

b) $r = -0.04$

c) $r = 0.49$

d) $r = 0.88$

R Problem

Data set named *Rotten Tomatoes Reviews* is available at https://artofstat.com/datasets Chapter 3: H -T Import the data using the following R command. Be sure to put the url address in one line before running R command.

```
##Warning: Make sure to put url address in one line in R script / console ##
review<-read.csv("https://img1.wsimg.com/blobby/go/bbca5dba-4947-4587-b40a-db346c01b1b3/downloads
/RottenTomatoes.csv?ver=1668691567467")
```

There are three variables :

- `Movie` : Title of a movie

- `Critic.Score` : On a scale of 100, it represents the percentage of professional critic reviews that are positive.

- `Audience.Score` : On a scale of 100, it represents the percentage of audience reviews that are positive.

where `Critic.Score` is the explanatory variable and `Audience.Score` is the response variable.

a) Construct a scatter plot of Critic.Score and Audience.Score. What is correlation between variables? Describe the association between two variables. Submit R commands and output and scatter plot.

b) Use R command to fit the linear regression model of Audience.Score and Critic.Score. Write down the estimated least square regression equation.

c) *Captain Marvel* has Critic.Score 78 and Audience.Score 55. What is the predicted Audience.Score using the estimate least square regression equation? What is the residual?