

Today: Assignment 6

More Ch 5 – Ozone Data, Organ Data, Iris Data

Assignment 6: (Due Nov 3) Analyze the Ozone data using regression methods, with an aim of determining a good model for predicting Ozone Concentration from Solar Radiation, Temperature and Wind Speed. Use coplots to analyze the residuals. Provide a verbal summary of your results. (Note that the Cleveland text gives some information about this data – but your report is on your own analysis of the data.) Your report should include your final regression output and the coplots.

Ozone Data:

Coplots p 277-281 Know how to look for and describe the interactions in words (Other than by quoting the book!)

Coplots of Surfaces & Cropping pp 282-289 Cleveland suggests specifying a parametric form for the O₃ conditional on T, as a function of R and W. Quadratic polynomial.

What advantage gained by using partial parametrization?

Cropping is needed in representing the fit – one looks at the factor plane and cuts out empty areas – the technical problem is to realize how these limitations affect the ranges of the factor plane variables in the surface coplots. This is complex and you should just understand the need for it and not worry so much about the details. See p 287-289.

Residual Analysis: p 290-292 The s-l plot is a good way to check for monotone spread. But beware of the “range effect” - it can mislead you in this visual check. In this case it confirms that the cube root is the best transformation of the O₃ variable to reduce the technical problems that would be created by monotone spread.

Note the normal Q-Q plot on p 292 – the reason it was done was to see if bisquare was needed. What can you say about the distribution of residuals if this plot is concave (like a cup)? Ans: too “peaked”. How could you test this idea?

What is the purpose of the r-f plot on p 292? Graphical option to computation of R-squared, but better!

Organ Data, Iris Data: Example of use of brushing to improve utility of matrix plot.

Outlier in organ data determined to be an enlarged spleen (liver OK)

Iris Data – suggest how an index might be constructed to aid classification. (Species is the dependent variable here). Size is length x width. Shape is length/width. Think logs and the m-d plot. Gives Fig on p 300. Example of Improvisation – another way to say it is that the context of the data is used to do a sensible summary of the data. (in 2 D).