7.9 Exercises

7.9.1. To see the effect on fits that "good" and "bad" points of high leverage can have, consider the following dataset:

x	1	2	3	4	5	6	7	8	9	10	20
y	5	7	6	14	14	25	29	33	31	41	75
y_2	5	7	6	14	14	25	29	33	31	41	20

The point x = 20 is a point of high leverage. The data for **y** (rounded) are realizations from the model $\mathbf{y} = 4x + e$, where e has a N(0, 9) distribution. Hence, the the value of y follows the model and is a "good" point of high leverage. Notice that $\mathbf{y_2}$ is the same as **y**, except the last component of $\mathbf{y_2}$ has been changed to 20 and, thus, is a "bad" point of high leverage.

(a) Obtain the scatterplot for x and y, the Wilcoxon and HBR fits, and overlay these fits on the scatterplot.

(b) Obtain the scatterplot for x and y_2 , the Wilcoxon and HBR fits, and overlay these fits on the scatterplot.

(c) Comment on the differences among the fits and plots.

7.9.3. There is some loss of efficiency when using the HBR fit instead of the Wilcoxon for "good" data. Verify this for a simulation of the model y = 4x+e, where e has a N(0, 625) distribution and x = 1 : 20, using 10,000 simulations.

7.9.4. Using the set up Exercise 7.9.3, check the validity of the 95% confidence intervals for β1 obtained by the Wilcoxon and HBR fits.