

Stat 401 A XM Homework 9
On-Campus Due Date: Wednesday, November 12
Off-Campus Due Date: November, November 19

1. It is generally thought that the percentage of fruits attacked by codling moth larvae is greater on apple trees bearing a small crop. Apparently the density of the flying moths is unrelated to the size of the crop on a tree, so the chance of attack for any particular fruit is increased if few fruits are on the tree. Data on a sample of 10 trees gives a sample linear correlation coefficient of -0.8 . Other summary statistics obtained from the sample of 10 trees are provided below.

Variable	Sample Mean	Sample Standard Deviation
Crop Size (number of fruit)	110	40
Percentage of Wormy Fruits	45	12

- (a) Find the equation of the least-squares regression line for predicting percentage of wormy fruits from crop size.
 - (b) Estimate the mean percentage of wormy fruits for trees with a crop size of 150 fruit.
 - (c) Provide a 95% confidence interval for the mean percentage of wormy fruits for trees with a crop size of 150 fruit.
 - (d) Suppose you wish to test the hypothesis: “Trees with a crop size of 150 have, on average, 50% wormy fruit.” Would you accept or reject the null hypothesis at significance level 0.05?
 - (e) Suppose a tree with 150 fruit is randomly selected from all trees with 150 fruit. Provide an interval that will contain the percentage of wormy fruit for this tree with 90% confidence.
 - (f) Write down the ANOVA table for the simple linear regression of percentage of wormy fruit on crop size. Compute the F statistic and find a p -value.
 - (g) Based on the F statistic and p -value that you computed in part (f), would you say the slope of the regression line is significantly different from 0?
 - (h) What proportion of the variability in percentage of wormy fruit is explained by the regression of percentage of wormy fruit on crop size?
 - (i) Find a 99% confidence interval for β_1 , the slope of the least squares regression line for predicting percentage of wormy fruit from crop size.
2. Read Section 8.4 from your text and the handout on *Interpreting the Slope of the Least Squares Regression Line* available at

<http://www.public.iastate.edu/~dnett/S401/handouts.shtml>.

Suppose a researcher is studying the effect of a chemical on the concentration of a certain bacteria. She prepares 20 identical bacteria cultures in 20 petri dishes. She randomly assigns each of 5 chemical concentrations (1,2,4,8,16 mg/L) to 4 petri dishes. A bacteria count is obtained for each dish after 24 hours of exposure to the chemical. The researchers notices that the log of the bacteria count seems to be linearly associated with the log of the chemical concentration. The equation of the least-squares regression line for predicting log bacteria count from log chemical concentration is determined to be $\log(\hat{Y}) = 12 - 1.7 \log(X)$. The 95% confidence interval for β_1 was determined to be -0.2 to -3.2 .

- (a) Fill in the blanks in the following sentence.

A doubling of the chemical concentration was estimated to cause a multiplicative change of _____ (95% confidence interval _____ to _____) in the median bacteria count.

- (b) Estimate the median bacteria counts for chemical concentrations of 6 and 12 mg/L, respectively. Show that these estimates are consistent with your statement in part (a).
3. Consider the data described in problem 23 of Chapter 8 on wine consumption and heart disease mortality in 18 countries. Let X denote the wine consumption in liters per person per year. Let Y denote the heart disease mortality in deaths per 1000 persons. Examine and run the program *wine.sas* and answer the following questions.
- (a) Examine plots of Y vs. X , $\log(Y)$ vs. X , Y vs. $\log(X)$, and $\log(Y)$ vs. $\log(X)$. For each plot, state whether a linear relationship is potentially appropriate or inappropriate.
- (b) Give the equation of the least-squares regression line for the most appropriate of the four potential regressions relating Y to X . (The most appropriate choice is debatable, so full credit will be given for multiple answers.)
- (c) Write a statement that provides an interpretation of the slope of the most appropriate least-squares regression line. Include a confidence interval along with any estimates in your statement.
4. The SAS program *meatproc.sas* contains the 12 data points described in problem 16 of Chapter 8 regarding the pH level in steer muscle at various hours after slaughter. Test for a lack of fit in the regression of pH on $\log(Hour)$. Give the test statistic, a p -value, and a conclusion.