Instructions:

1) Make sure you have the correct number of pages. There are 9 pages and 17 questions. Questions are worth varying points, and the amount is listed at the question.

2) Please use a pencil.

3) Questions can be clarified, but no hints will be provided.

4) Please hand-in your formula sheet with your exam.

5) You may begin, good luck.
Question 1.

1. What are the four assumptions of the simple linear regression model we have discussed in class. [8 pts]

   i)  

   ii)  

   iii)  

   iv)  

Question 2.

2. In the context of simple linear regression, the point \((\bar{X}, \bar{Y})\) _____________. [5 pts]

Circle ALL answers that apply to the blank above:

a) will always be one of the points in the data set.

b) will always fall on the fitted line.

c) is not informative.

Question 3.

3. In simple linear regression, when we estimate \(\sigma^2\) we use \(\hat{\sigma}^2 = \frac{RSS}{n-2}\) where \(n\) is the number of data points and RSS is the residual sums of squares.

What does the number 2 in the denominator refer to? [5 pts]
Question 4.

4. If the range of the x-values are from x=100 to x=150, why might we be reluctant to use our data to perform a hypothesis test on \( \beta_0 \)? [8 pts]

Question 5.

5. A hypothesis test was performed using a \( t \)-test. Using the observed \( t \)-statistic, a p-value of 0.009 was computed. You have chosen a significance level of \( \alpha = 0.01 \) for the test. [5 pts]

Circle ALL the following statements that are true:

a) There is not strong evidence against the null.

b) Under the null being true, there’s a 0.009 chance of observing a \( t \)-statistic this extreme or more extreme.

c) There is a 0.009 probability that the alternative hypothesis is true.

d) You reject the null.

e) You have proven that the alternative is true.
Question 6.

6. Is it appropriate to fit a simple linear regression model that predicts survival (live or die) of a mouse using the quantitative variable of drug dosage (in milligrams) as a predictor variable? Briefly justify your answer. [5 pts]

Question 7.

7. How does the strength of a linear relationship in simple linear regression change if the units of the data are converted, say from feet to inches? [5 pts]

Question 8.

8. Why does it make sense that the variability in the estimated slope $\beta_1$ is smaller when the x-values are more spread out? Feel free to include a graph in your answer. [8 pts]
Question 9.

9. The plot on the left shows the histogram of a variable called Population. The plot on the right shows the variable after I transformed it.

What transformation(s) might have gotten me to this more symmetrical distribution on the right? Supplying one appropriate transformation is fine. [5 pts]

Question 10.

10. When estimating $\beta_0$ and $\beta_1$ through least squares estimation, what tactic do we use? In other words, what criterion do we use to find $b_0$ and $b_1$? [5 pts]

*A general answer is fine, but you can include some notation if you’d like.
Questions 11-17.

In 1965, data on the connection between radioactive waste exposure and cancer mortality was published. The data was collected from 9 counties that were located near an Atomic Energy Commission facility in Hanford, Washington.

The data give the index of exposure and the cancer mortality rate during 1959-1964 for the nine counties affected. Higher index of exposure values represent higher levels of contamination.

Variable Description:
- **County**: Name of county
- **Exposure**: Index of exposure
- **Mortality**: Cancer mortality per 100,000 man-years

The data is as follows:

<table>
<thead>
<tr>
<th>County</th>
<th>Exposure</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Umatilla</td>
<td>2.49</td>
<td>147.1</td>
</tr>
<tr>
<td>2 Morrow</td>
<td>2.57</td>
<td>130.1</td>
</tr>
<tr>
<td>3 Gilliam</td>
<td>3.41</td>
<td>129.9</td>
</tr>
<tr>
<td>4 Sherman</td>
<td>1.25</td>
<td>113.5</td>
</tr>
<tr>
<td>5 Wasco</td>
<td>1.62</td>
<td>137.5</td>
</tr>
<tr>
<td>6 HoodRiver</td>
<td>3.83</td>
<td>162.3</td>
</tr>
<tr>
<td>7 Portland</td>
<td>11.64</td>
<td>207.5</td>
</tr>
<tr>
<td>8 Columbia</td>
<td>6.41</td>
<td>177.9</td>
</tr>
<tr>
<td>9 Clatsop</td>
<td>8.34</td>
<td>210.3</td>
</tr>
</tbody>
</table>

The scatterplot:

* This is a measure of cancer rate per 100,000 people for a certain amount of time
Output from fitting the simple linear regression for predicting Mortality from Exposure is shown below:

```r
> lm.out=lm(Mortality~Exposure)

> summary(lm.out)

Call:
  lm(formula = Mortality ~ Exposure)

Residuals:
       Min       1Q   Median       3Q      Max
-16.2950 -12.7552   4.0113   9.3983  18.5944

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)         114.7160     8.0458  14.258 1.98e-06 ***
Exposure             9.2314     1.4185   6.507 0.000332 ***
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 14.01 on 7 degrees of freedom
Multiple R-Squared: 0.8581, Adjusted R-squared: 0.8378
F-statistic: 42.34 on 1 and 7 DF,  p-value: 0.0003321

> anova(lm.out)

Analysis of Variance Table

Response: Mortality
            Df  Sum Sq Mean Sq  F value Pr(>F)
Exposure    1 8309.6  8309.6   42.34   0.0003321 ***
Residuals   7 1373.9  196.3
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
11. What is the expected mortality rate for a county with an exposure index of 3? [5 pts]

12. Calculate two points that fall on the fitted line (and would fall in the window of the scatterplot shown), draw the two points on the scatterplot, and connect them to show the fitted line. [6 pts]

   Show your work for calculating the points.

13. Interpret the estimated slope of the fitted model. [5 pts]

14. What is the correlation between Mortality and Exposure? [5 pts]
   (You can get this from data provided.)
15. Is there a significant linear relationship between Mortality and Exposure? Provide a null hypothesis, a test statistic, p-value, and conclusion. [10 pts]

16. What is the estimated variance of the observations at a conditional mean? [5 pts]

17. What is the Total Sums of Squares (TSS) for this data? [5 pts]