Department of Statistics

COURSE STATS 330

Assignment 3, 2007

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on Thursday 13th September.

Question 1

The data in the file **`cows.txt`** contain the results of livestock sales. Each line in the file relates to the sale of a single cow. The variables are

Price: Price of the cow in dollars.

Age: Age of the cow in whole years

Bred: Has the cow been "bred" ? (i.e. is it carrying a calf, which is included in the sale for free) 1=Yes, 0=No

Angus: Is the cow an Angus (a breed of cow favoured for lean meat) 1=Yes, 0=No

Frame: Is the cow large (i.e. large frame?) 1=Yes, 0=No

Weight: Weight of the cow (00's pounds i.e. 10.0 = 1000 pounds)

Cond: Is the cow conditioned (i.e. well fed) 1=Yes, 0=No

Reg : Is the cow registered (ie certified as a purebred cow) 1=Yes, 0=No

In this assignment, you are required to develop a prediction equation to predict the price of a cow, given the values of the other variables. To give the assignment a little added spice, there will be a prize for the best prediction, based on the prediction error for an actual cow. In the event of a draw, the winner will be chosen at random from those having the equal best prediction.

Specifically:

(a) Read the data into R, checking the data as usual. The data are in the file **`cows.txt`** is on the course web page. [5 marks]

(b) You are required to fit a suitable model to these data, taking note of the following points:

- Are all the variables required in the regression? Use variable selection techniques to choose a suitable subset if not. For your chosen subset:

- Is the relationship between the response and the explanatory variables linear? If not, can a suitable transformation be made?
- Are there any outliers or influential points in the subset?
- Is the normality assumption satisfied?
- Is the equal variance assumption satisfied?

If you see any problems take appropriate corrective action until the model is satisfactory. [15 marks]

(c) Use your final model to predict the price of a 3 year old cow weighing 1100 pounds with the following characteristics: Bred =Yes, Angus=Yes, Frame=No, Cond=Yes, Reg=No. Express your prediction in the form of a 95% prediction interval.
[5 marks]

Question 2.

How big do you expect the maximum externally studentized residual to be in a regression where all the usual assumptions are satisfied? In Lecture 11, we mentioned in passing that the externally studentized residuals have a t-distribution with n – k – 1 degrees of freedom. Can we use this fact to get an idea of how big the maximum residual might be?

Here is a suggestion: The residuals are slightly correlated, but not excessively so – we will ignore this in our approach.

Let's take the case of a regression with 100 observations and 5 independent variables, so that n = 100 and k = 5.

You are required to do the following:

a) Draw a sample of size 100 from a t distribution with n-k-1 = 94 degrees of freedom, and calculate the maximum of the absolute values of the residuals. This is simple in R: use the code

```
max(abs(rt(100,94)))
```

Repeat the process 10000 times, saving the result each time. See Tutorial 4 for some ideas on how to do this. [5 marks]

b) Summarize the results. What do you conclude? What magnitude would you expect the largest residual to have? [5marks]

c) Would you be surprised if you got a maximum externally studentized residual of 3.0 in a regression with 100 observations? Would this indicate any problems with the regression? [5marks]

Question for bonus marks: We have ignored the correlation between the residuals in our simulation approach above. Given a specific regression, can you suggest how we could modify our approach to take these correlations into account? [5 marks]

Data for Question 1 follow:

Price	Age	Bred	Angus	Frame	Weight	Cond	Reg
1000	3	1	1	1	10.15	1	0
1250	3	1	1	1	11.00	1	0
980	5	1	0	0	11.15	1	0
1015	4	0	1	0	11.00	1	0
995	5	1	0	1	10.00	0	0
825	7	1	0	0	9.80	0	0
850	6	1	0	0	10.25	0	1
1150	2	1	1	0	10.50	1	1
1150	2	1	1	1	10.75	1	1
1200	3	1	1	1	11.75	1	0
1200	2	1	1	1	11.60	1	0
1000	6	1	1	0	11.00	1	0
1000	7	0	1	0	10.00	0	0
1050	6	1	0	0	10.00	0	1
1075	4	1	1	1	9.90	0	1
1165	5	1	1	0	11.35	1	0
780	2	0	1	0	8.85	0	0
800	2	0	1	0	9.50	0	1
1180	3	0	0	1	11.45	1	1
1000	4	0	0	1	11.45	0	1
1200	2	1	0	1	11.50	1	1
1000	3	1	1	0	10.00	1	0
1025	3	1	1	0	9.75	1	0
1175	2	0	1	1	11.35	1	1
800	5	1	0	0	12.00	1	0
915	4	0	1	0	11.85	1	1
1185	6	1	1	1	12.00	0	0
1020	5	1	1	0	11.25	0	0
775	7	1	1	0	12.00	1	0
850	7	1	0	1	12.00	1	1
1200	2	1	1	0	10.00	1	1
1200	3	1	1	0	10.15	1	1
775	8	1	0	0	12.50	1	0
775	7	1	0	0	11.85	1	0
1200	3	1	0	1	11.85	0	1
1135	4	1	0	0	12.00	1	0
1000	7	1	0	0	11.50	1	0
1185	3	0	0	1	12.00	0	1
1155	3	1	0	0	11.85	1	0
1155	2	0	0	1	12.00	1	1
1175	2	1	1	1	11.75	1	0
1200	2	1	0	1	11.50	1	0
1165	3	1	1	0	11.65	1	0
1000	6	1	1	0	12.25	1	0
1200	2	0	1	1	10.25	0	0
1175	2	1	0	0	10.25	0	1
1000	5	0	0	0	10.00	0	1
1125	5	0	0	0	10.00	1	0
1150	4	1	1	1	12.25	1	0
1125	2	0	0	1	11.25	1	1
875	6	0	0	0	10.00	0	0
1000	6	1	0	1	11.25	0	0
1185	3	1	1	0	10.00	0	0
1185	4	1	1	0	10.00	1	1

1000	4	1	1	0	11.00	1	0
950	4	1	1	1	10.25	0	0
875	6	1	0	0	11.25	1	0
775	6	0	0	0	10.15	0	0
1100	2	0	1	0	9.85	0	1
1125	2	0	1	0	10.35	1	1
1200	2	0	0	0	10.65	1	1
1175	4	1	1	1	12.30	1	0
1000	4	1	0	1	11.75	0	0
1000	3	1	0	1	11.85	0	0
1000	6	1	1	0	11.25	0	0
1000	6	1	0	1	11.25	1	0
985	5	0	1	0	10.50	0	0
1150	3	1	1	0	11.75	1	1
1150	4	1	1	0	11.50	1	0
1075	7	1	0	0	11.45	1	0
1050	6	0	0	0	11.50	0	0
885	6	0	0	0	10.00	0	0
1200	4	1	1	1	11.50	1	1
1150	3	1	1	0	11.50	1	1
1000	4	1	0	0	10.85	0	0
1075	2	1	1	0	10.85	0	1
980	4	0	0	0	12.50	1	0
980	3	0	0	0	12.45	1	0
1000	6	1	1	1	12.00	1	0
1085	3	1	0	0	11.25	1	0
1175	5	1	0	0	11.55	1	0
1150	5	1	0	1	11.15	0	1
1175	3	1	0	0	11.15	1	0
1000	2	0	1	0	10.00	1	0
875	7	1	1	1	11.65	1	0
995	5	1	1	1	11.55	1	1
1000	3	1	1	1	10.15	1	1
1250	3	1	1	1	11.00	1	1
1150	2	1	0	1	11.00	1	0
1150	5	1	0	1	10.85	1	0
1200	5	1	0	1	10.95	1	1
980	8	1	1	0	12.75	1	0
995	3	0	0	0	10.00	0	0
1000	3	1	1	1	10.35	1	0
1165	3	1	0	1	10.35	1	0
1175	5	0	1	0	11.25	1	1
1000	6	1	1	1	11.20	1	0
775	5	0	1	1	10.00	0	0
1000	3	1	1	1	11.65	1	0
1015	4	1	0	1	11.50	1	0
1200	4	1	1	1	11.50	1	0
1175	6	1	1	1	11.75	1	1
1095	4	0	1	1	11.75	1	0
1100	3	1	1	0	12.35	1	1
980	3	0	0	0	10.00	0	0
1110	5	0	1	1	10.35	0	0
1200	2	1	1	1	11.00	1	0
1185	4	0	1	1	11.25	1	0
1185	3	0	1	0	11.00	1	0
1170	4	0	1	0	11.25	1	0
1150	3	1	0	0	11.10	1	0
1000	4	1	1	1	11.00	1	1
975	6	1	0	0	10.00	0	0
1200	4	0	1	0	11.00	1	0