STAT 400 Data Analysis October 7, 2005

Today: Ch 4 of Cleveland – review pp 181-189 and continuation of Ch 4

But first,

What did you learn from the Bimbo assignment?

1. Simulation is not only for exploring theory, but useful for data analysis.
2. Use of Intuition (Subjectivity?) is useful and inevitable in modeling.
3. Graphical methods are often the best way to portray results of analysis.
4. Simple probability methods can be used to solve practical problems.
5. Programming requires good knowledge of syntax but modifying a program is not so demanding.
6. Doing anything properly on computer takes more time than anticipated!

Back to Ch 4:

We introduced coplots (conditioning plots) to study the dependence of the relationship between two variables on a third variable. (Rubber data, p 185 and p 187)
Quantitative variables for this.

Possible confusion – alpha for loess in coplot (proportion of data used per grid point)
_ alpha for slices for coplot (proportion overlap in adjacent slices)

Much more informative than matrix plots (p 183) – signal amplification – how?
Ability to visualize the nature of interactions – much more flexible and informative than analysis of variance.

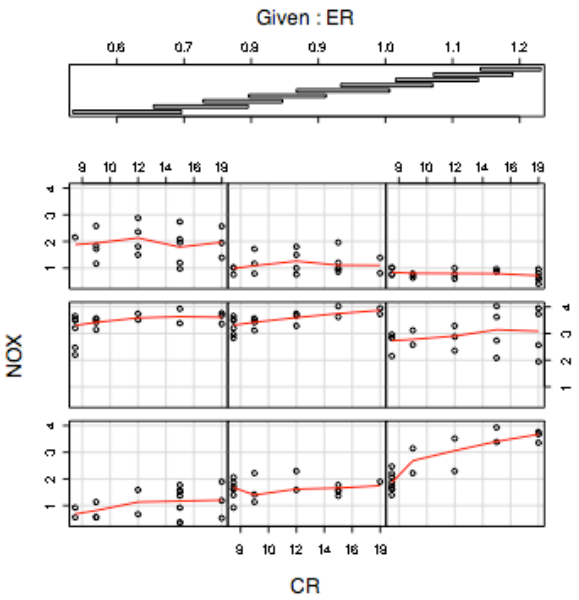
Rubber data shows little interaction between Tensile Strength and Hardness in predicting Abrasion Loss.

Compression data (p 188): understand as well as possible from matrix plot

Coplot on p 189 shows strong interaction between ER and CR in predicting NOX.
Review definition of interaction. Check that you understand the following:

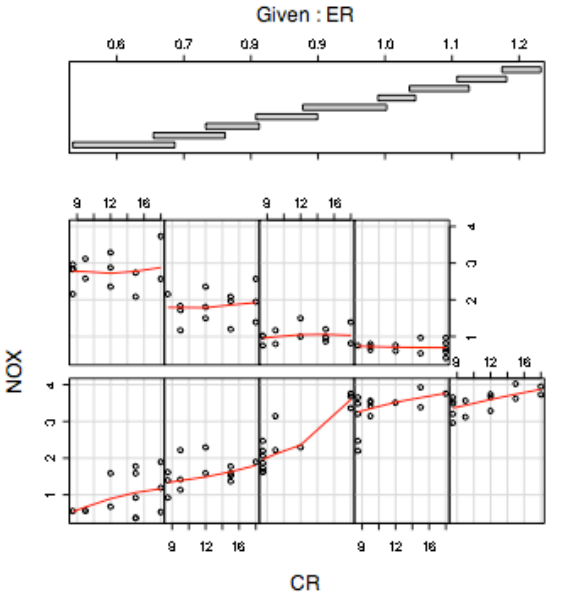
P 191: If the 5 lower panels, or the 4 upper panels (but not both), were all the data you had, you would conclude that the interaction between ER and CR in predicting NOX was negligible.

```
attach(ethanol.df)
coplot(NOX~CRIER,number=9,panel=panel.smooth,overlap=.5)
```



We can get a bit closer to the books graph 4.6 on p 189

coplot(NOX~CR | ER,number=9,panel=panel.smooth,span=1,overlap=.3,rows=2)



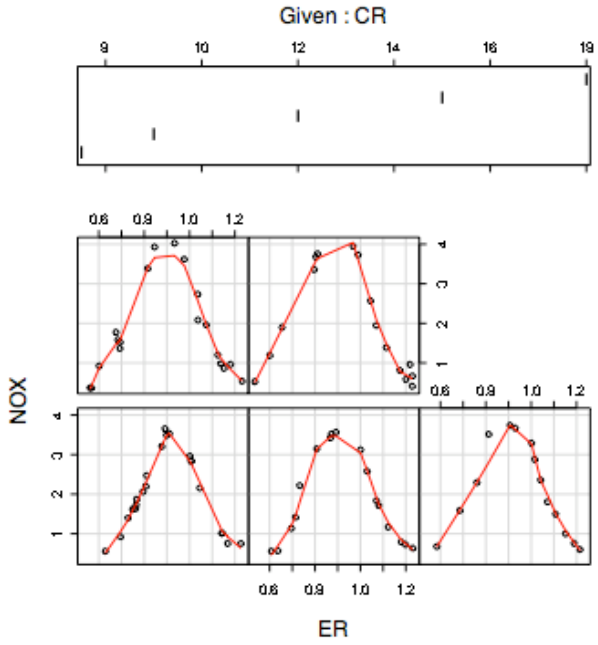
The coplot of the Compression data on p 190 should also show interaction – can you see it?

You need to specify

> int

```
     [,1] [,2]
[1,]  7.5  7.5
[2,]  9.0  9.0
[3,] 12.0 12.0
[4,] 15.0 15.0
[5,] 18.0 18.0
```

coplot(NOX~ER|CR,given.values=int,panel=function(x,y,...)panel.smooth(x,y,span=.4,...), number=5)



Its not quite like the one in the book – really needs panel.loess with degree =2 not panel.smooth

Result p 197. Shows a more smoothly described interaction

Brushing – Direct Manipulation

-----lecture ended here! But see exercise below:

Coplots of Fitted Surfaces: Suppose you fit a surface to the relationship LOX=f(ER,CR). How do you graph it? Later – 3D options (not perfect because of projection ambiguity)

see plots 4.12,4.13 for the ethanol data, 4.24, 4.25 for the rubber data.

Cropping: p 201 You questioned the blank regions in Fig 4.4 p 187. Note that if surface fitted, might look like we have info about surface here – but not justified. Hence crop for graph of fit.

pp 218-227 "45°" experiment. Of interest but no new techniques ...

Next pp 228 ff Level Plots, Contour Plots, Wireframe Plots

Exercise: (for Friday Oct. 14) Analyze the "ethanol" data (NOX,ER,CR) using parametric methods (regression and/or anova). Comment on the strengths and weaknesses of this method, compared to the coplot approach, for identifying and describing the interaction between ER and CR in predicting NOX.