Stat 440/540 Regression Analysis. Fall 2006. Final Exam. Due Wednesday, December 13

Directions: The exam has one question. Your answer should be a word-processed report divided in sections and no longer than five pages. An appendix is allowed which may include relevant graphs, tables, output or computer code. An output only report will be acceptable but diminishing to your grade.

This is an open question but be sure to base your answers on principles of exploratory data analysis, regression modeling, comparison of models, variable selection, residual analysis/diagnostics, transformations, outlier/leverage points, etc. Try to be clear and concise. To prepare your report you may use any notes, books, articles, etc. that you may have available. However, you are not allowed to discuss the exam with other classmates or any other students.

1. Anthropologists conducted a study to determine the long-term effects of an environmental change on systolic blood pressure. They measured the blood pressure, and several other characteristics, of 39 Indians who migrated from a very primitive environment high in the Andes into the mainstream of Peruvian society at a lower altitude. All of the Indians were males at least 21 years of age, and were born at a high altitude.

    The data are given below, and a text file version can be found on the WWW from the class web page. The file contains 39 rows, one for each individual, and 12 columns corresponding to the variables, from left to right:

```
n = individual id
v1 = age in years             v2 = years since migration
v3 = weight in kilos          v4 = height in mm
v5 = chin skin fold in mm     v6 = forearm skin fold in mm
v7 = calf skin fold in mm     v8 = pulse rate-beats/min
v9 = systolic bp              v10 = diastolic bp
v11 = years since migration divided by age
```

n	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11
1	21	1	71.0	1629	8.0	7.0	12.7	88	170	76	0.04762
2	22	6	56.5	1569	3.3	5.0	8.0	64	120	60	0.27273
3	24	5	56.0	1561	3.3	1.3	4.3	68	125	75	0.20833
4	24	1	61.0	1619	3.7	3.0	4.3	52	148	120	0.04167
5	25	1	65.0	1566	9.0	12.7	20.7	72	140	78	0.04000
6	27	19	62.0	1639	3.0	3.3	5.7	72	106	72	0.70370
7	28	5	53.0	1494	7.3	4.7	8.0	64	120	76	0.17857
8	28	25	53.0	1568	3.7	4.3	0.0	80	108	62	0.89286
9	31	6	65.0	1540	10.3	9.0	10.0	76	124	70	0.19355
10	32	13	57.0	1530	5.7	4.0	6.0	60	134	64	0.40625
11	33	13	66.5	1622	6.0	5.7	8.3	68	116	76	0.39394
12	33	10	59.1	1486	6.7	5.3	10.3	73	114	74	0.30303
13	34	15	64.0	1578	3.3	5.3	7.0	88	130	80	0.44118
14	35	18	69.5	1645	9.3	5.0	7.0	60	118	68	0.51429
15	35	2	64.0	1648	3.0	3.7	6.7	60	138	78	0.05714
16	36	12	56.5	1521	3.3	5.0	11.7	72	134	86	0.33333
17	36	15	57.0	1547	3.0	3.0	6.0	84	120	70	0.41667
18	37	16	55.0	1505	4.3	5.0	7.0	64	120	76	0.43243
19	37	17	57.0	1473	6.0	5.3	11.7	72	114	80	0.45946
20	38	10	58.0	1538	8.7	6.0	13.0	64	124	64	0.26316
21	38	18	59.5	1513	5.3	4.0	7.7	80	114	66	0.47368

22	38	11	61.0	1653	4.0	3.3	4.0	76	136	78	0.28947
23	38	11	57.0	1566	3.0	3.0	3.0	60	126	72	0.28947
24	39	21	57.5	1580	4.0	3.0	5.0	64	124	62	0.53846
25	39	24	74.0	1647	7.3	6.3	15.7	64	128	84	0.61538
26	39	14	72.0	1620	6.3	7.7	13.3	68	134	92	0.35897
27	41	25	62.5	1637	6.0	5.3	8.0	76	112	80	0.60976
28	41	32	68.0	1528	10.0	5.0	11.3	60	128	82	0.78049
29	41	5	63.4	1647	5.3	4.3	13.7	76	134	92	0.12195
30	42	12	68.0	1605	11.0	7.0	10.7	88	128	90	0.28571
31	43	25	69.0	1625	5.0	3.0	6.0	72	140	72	0.58140
32	43	26	73.0	1615	12.0	4.0	5.7	68	138	74	0.60465
33	43	10	64.0	1640	5.7	3.0	7.0	60	118	66	0.23256
34	44	19	65.0	1610	8.0	6.7	7.7	74	110	70	0.43182
35	44	18	71.0	1572	3.0	4.7	4.3	72	142	84	0.40909
36	45	10	60.2	1534	3.0	3.0	3.3	56	134	70	0.22222
37	47	1	55.0	1536	3.0	3.0	4.0	64	116	54	0.02128
38	50	43	70.0	1630	4.0	6.0	11.7	72	132	90	0.86000
39	54	40	87.0	1542	11.3	11.7	11.3	92	152	88	0.74074

The question of interest here is to determine which factors are the most important for explaining systolic pressure. Answer this question by identifying the appropiate response variable and by fitting an appropiate model (or models), that include relevant explanatory variables. You may find that some simple graphs and summary statistics will be helpful in thinking about the relationship between systolic pressure and factors. Structure your report by including sections that describe your analysis and your results in technical terms. Also, write some general conclusions that could be understood by a non-statistician.