4.9 Exercises

4.9.10. Hettmansperger and McKean (2011) discuss a dataset in which the dependent variable is the cloud point of a liquid, a measure of degree of crystallization in a stock, and the independent variable is the percentage of `I-8` in the base stock. For the readers' convenience, the data can be found in the dataset cloud in the package `npsm`.

(a) Scatterplot the data. Based on the plot, is a simple linear regression model appropriate?

(b) Show by residual plots of the fits that the linear and quadratic polynomials are not appropriate but that the cubic model is.

(c) Use the R function `polydeg`, with a super degree set at 5, to determine the degree of the polynomial. Compare with Part (b).

4.9.11. Devore (2012) discusses a dataset on energy. The response variable is the energy output in watts while the independent variable is the temperature difference in degrees K. A polynomial fit is suggested. The data are in the dataset energy.

(a) Scatterplot the data. What degree of polynomial seems suitable?

(b) Use the R function `polydeg`, with a super degree set at 6, to determine the degree of the polynomial.

(c) Based on a residual analysis, does the polynomialifit of Part (b) provide a good fit?

4.9.13. As in the last problem, consider the weather dataset, weather. One of the variables is total snowfall (in inches), totalsnow, for the month of January.

(a) Scatterplot total snowfall versus year. Determine the years of maximal and minimal snowfalls.

(b) Obtain the local `LS` and robust `loess` fits of the data. Compare the fits.

(c) Perform a residual analysis on the robust fit.

(d) Obtain a boxplot of the residuals found in Part (c). Identify the outliers by year.