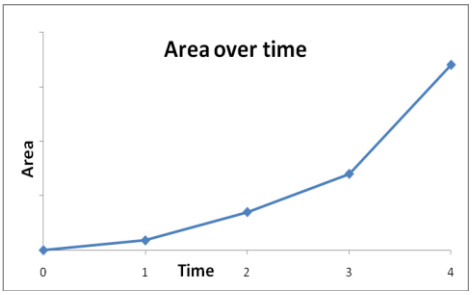
Lessons Not Learned at University #7

When simple linear regression analysis is actually simple

The Consultant was asked to assist *The Specialist* with the interpretation of the statistics in a post-graduate thesis for which *The Specialist* was the examiner. *The Postgrad* had done a quite detailed and complicated laboratory analysis looking at how various substrates, fungicides, temperature, humidity, fertilisers and so on had affected the spread on agar plates of a particular fungus. The study was well replicated and *The Postgrad* had carefully measured everything at exactly consistent intervals. The area of spread of the fungus was measured at each point in time and the rates of spread differed considerably. There were many alternatives to analyse. *The Specialist* was also a specialist in this particular fungus but had little statistical knowledge, that is why he asked for help.

The data for one plate looked like the graph, but this pattern was typical of all the data.



The first thing *The Postgrad* had noted was that there was a lovely straight line between the observations at times 1-3. The value at time 4 seemed aberrant so he culled all these data out of his data base. He then fitted a linear regression to the remaining three points as in:

$$Y = b_0 + b_1 X$$

It was a simple linear regression. The intercept value b_0 was generally significantly different from zero.

Now *The Postgrad* had a problem. He needed to analyse the multitude of alternatives and he lacked the essential multi-variate statistical skills. After all he was a mycologist not a statistician. So he prepared a myriad of tables and interpreted the results visually. The final analysis presented in the thesis could only be described as a dogs breakfast. But this had not fazed *The Specialist.* He was also a mycologist and he could make sense of the inferences that *The Postgrad* had made. *The Postgrad* gained his degree.

But, *The Consultant* noted, the student had discarded 25% of his data, basically on a whim. Regressions should be based on more than three data points. How could *The* Postgrad justify deleting 25% of the measurements? Further, how was it possible for the area of the fungus to be negative at age 0?

The answer was both simple and obvious. *The Postgrad* should have plotted the square root of the area, or distance, against time. Irrespective of treatment, the rate of fungus expansion across the agar plate should be linear. Looking at some of the data and changing Y to the square root of the area showed that the regression constant b_0 was not significantly different from zero. If *The Postgrad* had done that then he would have had a single variable regression through the origin, using all 4 measurement points. With a single standard error for the parameter estimate, and many replicates, a neat analysis of variance would have been possible that would have added greatly to the power of *The Postgrad*'s study.

The Consultant, who was not a mycologist, concluded that it was a great pity this had not been done.

Lesson: Statistics is an art, an arcane art to most, but if data are to be analysed, the analysis should be done properly. The application of fairly simple, but appropriate, statistical analysis techniques can generally ensure this occurs.

Jerry Leech

jleech@ozemail.com.au