STAT-S 631 **Assignment 4** **Due date: 9/22/17**

If R is used, present the required output and the (relevant) syntax.

1. From ALR Problem 2.16.

2. The data `Sahlins.txt` that you can download from our Canvas page[1], were compiled by Sahlins (1972) from information presented in Scudder's (1962) report on the Gwenba valley of Central Africa. The data describe agricultural production in Mazulu village. The explanatory variable (Consumers/Gardener) is the ratio of consumers to productive individuals in each of 20 households, making suitable adjustments for the consumption requirements of different household members. The response variable (Acres/Gardener) is a measure of domestic-labor intensity, based on the amount of land cultivated by each household. Think of Consumers/Gardener as representing the relative consumption needs of the household, and Acres/Gardener as representing how hard each productive individual in the household works. Sahlins was interested in production, consumption, and redistribution of the social product in "primitive" communities.

   a) Draw a scatterplot of Acres/Gardener (Y) versus Consumers/Gardener (X). What relationship, if any, do you discern in this plot –does the relationship appear to be positive or negative (or neither), linear or nonlinear, strong or weak? Is there anything else noteworthy about the data– for example, do any households appear to be unusual?

   b) Analyze the data by regressing Acres/Gardener on Consumers/Gardener. In a society characterized by primitive communism, the social product of the village would be redistributed according to need, while each household would work in proportion to its capacity, implying a regression slope of zero. In contrast, in a society in which redistribution is purely through the market, each household should have to work in proportion to its consumption needs, suggesting a positive regression slope and an intercept of zero. Interpret the results of the regression in light of these observations. Examine and interpret the values of $\hat{\beta}_0, \hat{\beta}_1$, $\hat{\sigma}^2$. Do the results change if the fourth household is deleted? Plot the regression lines calculated with and without the fourth household on a scatterplot of the data. Does either regression do a good job of summarizing the relationship between Acres/Gardener and Consumers/Gardener? (see your response in part (a))

   c) Find the standard errors of the intercept and slope. Can we conclude that the population slope is greater than zero? Can we conclude that the intercept is greater than zero? Obtain both, confidence intervals and perform hypothesis tests to answer these questions. Use some reasonable significance level (or corresponding confidence levels). Repeat these computations omitting the fourth household. Provide your conclusions for both scenarios.

   d) Use the regression coefficients for the entire data (20 households). What do you expect to be the Acres/Gardener ratio for a household with a Consumers/Gardener ratio equal to 1.5. To answer this question, obtain an interval with a 98% confidence level. Would your answer change if instead you are asked to determine the mean Acres/Gardener ratio for all those households with a Consumers/Gardener ratio equal to 1.5? Explain why or why not.

[1]The data and questions were constructed based on the supplementary material of "Applied Regression Analysis and Generalized Linear Models" 3rd Ed by Fox.