STAT 400 Data Analysis October 10, 2003

Today: Level Plots of Surfaces
Wireframe Plots
Example Midterm

Level Plots – p247
Wireframe plots – p 250

Midterm from last year:

Test from last years version of this course, with answers:

1. In the "ubi" assignment (A1), why was an ordinary one-way analysis of variance an inadequate method for studying the dependence of the shape index on species?
Ans. Mainly because the species effect was confounded with spatial position. Lesser items that might be mentioned are the differing variabilities of the species and the skewness of the index distribution. (full marks for the main idea).

2. In the "Mercedes" assignment (A2), what was the role of the loess procedure in forecasting the future?
Ans. The main use of loess was to make clear the seasonal effect – another possible use was to focus attention (via the residual) on the seasonally adjusted data, especially for the last few data values.

3. a) In the "histo" assignment (A3), describe the weighting function that was used in the original program for producing a density estimate.
b) What was the advantage of using all the data to estimate the density at each grid point?
Ans. a) It was 1 within d of the grid point, and 0 outside of this interval.
b) The weights could be adjusted so that they were less drastic than 1 or 0, starting at 1 perhaps, and decreasing gradually to 0. This allowed for a more flexible procedure, which includes the 0-1 weighting as a special case.

4. In the "bimbo" assignment (A4), how was the guessed demand distribution checked against the data (just the general strategy, no details needed)?
Ans. The guessed demand for 53 days was compared with the actual 53 delivery quantities to determine the 53 simulated sales, then this simulated sales distribution was compared with the actual sales distribution using ecdf plots.

In each of questions 1 to 4 above, my aim was to find out if you understood the basic idea underlying the exercise.

5. Refer to p 16, Fig 2.1. Identify the largest quantile in the Alto 1 group by giving its coordinates.
Ans. There are 35 points, so n=35. the largest f=34.5/35 = .986. The ordinate is 72,

so the coordinates are (.986, 72).

In this question, I just want to know if you know how a quantile plot is constructed.

6. Refer to Fig 3.7 on p 95. Where would the dashed vertical lines lie in Step 1 if the loess were to be computed at x=20 and for alpha = 0.3?
Ans. At 17 and 23.

We spent quite a bit of time in class going over the detail of the loess procedure illustrated on p 95. I wanted to know if you understood this.

7. Refer to Fig 3.59 on p 148.
a) Why are there two loess fits in this diagram?
b) Why is jittering used in this display?
c) Would a Q-Q plot of this data be a useful additional display? Explain.
Ans. a) Because of the possible reversal of the "X-Y" roles in this case.
b) So that measurements made identical by rounding can be seen as separate points.
c) No, because a Q-Q plot is useful for independent samples of commensurate data.

a) and b) were straightforward, but part c) was tricky. You had to think about the use of the Q-Q plot following Cleveland's introduction of it pp21-25. However, there was only 1 mark for part c).

8. Refer to pp 185, Figure 4.3.
What does one look for in a graph of this type?
Ans. Interaction – and if so, what is it?
I wanted to know if you knew why we were spending so much time considering coplots. I certainly mentioned the search for interactions being what coplots were so good at, several times.

9. Refer to pages 198-199. Explain the relationship between Fig 4.14 and Fig 4.13.
Ans. Each vertical line in Fig 4.14 corresponds to a panel in Fig 4.13. 4.14 is the plane over which a loess surface has been fit, and the intersection of this surface with a vertical plane on one of the vertical lines in 4.14 produces the curve in a panel of 4.13.

I had to ask something from Monday's lecture, and this was closely related to the main topic discussed. That is, fitting a loess surface to the data, and then using this fitted surface to draw the coplots. Some people asked me to explain this further after class, and this was a good idea. Unless one has the main idea (coplotting the fitted surface), there is no way to understand the relationship of the two graphs.

10. What name would Cleveland give to the pattern shown in the following graph?
Ans. Monotone spread.

Monotone spread is explained on p 47 and p 50. However, I admit that the book is

not too clear on this. But I did describe it in class. This is a simple idea that is most often solved by taking logs of the data, and something you should have run into in other stat courses. And if you didn't understand it in the book, you should ask me (or Pritam) about it.

LET NO CONFUSION SLEEP – WAKE IT UP BY ASKING QUESTIONS!