HW of Regression Diagnostics

1. Consider the computer repair problem discussed in Section 2.3. In a second sampling period, 10 more observations on the variable Minutes and Units were obtained. Since all observations were collected by the same method from a fixed environment, all 24 observations were pooled to form one data set. The data appear in Table 4.6.

Table 4.6 Expanded Computer Repair Times Data: Length of Service Calls (Minutes) and Number of Units Repaired (Units)

Row	Units	Minutes	Row	Units	Minutes
1	1	23	13	10	154
2	2	29	14	10	166
3	3	49	15	11	162
4	4	64	16	11	174
5	4	74	17	12	180
6	5	87	18	12	176
7	6	96	19	14	179
8	6	97	20	16	193
9	7	109	21	17	193
10	8	119	22	18	195
11	9	149	23	18	198
12	9	145	24	20	205

(a) Fit a linear regression model relating Minutes to Units.

(2) Check each of the standard regression assumptions and indicate which assumption(s) seems to be violated.

2. Consider the data in Table 4.8, which consist of a response variable Y and six predictor variables. Consider fitting a linear model relating Y to all six X-variables.

(a) What least squares assumptions (if any) seem to be violated?

(b) Compute $r_i, C_i, DFITS_i$.

(c) Construct the index plots of $r_i, C_i, DFITS_i$.

(d) Identify all unusual observations in the data and classify each according to type (i.e., outliers, leverage points, etc.)

Suppose that we fit a linear model relating Y to the first three X-variables. Justify your answer to each of the following questions with the appropriate added-variable plot:

(e) Should we add X_4 to the above model? If yes, keep X_4 in the model.

(f) Should we add X_5 to the above model? If yes, keep X_5 in the model.

(g) Should we add X_6 to the above model?

(h) Which model(s) would you recommend as the best possible description of Y? Use the above results and/or perform additional analysis if needed.

Table 4.8 Data for Exercises 4.12–4.14

Row	Y	X_1	X_2	X_3	X_4	X_5	X_6
1	443	49	79	76	8	15	205
2	290	27	70	31	6	6	129
3	676	115	92	130	0	9	339
4	536	92	62	92	5	8	247
5	481	67	42	94	16	3	202
6	296	31	54	34	14	11	119
7	453	105	60	47	5	10	212
8	617	114	85	84	17	20	285
9	514	98	72	71	12	−1	242
10	400	15	59	99	15	11	174
11	473	62	62	81	9	1	207
12	157	25	11	7	9	9	45
13	440	45	65	84	19	13	195
14	480	92	75	63	9	20	232
15	316	27	26	82	4	17	134
16	530	111	52	93	11	13	256
17	610	78	102	84	5	7	266
18	617	106	87	82	18	7	276
19	600	97	98	71	12	8	266
20	480	67	65	62	13	12	196
21	279	38	26	44	10	8	110
22	446	56	32	99	16	8	188
23	450	54	100	50	11	15	205
24	335	53	55	60	8	0	170
25	459	61	53	79	6	5	193
26	630	60	108	104	17	8	273
27	483	83	78	71	11	8	233
28	617	74	125	66	16	4	265
29	605	89	121	71	8	8	283
30	388	64	30	81	10	10	176
31	351	34	44	65	7	9	143
32	366	71	34	56	8	9	162
33	493	88	30	87	13	0	207
34	648	112	105	123	5	12	340
35	449	57	69	72	5	4	200
36	340	61	35	55	13	0	152
37	292	29	45	47	13	13	123
38	688	82	105	81	20	9	268
39	408	80	55	61	11	1	197
40	461	82	88	54	14	7	225

Source: Chatterjee and Hadi (1988)