Exercise 13: Predicting water salinity

The data are taken from[6]. The first column (y) gives the water salinity of a river discharge in North Carolina's Pamlico Sound whereas the other columns (x_1, x_2, x_3) are predictor variables, respectively salinity two weeks earlier, a linear time trend computed from the history of the series and the amount of river discharge.

y	x_1	x_2	x_3
7.6	8.2	4	23.005
7.7	7.6	5	23.873
4.3	4.6	0	26.417
5.9	4.3	1	24.868
5.0	5.9	2	29.895
6.5	5.0	3	24.200
8.3	6.5	4	23.215
8.2	8.3	5	21.862
13.2	10.1	0	22.274
12.6	13.2	1	23.830
10.4	12.6	2	25.144
10.8	10.4	3	22.430
13.1	10.8	4	21.785
12.3	13.1	5	22.380
10.4	13.3	0	23.927
10.5	10.4	1	33.433
7.7	10.5	2	24.859
9.5	7.7	3	22.686
12.0	10.0	0	21.789
12.6	12.0	1	22.041
13.6	12.1	4	21.033
14.1	13.6	5	21.005
13.5	15.0	0	25.865
11.5	13.5	1	26.290
12.0	11.5	2	22.932
13.0	12.0	3	21.313
14.1	13.0	4	20.769
15.1	14.1	5	21.393

a) Read the data into some software package and run an ordinary linear regression with three covariates.

b) Compute the fitted values and the residuals. How are these defined?

c) Various plots of the residuals may be used to check the model assumptions. Make plots for checking (i) normality, (ii) homoscedasticy (equal variances), and (iii) linearity. Comment on what each of the plots tell you. Also check if some of the observations have a large influence on the estimates of the fitted model.

d) Reanalyse the data in some modified way if you find it necessary. What is your suggested predictor for water salinity?

[6]Staudte R. G. and Sheather S. J. (1990). Robust estimation and testing. Wiley.