

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

1. The data `Sahlins.txt` that you can download from our Canvas page¹, 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) Create 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) Notice that the households are ordered by the values of Consumers/Gardener (X). Divide the 20 households into three groups, placing the first seven households in the first group, the next six in the second group, and the last seven in the third group. Add a column to the data frame to represent the group assignment. Use dplyr tools (e.g. `group_by`, `summarize`) to calculate the mean Y and mean X in each of the three groups. Plot these means on the scatterplot and connect them with a simple nonparametric regression line using `geom_line`. Does the regression line help you to interpret the relationship between the variables?
 - c) Two of the households, one in the first group, one in the second (which ones?), stand out from the others. Recalculate the means in groups 1 and 2 omitting those observations, and redraw the nonparametric regression line. How, if at all, do the new means differ from the original values?

¹The data and question are based on the supplementary material of "Applied Regression Analysis and Generalized Linear Models" 3rd Ed by Fox.