

TEST 2 STAT 572

Spring 2001

1. It generally perceived that military retirees like to retire in areas that are close to military bases and also have pleasant climate. Data on each state were collected to attempt to determine the factors that affect the choice of a state for military retirees. The variables are:

RTD: Total military retirement pay, a proxy for number of retirees
 ACT: Total active military pay, a proxy for total military population
 DOD: Total defense spending in state
 POP: Total population in state
 CLI: Cost of living index
 LAT: Latitude (north in degrees)
 PCP: Days of precipitation

To explore the data, a model with only main-effects was fitted. Values of Cook's D were obtained and given its extremely large value, the state of California was removed from the analysis. After including interactions and doing variable selection on the remaining data, the results of the final model appear below

Summary of Fit

RSquare	0.843155
RSquare Adj	0.780417
Root Mean Square Error	143293.9
Mean of Response	197983
Observations (or Sum Wgts)	50

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t	VIF
Intercept	2183707.4	909881.4	2.40	0.0219	.
ACT	-1.203817	0.407997	-2.95	0.0056	172.17715
DOD	0.3613531	0.157982	2.29	0.0283	406.86748
ACT*DOD	6.1651e-8	1.88e-8	3.28	0.0024	18.114217
POP	-171.3711	167.1884	-1.03	0.3124	1118.6211
ACT*POP	-0.000054	0.000021	-2.63	0.0127	46.57396
DOD*POP	-0.00001	0.000002	-4.96	<.0001	13.147937
CLI	-5929.798	3082.878	-1.92	0.0626	2.9534854
DOD*CLI	-0.003116	0.001484	-2.10	0.0431	400.82311
POP*CLI	5.9371676	1.47015	4.04	0.0003	948.68746
LAT	-32594.29	20474.08	-1.59	0.1204	36.813358
POP*LAT	-9.85952	1.859493	-5.30	<.0001	192.38783
PCP	-18849.49	8111.601	-2.32	0.0261	130.28675
ACT*PCP	0.0122576	0.003322	3.69	0.0008	149.16365
LAT*PCP	412.18272	202.7231	2.03	0.0497	214.3862

a) Test the overall utility of the model (5 p)

b) Determine the statistical significance and the nature of the overall effect of POP on RTD. (5 p)

c) The original variables did not exhibit any problems with collinearity. Are you surprised and/or concerned by the apparent high level of collinearity in this model? Explain.

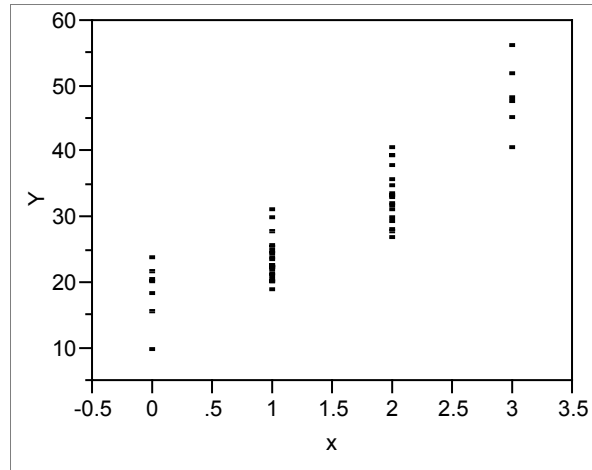
d) The R -student for California is 5.32. Does this value justify the removal of this observation from the analysis? Explain.

e) Speculate as to why JMP does not produce a test for lack of fit for this model.

f) Someone criticizes your analysis by pointing out that the ratio of regressors (i.e. terms in the model) to observations (k/n) is too large. Briefly respond.

g) Your boss is confused about the results and asks you to run the model without interactions. After analyzing the main effects it is determined that collinearity is not a problem. Using all the types of stepwise selection, you find that the only significant one is ACT with a slope of 0.2824 and $R^2 = 0.3698$. Write a short report describing the pros and the cons of these two models and give a recommendation.

2. A certain data set contains the results of an experiment. A certain drug was tested at four levels and multiple subjects at each level. The number of subjects observed at level i were n_i for $i = 1, 2, 3, 4$. A scatterplot of the data is shown below.



Delineate the ANOVA table (including lack of fit) for the following models:

a) a linear model

b) a quadratic model

c) a cubic model.

d) Would you consider any of the above models the "saturated" model? Why.