

# FINAL EXAM STAT 461

Fall 1996

- An accountant was interested in finding out which accounting variables were associated with the probability of bankruptcy. He collected data on 46 companies. The variables collected were four accounting ratios and the response took on the value of 1 if the company were bankrupted and 0 if not. The results were as follows:

**Response: Bnkruptcy**

### Iteration History

Converged by Gradient

### Whole-Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	17.943731	3	35.88746	<.0001
Full	13.766907			
Reduced	31.710637			
RSquare (U)		0.5659		
Observations (or Sum Wgts)		46		

### Lack of Fit

### Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-5.0723381	2.0686689	6.01	0.0142
CF/TD	5.79716941	3.0189027	3.69	0.0548
CA/CL	3.29225511	1.0848813	9.21	0.0024
CA/NS	-2.930697	3.0360868	0.93	0.3344

### Effect Test

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
CF/TD	1	1	3.6875146	0.0548
CA/CL	1	1	9.2092127	0.0024
CA/NS	1	1	0.9317802	0.3344

a) Is there evidence that the accounting variables are useful in explaining bankruptcy?

b) Which of the variables seems to have a larger effect on the response?

c) Interpret the value of the coefficient for CA/CL.

d) Estimate the probability of bankruptcy for a company with the following values:  $CF/TD = 0.10$ ,  $CA/CL = 2.0$ ,  $CA/NS = 0.50$ .

e) For values of  $CF/TD = 0.10$ ,  $CA/NS = 0.50$ , what is the value of  $CA/CL$  that will give the company a 0.20 probability of bankruptcy?

2. The following model was fitted to data on record times in 1984 for 35 Scottish hill races. The response is the time to run the race (minutes) and the regressors are the distance (miles) and the height from the start to the finish points of the race (feet).

**Response: Time**

**Summary of Fit**

RSquare	0.939185
RSquare Adj	0.933299
Root Mean Square Error	12.92297
Mean of Response	57.87954
Observations (or Sum Wgts)	35

**Lack of Fit**

**Parameter Estimates**

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	9.4117016	6.87815	1.37	0.1810
Distance	4.148492	0.835143	4.97	<.0001
Height	-0.000981	0.004164	-0.24	0.8153
Distance*Height	0.0009835	0.000307	3.20	0.0031

**Effect Test**

Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Distance	1	1	4120.8216	24.6751	<.0001
Height	1	1	9.2647	0.0555	0.8153
Distance*Height	1	1	1715.0088	10.2693	0.0031

a) Explain the effect of distance on the time for a "steep" race (height of 2500 feet).

b) Do the same as a) for a "flat" race (height of 800 feet).

c) Does height affect the time? Explain carefully.

The following are the values of leverage and Cook's D for two observations

<b>Observation</b>	<b>Residual</b>	<b>Leverage</b>	<b>Cook's D</b>
11	11.32	.710	1.73
18	56.10	.102	.60

d) Explain the role of observations 11 and 18. For example, if either of these observations are deleted from the model, what would be affected?  
[Be as thorough as possible].

3. Determine the truthfulness or falseness of the following statements. Make sure that you justify your answer.

T F      If an interaction effect is significant, the main effects involved should also be included in the model

T F      The independence of errors is frequently violated with time series data.

4. Suppose that we want to fit a piecewise linear model with one knot. Explain carefully how this is done when:

a) The knot is known

b) The knot is unknown and is estimated from the data.