Assignment #1 — Due: Monday, October 5, 2020, by 09:00 a.m.

Use R to answer the questions. Solve by hand only when required. Your should include R command and relevant outputs, but not necessarily all outputs.

1. [20 points] Consider the following data set:

X	2	3	3	3	4	5
Y	0	1	1	2	2	3

   (a) Fit the following model (both by R and by hand): (Find $\hat{\beta}_0$ and $\hat{\beta}_1$.) $Y = \beta_0 + \beta_1 X + \varepsilon$.

   (b) Fit the following model (both by R and by hand): (Find $\hat{\beta}_1$.) $Y = \beta_1 X + \varepsilon$.

   (Write some of your working process for calculating the estimated coefficients when solving by hand.)

2. [20 points] We assume our income increases as we get more education. We are interested in investigating how years of education (EDU) relates to income. The data set below shows the EDU (in years) and annual income (in 1000 US dollars) of 12 randomly selected workers.

EDU (years)	12	20	20	14	16	16	18	14	12	16	15	10
Income (1000 USD)	35	80	78	45	57	65	59	63	57	66	73	23

   (a) Plot the data. Explain the relationship between the EDU (in years) and annual income (in 1000 USD).

   (b) Fit a regression line and add it to the plot in (a). Report the fitted regression equation using 'EDU' and 'Income'.

   (c) Compute the covariance and the standard deviations of the two variables.

   (d) Using the sample statistics in (c), calculate the estimated slope (Solve this question by hand).

3. [30 points] The following data represents two measurements (*Length* and *Width*) from the flower *Iris setosa*.

sample	1	2	3	4	5	6	7	8	9	10	11	12
Length	5.1	4.9	4.7	4.6	5.0	5.4	4.6	5.0	4.4	4.9	5.4	4.8
Width	3.5	3.0	3.2	3.1	3.6	3.9	3.4	3.4	2.9	3.1	3.7	3.4

   (a) Compute the mean and variance of both *Length* and *Width*. Compute the correlation between *Length* and *Width*. Comment on the strength and direction of the linear relationship between the variables.

   (b) Fit a simple linear regression using *Length* as a predictor variable and *Width* as a response variable. What is the estimated intercept and slope of the regression line? Write in words the interpretation of the slope.

   (c) Find the standard deviation around the regression line, i.e., estimate of σ or $\hat{\sigma}$.

   (d) Test the hypothesis that there is no linear relationship between *Length* and *Width*. What is the p-value of the test?

   (e) Construct a 95% confidence interval for the estimated mean *Width* when *Length* is 4.8.

   (f) Construct a 95% prediction interval of the *Width* for a specific sample when *Length* is 4.8.