Pages 339-358: Assuring Product Reliability and Safety. Exploratory Graphics. Reliability. Components in Series and in Parallel. Distribution Models (Normal, Weibull). Confidence.

This paper starts with the dramatic case of the failed Challenger Flight and its cause dues to some naïve statistical work. See Figure 1.

Q1. Why was the cold weather fatal to the Challenger Flight? Was this a "reliability" issue?

Q2: How is "reliability" expressed numerically?

Q3: See Figure 2 on p 344. Under what conditions is the reliability of an item the product of the reliabilities of each component in the item?

Q4: How was it possible to test the 10 year reliability of a washing machine in 6 months?

(We may return to the details of Fig 3 later in the course.)

Page 373-389: Advertising as an Engineering Science. Process of Data Based Information Collection. Bar Charts. Covariates. Percentages. Combinations.

The tone in this article is a bit like cheerleading for Engineering. However, I think the real engine in this article is Statistics! However the article is useful in the sense that it shows how real marketing people make use of basic statistics strategies.

The article is about an email experiment for advertising the association of Butterball Turkeys with Thanksgiving, presumably to support long term sales.

The factors examined in the study are:

1. Subject of email(Planning, Elegance, or Festive) - 3 choices
2. Day of week (midweek or Friday) - 2 choices
3. Weeks before Thanksgiving (3,2,1,0) - 4 choices

Design of the email Targets:
The three "Subjects of email" are used during the first three weeks (but all possible combinations are used including Planning, Planning, Planning and Planning, Festive, Planning, for example). Each subject gets 4 emails – one per week. In addition to the assignment of a certain subject order to a client, there is the assignment of "Day of Week" for each email (so one client might be assigned Planning-midweek in week 1, Festive-Friday in week 2, Festive-Friday in week 3, and Planning-midweek in week 4.) Any of the 6=3 x 2 combos is possible on weeks 1,2,3.) But on week 4 the

only combo that was allowed in the design is Planning-midweek and Planning – Friday.

Q5: What strategy was used to try to reduce the variability of responses due to known characteristics of the registrants?

You need to understand why there are 432 different email patterns. The basic rule in combinatorics is that if there are m ways to do something and n ways to do something else, then the combination of things to do can be done in m x n ways. This is sometimes called the "mn rule".

So if there are three ways to specify a subject and two ways to specify a "Day of Week" then there are 3 x 2 =6 ways to specify both. Also, if there are 6 ways to specify the subject-day combo on a particular week, there are 6 x 6 x 6 =216 ways of specifying the combo for the first 3 weeks. For each of these, there are just two possibilities for the "day" on the fourth week so the total number of possibilities is 216 x 2 = 432.

This is one of the few calculation ideas that you do need to know for this course.

Q6: How do you randomly assign the 57584 registrants (clients) to the 432 email streams?

Conceptually you need a way to assign a number from 1 to 432 to a registrant, and it has to be done independently from one registrant to the next. Like rolling a die with 432 faces! Actually, in practice, one uses a computer. It uses a list of the 432 streams, and numbers the list 1 to 432, then generates 57584 numbers in the range 1 to 432 to generate the stream characteristics for each registrant. Of course, there is still lots of work getting the appropriate emails organized and sent, and this is another computer job that we will ignore for our discussion.

Another big item of analysis that we will gloss over is how you take all the feedback (the click through info) for the 57584 registrants and summarize it. There is mention of "logistic regression" on p 386 – that is the main analysis engine – but the article and our discussion will stay with the simpler graphical summaries for details. The graphical summaries would have to be looked at anyway to make sure the "logistic regression" does not bury some important, but unexpected, information. Remember the basic rule of data analysis "First, plot the data".

The charts on pp 381-386 are called "bar charts" or "histograms" – they show frequency (or relative frequency or percent) for various values or groups of values of a variable. Note the logical difference between the bar chart for a categorical variable and a histogram for a quantitative variable. For the former only, rearranging the values of the variable would not destroy the graph.

Q7: Which graphs in this article plot categorical data frequencies and which graphs plot quantitative frequencies?

It is customary to call the categorical frequency charts Bar Charts and the quantitative frequency charts Histograms.   The reason for this is that the charts for histograms usually have the vertical bars touching since the quantity could be any value in the associated interval of the x-axis, and the appearance of "bars" is lost in this case. In this article, this convention is not followed – see the "Age" graph on p 383. The sequence graphs (labeled 1,2,3) are properly bar charts since numbers like "1.5" are not possible here.

Q8: In the interpretation of the plots on pp 381-386, there is no mention of p-values. Might the observed differences be due to chance and not reflect real differences?

Q9: Is the method used in this article "Engineering" or "Statistics" or "Something Else"?