

Prediction, estimation and attribution

Exercises

1 CASI

Many of these exercises use data used in the book. These datasets can be found on the book webpage <https://web.stanford.edu/~hastie/CASI>.

Chapter 1 Exercises

- 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.
 - How do the results compare with those in Table 1.1?
- 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.
- 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: $1/7128 = 0.00014$.