Please include a cover sheet that provides a complete sentence answer to each of the following three questions:

(a) In your opinion, what were the main ideas covered in the assignment?
(b) Were there any topics that you believe should have been covered, but were not?
(c) What problems did you have with the assignment, if any?

Write your solutions carefully and neatly, preferably in complete and grammatically correct sentences. Justify all of your arguments.

1. Question 11.3.8, page 568. (The relevant test statistic is $t = 6.51$.)

2. In Question 11.2.28, page 554, we observed that $1/x$ and $y$ are linearly related in the model $y = \beta_0 + \beta_1 \cdot \frac{1}{x}$, and this enables us to apply linear regression to determine the parameters $\beta_0$ and $\beta_1$ for a given data set. Assuming that the data points $(1/x_1, y_1), \ldots, (1/x_n, y_n)$ fit the Simple Linear Model, use this data to produce 90% confidence intervals for the parameters $\beta_1$ and $\sigma^2$. (The answers are $[19.21, 20.44]$ and $[0.51, 4.91]$.)


4. Question 11.3.20, page 568. (The answers are $[179.40, 216.18]$ and $[159.90, 235.68]$.)


7. Question 11.5.9, page 589.

8. Question 11.5.11, page 589.
Solutions to Selected Problems.

Solution (11.3.8). We want to determine if elevated index of exposure has an effect on cancer mortality. We assume that our data follows the simple linear model

\[ y = \beta_0 + \beta_1 x. \]

In the previous assignment, you computed maximum likelihood estimates for these parameters:

\[ \hat{\beta}_0 = 114.72, \quad \hat{\beta}_1 = 9.23. \]

Here is a plot of the data (blue) and the best fit line \( y = \hat{\beta}_0 + \hat{\beta}_1 x \) (red):

\[ T_{n-2} = \frac{\hat{\beta}_1 - 0}{s / \sqrt{\sum (x_i - \bar{x})^2}} \]

which is student-\( t \) distributed with \( n - 2 \) degrees of freedom (\( n = 9 \)). And we will reject the null hypothesis if \( t > t_{0.05,7} = 2.015 \).

Plugging our measurements in, we find

\[ s = \sqrt{\frac{1}{n-2} \sum (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2} = 14.01, \quad t = \frac{\hat{\beta}_1 - 0}{s / \sqrt{\sum (x_i - \bar{x})^2}} = 6.51. \]

Since \( t > 2.015 \), we reject the null hypothesis and conclude that increased index of exposure does correlate with increased incidence of cancer mortality.

Solution (Problem 2). The computations for this one are in the Sage worksheet for Homework 6. Recall that in order to apply linear regression to the model \( y = \beta_0 + \beta_1 \cdot x \), we will need to use the modified data points \((1/x_1, y_1), \ldots, (1/x_n, y_n)\). With this data, we get

\[ \hat{\beta}_0 = 1.54, \quad \hat{\beta}_1 = 19.83. \]
We obtain the sample variance
\[ s^2 = \frac{1}{n-2} \sum \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2 = 1.12. \]

The confidence interval for \( \beta_1 \) is
\[ \hat{\beta}_1 \pm t_{0.05, n-2} \cdot \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = [0.51, 4.91]. \]

The confidence interval for \( \sigma^2 \) is
\[ \left[ \frac{(n-2)s^2}{\chi^2_{0.05,n-2}}, \frac{(n-2)s^2}{\chi^2_{0.95,n-2}} \right] = [19.21, 20.44]. \]

**Solution** (11.4.10). We computed the sample correlation in the accompanying Sage worksheet:
\[ r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} = 0.7295. \]
Hence \( r^2 = 0.5322 \). Our interpretation is that 53.2\% of CHD mortality is explained by cigarette consumption.

**Solution** (11.5.9). See the accompanying Sage worksheet for calculations. Here we test the null hypothesis \( H_0 : \rho = 0 \) versus the alternate hypothesis \( H_1 : \rho \neq 0 \). Let
\[ t = \frac{\sqrt{n-2} \cdot r}{\sqrt{1 - r^2}} \]
be our test statistics, which behaves like a measurement from a student-\( t \) distribution with \( n - 2 \) degrees of freedom. We will reject the null hypothesis if \( t < -t_{0.025, n-2} = -2.23 \) or if \( t > t_{0.025, n-2} = 2.23 \). We compute that \( r = -0.0296 \), so that \( t = -0.094 \). Hence we fail to reject the null hypothesis. It appears that we do not have sufficient evidence to assert that home run frequency and home park altitude are dependent.