AMS 315 Data Analysis: Homework 4
Due: March 21, 2022

1. (10 pts) Use the data given here to answer the following questions.

x	4	9	10	17	23	31
y	13	27	42	61	78	86

   (a) Plot the data values in a scatter diagram.
   (b) Find the least-squares estimates for the regression line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.
   (c) Use the least-squares prediction to predict y when $x = 12$.

2. (15 pts) Use the dataset given in Q1.

   (a) Find the 95% confidence interval for β_0.
   (b) Find the 95% confidence interval for the slope β_1.
   (c) Conduct the F test for testing $H_0 : \beta_1 = 0$

3. (15 pts) Use the dataset given in Q1.

   (a) Find the 95% confidence interval for $E[y_{n+1}]$ based on an assumed value for x_{n+1} of 12.
   (b) Find the 95% confidence interval for y_{n+1} based on an assumed value for x_{n+1} of 12.

4. (15 pts) A research team collected data on $n = 450$ students in a statistics course. The observed average final examination score was 520, with an observed standard deviation of 127.0 (the divisor in the estimated variance was $n - 1$). The average first examination score was 390, with an observed standard deviation of 96.0. The correlation coefficient between the first examination score and the final examination score was 0.53.

   (a) Report the analysis of variance table and result of the test of the null hypothesis that the slope of the regression line of final exam score on first exam score is zero against the alternative that it is not. Use the 0.10, 0.05, and 0.01 levels of significance.

   (b) Determine the least-squares fitted equation and give the 99% confidence interval for the slope of the regression of final examination score on first examination score.

   (c) Use the least-squares prediction equation to estimate the final examination score of students who scored 550 on the first examination. Give the 99% confidence interval for the expected final examination score of these students.

   (d) Use the least-squares prediction equation to predict the final examination score of a student who scored 550 on the first examination. Give the 99% prediction interval for the final examination score of this student.

5. (15 pts) A research team conducted a longitudinal study of participants between 25 and 30 years of age. They measured each participant's level of education at age 25. They also measured each participant's earnings at age 30. The team collected data on $n = 312$ participants. The average level of education at age 25 was 15.0, with an observed standard deviation of 3.0 (the divisor in the underlying variance calculation was $n - 1$). The average earnings (in thousands of dollars) was 54.0, with an observed standard deviation of 14.0 (the divisor in the underlying variance calculation was $n - 1$). The Pearson product moment correlation coefficient between the two variables was 0.75. The research team seeks to estimate the regression of participant earnings at age 30 on participant education at age 25.

   What is the 95% confidence interval for the population correlation coefficient of level of education at age 25 and earnings at age 30?

   Note: Following two Problems are R programing problem.

6. (15 pts) The Department of Natural Resources (DNR) received a complaint from recreational fishermen that a community was releasing sewage into the river where they fished. These types of releases lower the level of dissolved oxygen in the river and hence cause damage to the fish residing in the river. An inspector from the DNR designs a study to investigate the fishermen's claim. Fifteen water samples are selected at locations on the river upstream from the community and fifteen samples are selected downstream from the community. The dissolved oxygen readings in parts per million (ppm) are given in the following table.

Upstream	5.2	4.8	5.1	5.0	4.9	4.8	5.0	4.7	4.7	5.0	4.6	5.2	5.0	4.9	4.7
Downstream	3.2	3.4	3.7	3.9	3.6	3.8	3.9	3.6	4.1	3.3	4.5	3.7	3.9	3.8	3.7

   (a) Calculate the difference of the mean dissolved oxygen readings for the two locations on the river.
   (b) Do the data provide sufficient evidence to indicate a difference in mean oxygen content between location above and below the town? Use $\alpha = 0.05$.
   (c) Estimate the size of the difference in the mean dissolved oxygen readings for the two locations on the river using a 99% confidence interval.

7. (15 pts) The risk of an investment is measured in terms of the variance in the return that could be observed. Random samples of 10 yearly returns were obtained from two different portfolios. The data are given next (in thousands of dollars).

Portfolio 1	130	135	135	131	129	135	126	136	127	132
Portfolio 2	154	144	147	150	155	153	149	139	140	141

   (a) Does portfolio 2 appear to have a higher risk than portfolio 1?
   (b) Give a p-value for your test, and place a confidence interval on the ratio of the standard deviations of the two portfolios.