STAT 400 Data Analysis October 6, 2003

Today: Feedback on Bimbo:
Level Plots and Contour Plots. (pp 228-248)
Tips for November mini-lecture
Next Assignment (coplots on galaxy data)

Bimbo Assignment Feedback:

Day	Mean	SD	Ratio for Max Profit
Thurs	143	40	1.85
Fri	115	38	1.5
Sat	145	32	1.4
Mon	130	50	1.8
Tue	103	17	1.4
Wed	110	20	1.4

What would communicate to the company on the basis of this?

1. Check cost parameters – rerun if necessary
2. Suggest increasing deliveries 40% (slope to right of max is steep)
3. Suggest more studies of other products!
4. Anything else?

Assignment was well done – couple of pitfalls:

Seasonal adjustment on the basis of 1 year is rough, and needs an lowess with a big alpha.

Adjustment for seasonality – location of result must be given to program

Variability of profit might depend on ratio – another consideration.

Decimal places – please be reasonable!

Q-Q plot – what shape does it have when one variable more skewed?

Level Plots and Contour Plots:

Here is the galaxy data used for this part of the text (whole file sent in e-mail):

E-W	S-N	angle	radius	velocity
-3.53	15.91	102.50	-16.30	1501
-4.03	18.16	102.50	-18.60	1491
-4.52	20.40	102.50	-20.90	1481
-5.02	22.65	102.50	-23.20	1468

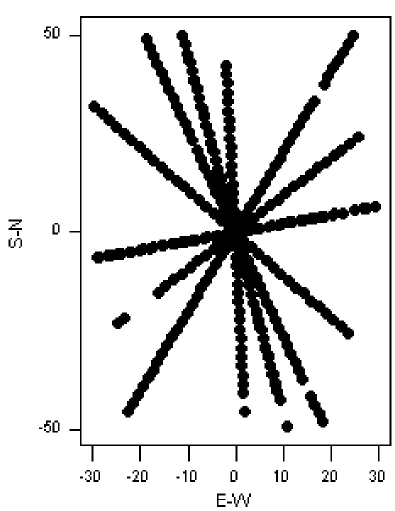
-5.52 24.90 102.50 -25.50 1455
.... and over 300 more rows

Imagine a spinning object, like a ball, that is receding from you. If you had a way of measuring its velocity along any line of sight, you could measure the velocity of each point on the ball at any instant. Would all such lines of sight return the same velocity? No. At one instant parts of the rotating ball are moving towards you relative to the centre of the ball, and other parts are moving away from you. So the line-of-sight velocity measurement will be a bit more or less than the velocity of the centre of the ball.

Now think of similar measurements of a galaxy – can we use the velocity measurements to detect the direction and speed of rotation?

The directions for which velocity measurements have been taken for the galaxy NGC 7531 are shown here:

Fig 4.43 - aspect ratio=2.2



At each point in this graph, there is a velocity measurement which might be represented as a coordinate measuring height above the page.

The velocities vary from about 1400 km/sec to about 1800 km/sec.

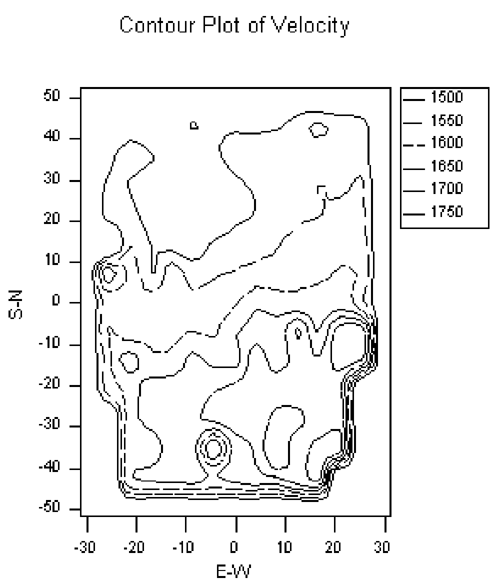
How can we graph this data? One way is with a "level plot". See p 231.

This is not a coplot – in which conditioning is usually on a factor or a measurement to be considered a factor. This is an instance where the E-W and S-N variables are not measurements but values selected by the experimenter, and the velocity is clearly the response, the dependent variable. Nevertheless, we can use the idea of conditioning to produce the level plot, by conditioning on the response.

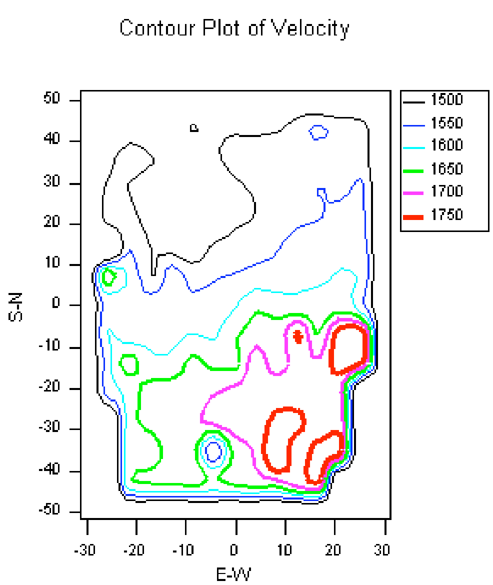
The bottom left graph shows which directions lead to the lower velocities small on E-W and big on S-N, whereas the upper right panel shows where the high velocities come from (the opposite side of the galaxy). We can already tell which way the galaxy is spinning ....

The data is collected along concentric lines of directions ("slits" as Cleveland describes them). It is possible to describe each direction in terms of polar coordinates – angle and radius. See p 233 for a coplot using of velocity against radius conditioned on slit angle. Can you see why these plots are consistent with the level plots?

What is Fig 4.46 on p 235. It is the same coplot as p 233 but based on the loess fit to the surface rather than individual loess fits to each slit's data. However, the resulting residual plot is not further analysed.



This can be improved with colour control and line-width control.



Why are the contours not as smooth as in Cleveland's p 239?

The loess here not robust?   The fitted surface not smoothed as much.

Which one is better?

Depends what you want to study ….

Spin only – Cleveland is better

Other anomalies – Minitab default is better

Why are we using level plots and contour plots instead of coplots?

The role of the variables is important – with this data, both E-W and S-N are selected by the experimenter. These two variables are really just arbitrarily selected coordinates – there is no interest to separating out the effect of one or other of these independent variables. We really want to look at the map of velocities across the galaxy. Level plots And contour plots are better for this. Read the last para on p 228 for more.

There are some interesting details involved in forming a contour plot from a surface that is defined only by a few points. When we use loess to fit a surface to some 3-D points, we do it on a 2-D grid and so have the height above the surface defined at each grid point. Note that these loess points are now treated as on the surface itself, not merely something we are fitting to. So the contour plot construction reduces to figuring out where a contour of a given height crosses through a grid-square. See pp 240-242.

Once we have a height (a z-value) for each grid point in the 2-D plane ((x,y) say), the procedure is:

1. Identify squares for which the contour must cross the interior.
2. Use linear interpolation to place dots on the edge of the square where the contour would cross.
3. Connect the dots. (If too jagged, increase number of grid points).

Note: Ignore ambiguity problem discussed on p 242-243. Increasing the grid density should avoid this.

Limitations of Contour Plots:

1. Flat areas are hard to display
2. Speed bumps and potholes look the same

Level Plots from smoothed surfaces:

Note that the level plots discussed so far were level plots based on the data directly – we did not estimate a response surface and then use level plots of this surface. Recall the level plots of p 231. If we instead use a loess surface fit, which before any interpolation would really be a revised (and smoothed) trivariate data set, with as many rows as there are grid points.

We use this smoothed data set to draw the plots on p 247. Each panel corresponds to a very small range of velocities (of width (1775-1409)/48 = 7.6 km/sec). In the original level plot on p 231, we had to use much larger ranges. Also, we can increase the grid density until the curve formed by joining the dots with straight lines looks like the curves on p 247.

From Fig 4.55, we can again easily see that the velocity in the North-East portion of the galaxy is minimal and increases as we move across the galaxy toward the South-West. (The directions E and W are backwards from the usual since we are looking outward at the sky.)

Some tips for your November mini-lecture (7 minutes!)

Some topics you might choose:

Coplots
Loess – univariate data
Loess – extension to surfaces
Contour Plot
Level Plot – univariate data
Level Plot – extension to surfaces
Banking to 45 °
Kernel estimates of density
Dotplots
Quantile plot
Q-Q plot
Normal Probability Plot
Box Plot
Log transform
Monotone Spread
Robust Fitting and Bisquare
Slicing
Cropping
Seasonal adjustment
Brushing
Residual Analysis
Interaction

Please prepare a short handout to go with your seven-minute presentation – MAXIMUM length is one side of one page. This is a strict limitation! Shorter is OK.

Your objective will be to make the idea very clear to your peers.

Assignment 5: Analyze galaxy data using coplots. (code, graphs, and words necessary in report.) How does this compare with the level plot analysis for revealing important information about the data? This will not be due until Wednesday, October 22, 2003. (5 days after midterm). It would be a good idea to learn how to do this in R or Splus – it is possible in MINITAB but more difficult.