Chapter 1 Exercises

1. (a) Fit a cubic regression, as a function of age, to the kidney data of Figures 1.1 and 1.2, calculating estimates and standard errors at ages 20, 30, 40, 50, 60, 70, 80.

See answer in attached Jupyter Notebook

(b) How do the results compare with those in Table 1.1?

See answer in attached Jupyter Notebook

2. The lowess curve in Figure 1.2 has a flat spot between ages 25 and 35. Discuss how one might use bootstrap replications like those in Figure 1.3 to suggest whether the flat spot is genuine or just a statistical artifact.

Ans: One way I can think of is to Bootstrap sample only from observations from subjects with ages 25 to 35. On each bootstrap dataset I would compute a linear regression using age and and intercept. Then I would collect the value of the coefficient for age on each bootstrap iteration. Finally, I would create a histogram with the values of the age coefficients across bootstrap samples and compute the confidence interval on the age coefficients using it. If the CI ranges from negative to positive values, then I would conclude the flat spot it legitimate.

3. Suppose that there were no differences between AML and ALL patients for any gene, so that t in (1.6) exactly followed a student-t distribution with 70 degrees of freedom in all 7128 cases. About how big might you expect the largest observed t value to be? Hint: $1/7128$ = 0.00014.

Ans: If the distribution of the difference in the means from AML and ALL patients was indeed t with 70 degrees of freedom, then by the hint, we´d expect the largest value to be larger than the other 7127 just by chance (the probability that a random value is the largest would be 1/7128). If we look for such a value in the t distribution with 70 degrees of freedom, we get 3.826.

4. (a) Perform 1000 nonparametric bootstrap replications of ALL (1.5). You can use program bcanon from the CRAN library "bootstrap" or type in the little program Algorithm 10.1 on page 178.

(b) Do the same for AML.

(c) Plot histograms of the results, and suggest an inference.

See answer in attached Jupyter Notebook