

Homework 4

1. Use Sahlins data (in car library) for this problem. This is data from a 1962 report on the Gwenda valley of Central Africa describing agricultural production in a Mazulu village. The independent variable (Consumers/Gardener) is the ration of consumers to productive individuals in each of 20 households, making suitable adjustments for the consumption requirements of different household members. The dependent 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.

Draw a scatterplot (using statistical software) of Acres/Gardener (Y) versus Consumers/Gardener (X). What relationship, if any, do you discern in this plot? Does the relationship appear to be linear or nonlinear? Is it strong or weak? Is there anything else noteworthy about the data, for example do any households appear to be unusual?

Notice that the households are ordered by the values of Consumers/Gardener (X). Divide the 20 households into three groups (7 in the 1st group, 6 in the 2nd, and 7 in the third). Calculate the mean Y and X in each of the three groups. Transfer the means to the scatterplot and connect with a simple non-parametric regression line. Does the regression line help you interpret the relationship between the variables.

Estimate and plot a locally weighted regression. Compare and contrast this analysis with the previous two steps. What is your best interpretation of the relationship between the variables?

(adapted from Fox 1997)

2. Using the methods for displaying univariate distributions described in Chapter 3 of Fox (2008), examine the distributions of the variables Income and Infant Mortality for 103 nations of the world (Leinhardt in car library). Examine the data for skewness, nonnormality, multiple modes, and unusual values.

Consider the differences between regions for both variables using parallel comparison boxplots. In each case, compare levels and spreads across the parallel boxplots, and look for unusual data values. Use labels where appropriate.

(adapted from Fox 1997)