STAT 400 Data Analysis October 3, 2003

Today: More Ch 4 – Trivariate Data

Direct Manipulation: Brushing for example (p 192-3)

Fitting surfaces – parametric and nonparametric (p 194-5)
Loess works on surfaces –but one needs to deal with possible different ranges in the two (x&y) directions. One way is to standardize variables X and Y and then use Euclidean distance in the plane the same way we used 1-Dim distance for the simpler loess (as on p 95). Or, just use a factor based on range.

Result of fit – how do you visualize it? See Fig 4.12 p 197. Use coplot again, but this time the data is suppressed and the surface at the conditioned values of X allow the relationship of Z as a function of Y to be visualized.

Can you re-create the picture of the surface from the figure on p 197? Needs practice.

Use "surface" macro on this ethanol data. Look at the smoothing method in program.

Fig 4.14 – There are only 5 CRs – how can there be 16 curves in Fig 4.13 (p 198)?

Pp 200 ff: Concerns fitting the surface for the "rubber" data and plotting the coplots of the fitted service. The example is different than for the ethanol data because there is less data (30 points as opposed to 88 rows), and also because the "data rectangle" does not include all the data. The small sample suggests a parametric model may be useful, and the empty space in the data region suggests "cropping" is needed before the coplot is constructed.

Cropping: A technique for restricting the data region of examination by coplots – the fitted surface however still uses all the data – it's just the display of the variable relationship that needs to be restricted – the reason is that in regions where there is no data, you do not want to be deceived about the nature of the fitted relationship! See Fig 4.15 p 201.

The parametric fitting method in this case is not so important – the important thing is that for a small data set, the parametric fit is sometimes helpful.

The absence of interaction in Fig 4.17 is a result of the parametric model – the residual analysis will check if the model was adequate.

Pp 204-5 discuss the method for extracting the conditional plots from the fitted surface. This is a detail you can skim over. Note that the step of filling in the surface from the estimated loess fits on a rectangular grid is not discussed in detail. How should this be done? (triangles).

Pp 206-211 show that a coplot of residuals shows more than ordinary residual plots. And also, it demonstrates that a pattern in the residuals gives an opportunity for improving the fit
(improving the model).

Pp 212-217 more of the the same techniques.

Pp 218- 227 The experiment leading to “banking to 45° ” – Another example of coplot/loess and residual analysis.

Section 4.6 p 228ff Level Plots – an important new topic. Condition on response variable!

Interesting example – data on galaxy rotation.