Department of Statistics

COURSE STATS 330/762

Assignment 2, 2012

Instructions: Hand in your completed assignment to the Student Resource Centre by **4pm August 16th.**

The data sets for this assignment is in the file **ozone.txt** which is available on the course web page.

The data consist of atmospheric measurements made on 111 days from May to September 1973 in New York. The variables measured are

ozone:	ozone concentration (ppb),
radiation:	solar radiation (langleys)
temperature:	daily maximum temperature (degrees F),
wind:	wind speed (mph)**.**

Questions and tasks

1. Load the data into R, and make a data frame **ozone.df** to contain the data. Check for any typographical errors (the data below may be taken to be the correct data, but the data on the web may have been corrupted). Correct as necessary. Print out the last 10 lines of the data file. [5 marks]

2. Fit a regression model to these data, using **ozone** as the response. Do all variables appear to have an effect on ozone levels? What is the nature of these effects, if any? Give reasons.[5 marks]

3. Do you think that these data are suitable to be modeled by a linear regression model? If not, what action should be taken? Is the fit improved if this action is taken? [10 marks]

4. Are there any data points that have high leverage or large studentised residuals? If so, are these points having an undue influence on any aspect of the fitted model? Give a full discussion with reasons. [15 marks]

5. Do you think these data have collinearity problems? Give reasons. [5 marks]

Data follow – you may take these as the **correct** data

	ozone	radiation	temperature	wind
1	41	190	67	7.4
2	36	118	72	8.0
3	12	149	74	12.6
4	18	313	62	11.5
5	23	299	65	8.6
6	19	99	59	13.8
7	8	19	61	20.1
8	16	256	69	9.7
9	11	290	66	9.2
10	14	274	68	10.9
11	18	65	58	13.2
12	14	334	64	11.5
13	34	307	66	12.0
14	6	78	57	18.4
15	30	322	68	11.5
16	11	44	62	9.7
17	1	8	59	9.7
18	11	320	73	16.6
19	4	25	61	9.7
20	32	92	61	12.0
21	23	13	67	12.0
22	45	252	81	14.9
23	115	223	79	5.7
24	37	279	76	7.4
25	29	127	82	9.7
26	71	291	90	13.8
27	39	323	87	11.5
28	23	148	82	8.0
29	21	191	77	14.9
30	37	284	72	20.7
31	20	37	65	9.2
32	12	120	73	11.5
33	13	137	76	10.3
34	135	269	84	4.0
35	49	248	85	9.2
36	32	236	81	9.2
37	64	175	83	4.6
38	40	314	83	10.9
39	77	276	88	5.1
40	97	267	92	6.3
41	97	272	92	5.7
42	85	175	89	7.4
43	10	264	73	14.3
44	27	175	81	14.9
45	7	48	80	14.3
46	48	260	81	6.9
47	35	274	82	10.3
48	61	285	84	6.3
49	79	187	87	5.1
50	63	220	85	11.5
51	16	7	74	6.9

52	80	294	86	8.6
53	108	223	85	8.0
54	20	81	82	8.6
55	52	82	86	12.0
56	82	213	88	7.4
57	50	275	86	7.4
58	64	253	83	7.4
59	59	254	81	9.2
60	39	83	81	6.9
61	9	24	81	13.8
62	16	77	82	7.4
63	122	255	89	4.0
64	89	229	90	10.3
65	110	207	90	8.0
66	44	192	86	11.5
67	28	273	82	11.5
68	65	157	80	9.7
69	22	71	77	10.3
70	59	51	79	6.3
71	23	115	76	7.4
72	31	244	78	10.9
73	44	190	78	10.3
74	21	259	77	15.5
75	9	36	72	14.3
76	45	212	79	9.7
77	168	238	81	3.4
78	73	215	86	8.0
79	76	203	97	9.7
80	118	225	94	2.3
81	84	237	96	6.3
82	85	188	94	6.3
83	96	167	91	6.9
84	78	197	92	5.1
85	73	183	93	2.8
86	91	189	93	4.6
87	47	95	87	7.4
88	32	92	84	15.5
89	20	252	80	10.9
90	23	220	78	10.3
91	21	230	75	10.9
92	24	259	73	9.7
93	44	236	81	14.9
94	21	259	76	15.5
95	28	238	77	6.3
96	9	24	71	10.9
97	13	112	71	11.5
98	46	237	78	6.9
99	18	224	67	13.8
100	13	27	76	10.3
101	24	238	68	10.3
102	16	201	82	8.0
103	13	238	64	12.6
104	23	14	71	9.2
105	36	139	81	10.3

106	7	49	69	10.3
107	14	20	63	16.6
108	30	193	70	6.9
109	14	191	75	14.3
110	18	131	76	8.0
111	20	223	68	11.5