

# TEST 2 STAT 572

## Spring 2000

1. The city of Concord, N. H. needed to reduce the consumption of water during the summer. At the beginning of 1981, the city launched a campaign aimed to reduce the water usage. To investigate the effect of the campaign and to understand the factors associated with water usage, a survey of 496 households was conducted. Among the variables measured were:

water81    Summer 1981 water use  
 income    income in thousands  
 edu        education in years  
 retire     is head of household retired? (coded as no and yes)  
 peo81     # of residents in 1981

A model was run with water81 as a function of the other variables. The lack of fit test was significant ( $p = 0.07$ ) and it was decided to transform the response. The square root transformation of the response (water81 X) achieved no lack of fit, so it was used in the analysis. After using the stepwise procedure the results of the fit were:

Response: water81 X

Summary of Fit	
RSquare	0.471846
RSquare Adj	0.467543
Root Mean Square Error	911.1209
Mean of Response	4605.157
Observations (or Sum Wgts)	496

Lack of Fit				
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack of Fit	301	253199241	841193	1.0351
Pure Error	190	154400158	812632	Prob>F
Total Error	491	407599399		0.4001
				Max RSq
				0.7999

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob> t	VIF
Intercept	2818.1406	239.7265	11.76	<.0001	0
income	38.994979	7.307745	5.34	<.0001	5.429536
edu	-25.75091	14.14693	-1.82	0.0693	1.1398542
peo81	501.59923	52.98677	9.47	<.0001	4.597545
income*peo81	-3.780166	2.020818	-1.87	0.0620	11.011723

Effect Test					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
income	1	1	23637564	28.4741	<.0001
edu	1	1	2750507	3.3133	0.0693
peo81	1	1	74392795	89.6146	<.0001
income*peo81	1	1	2904820	3.4992	0.0620

a) Test the significance of the overall model (10 p)

b) Interpret the effect of education on water consumption (5 p)

c) Explain carefully how income and peo81 affect the water use (10 p)

[Hint: Use pictures] [For your information, the average income was 23 and the average peo81 was 3]

d) If the regressor retire is added to the model, the residual sum of squares is reduced by 740,543. Should this regressor be kept in the model?

The following information was obtained from observations 118 and 125:

<b>Observation</b>	<b>Leverage</b>	<b>DFBETAS</b>
118	.1229	-.4155 (for peo81)
125	.0871	.4088 (for income)

f) Which of these points can be considered to have high leverage? (5 p)

g) Interpret the values of the DFBETAS for the two points. (10 p)

3. Determine the truthfulness or falseness of the following statements. Carefully justify your answers and explain any assumptions made. (5 p each)

T      F      A large value of COVRATIO indicates that the corresponding regressor is significant

T      F      Autocorrelation causes the OLS estimators of the coefficients to be biased

T      F      An influential observation usually affects all the coefficients in a similar way

T F SSE(full model) > SSE(restricted model)

T F Collinearity is harmful to prediction.

T F If there are significant interactions then collinearity is usually strong