Question 3 (*Chapter 10 Explanatory relations*)

(a) The data set `electric` in the SSC data subdirectory gives the average daily electricity usage (in kilowatt hours) and the average temperature (in degrees Fahrenheit, °F) for 55 consecutive months from August 1989 to February 1994 in Westchester County, New York. The question of interest is how the temperature affects levels of electricity consumption.

(i) Draw an appropriate scatter plot to indicate whether average temperature might provide a useful predictor for electricity usage, and comment. [2]

(ii) Find the equation of the least squares regression line through these data, and interpret the slope of your fitted line. [3]

(iii) Find a 95% confidence interval for the slope of the regression line for electricity usage against temperature. Does this cast any light on the usefulness of temperature as a predictor for electricity usage? [3]

Hint Try `?betaci` to discover the syntax for confidence limits for the slope of a regression line.

(iv) Use the fitted regression line to predict the electricity usage in a month when the average temperature is 50°F, and find the corresponding 99% predictive interval. [3]

(b) The data set `target` in the SSC data subdirectory gives results from an experiment on gunnery carried out in the eighteenth century. A gun was fired at a mark on a target from various different ranges. The bullet struck the target at varying distances from the mark. The data set gives the results of 24 separate firings of the gun.

(i) To investigate how the distance from the mark varies with range, produce a scatter plot of distance against range. Include on your plot the fitted least squares regression line, making clear the regression model you have decided to fit. [3]

(ii) By reference to your plot, comment on whether the assumptions of the regression model appear to be satisfied in these data. [2]

(iii) Two observations, one at range 6 feet and the other at range 9 feet, seem to be very different from the other observations at those distances. Remove these outliers from the data set, and again plot the data together with the fitted least squares regression line. Explain how the slope of your line differs from that calculated in part (b)(i). [4]

(iv) After removing these outliers, does it appear that the assumptions of the regression model are satisfied? Briefly explain why or why not. [2]

(c) Six industrial heaters of a particular type were tested at each of four temperatures, and their life (in hours) at that temperature was recorded. Interest centred on how the operating temperature could be used to predict the lifetime. The data are available in the data set `heater` in the SSC data subdirectory. The variables are `temp` (operating temperature in °F) and `hours` (lifetime in hours).

(i) By plotting the data, comment on why it would not be very useful to fit a least squares regression line to these data without transformation. [2]

A suggested alternative regression model for these data is to transform `hours` to `log(hours)` and `temp` to `(temp + 460)^(-1)`. (Note that `temp + 460` is the temperature in degrees Fahrenheit above absolute zero, so that there is a physical reason for this odd-looking transformation.)

(ii) Perform this transformation. Plot the transformed data and comment on the appropriateness of fitting a straight line using least squares. Calculate the resulting least squares line. Use it to predict the lifetime of a heater of this type operating at 1600°F, and find the corresponding 90% predictive interval for the lifetime. [6]