Teaching Statistics:
Its not about Procedures, Its about Understanding Relationships

Session Facilitator and Moderator
Robert L. Andrews,
Virginia Commonwealth University, Department of Management,
Richmond, VA. 23284-4000, 804-828-7101, randrews@vcu.edu

Session Contributors and Panelists:
Barry Wray, University of North Carolina Wilmington
Weiyong Zhang, Old Dominion University
Dmitriy Nesterkin, Coastal Carolina University

ABSTRACT

The typical course description for a statistics class lists a set of covered analysis procedures. What most business students will need in their future is to understand relationships between variables and how to measure the nature of these relationships. Focus on procedures gets students mired in calculations without getting the big picture and leaves them inadequately prepared to analyze data to answer real business questions.

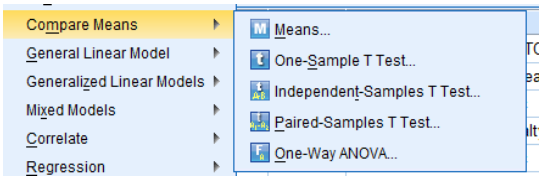
INTRODUCTION

Most of us want to do well when we do something, like teach a class, and doing well can easily take precedent over other objectives, like helping prepare students for their future. We can do best by focusing on the things we know best. Hence when it comes to teaching class, what we most often know best are the things and ways we were taught or the things we are currently doing in our research. Also textbook authors and publishers produce materials that match the way faculty teach their courses. The above logic and reasons support teaching statistics in the same way as we have been for years. But that does not mean that it is the best way to do it, certainly if determining best uses preparing students for the future as a major criterion.

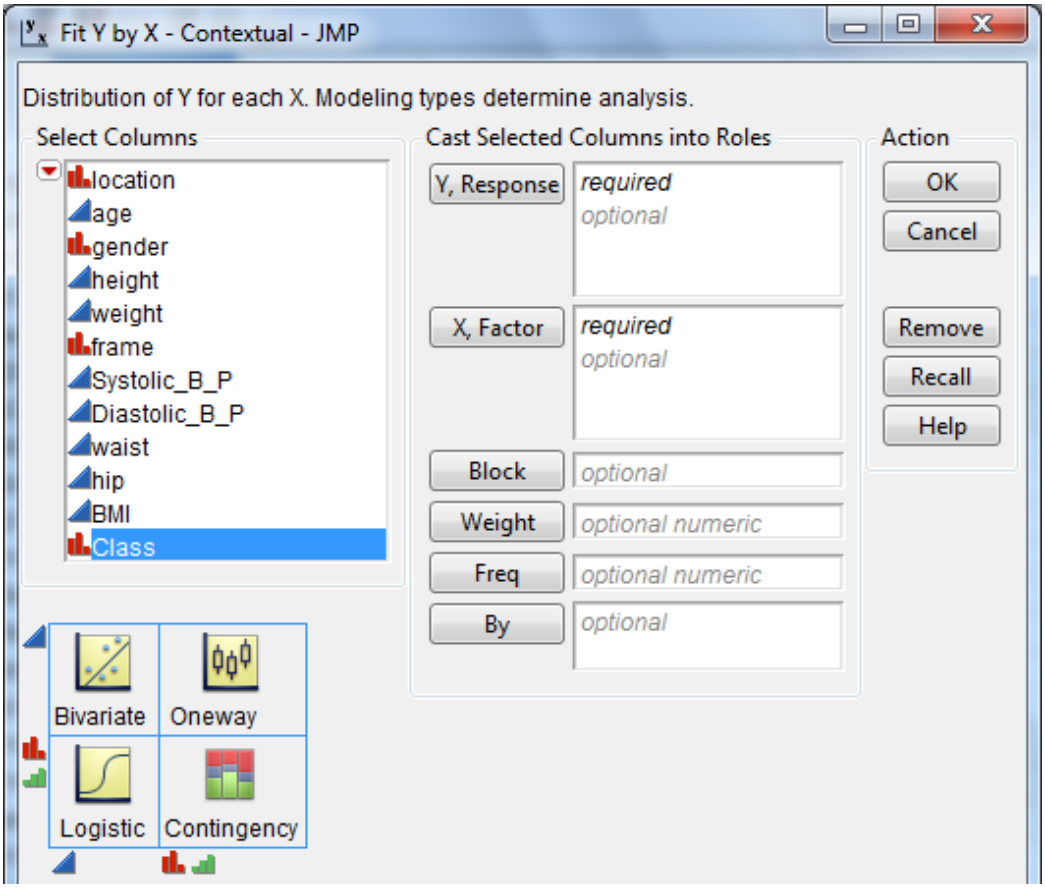
Since the publication of *Competing on Analytics* by Davenport and Harris in 2007, analytics has received increased focus and use in the business world. Businesses are using analytics to enhance company competitiveness by using data to improve decision making. These businesses need employees with business and analytical knowledge to interpret findings and develop relevant advice to guide their business decisions. In March 2010 IBM published *The New Value Integrator: Insights from the Global Chief Financial Officer Study* which found that the CFOs felt that a significant gap existed between skills required for today's business environment and the skills currently available in the workforce. This indicates that our present educational process is deficient for their perceived needs. For business students the business statistics class is one that has a clear focus on analysis of data. Historically this has included a set of analysis procedures listed in the course description with a primary focus on descriptive statistics, confidence interval estimation and testing for the statistical significance of hypotheses. Testing

of statistical hypotheses may match well with the analysis a faculty member is doing to get his/her research published but does not match well with the analysis that is part of analytics.

If the goal is to help prepare business students for their future then major consideration needs to be given to what they are most likely to be doing in the future. The chances are slim that they will be collecting data to test hypotheses so that they can get an article published in a refereed journal. We need to look at trends in business and the current focus on analytics is one clear trend. Looking at statistical software can also be helpful. SPSS is a software package that has been in existence since the days of computer punch cards. It has evolved into the point and click era of computing but the menu still has a list of classical procedures as shown in the SPSS image below which remains the same in SPSS 20.



JMP, a SAS statistical software product, is more recently developed. It tends to be more oriented to providing service to businesses with an analytics focus than what SPSS has done historically. Of course with IBM purchasing SPSS and its "Smart Planet" initiative, the look of SPSS may change. The JMP menu below focuses on the relationship between variables and automatically selects an appropriate procedure based on the type of variable for the response Y and the type of variable for the predictor or factor variable X.



Once the variables have been selected then JMP creates a graphical display corresponding to the appropriate one of the four quadrants in the box at the lower left in the above diagram based on the type of variables for X and Y. It may also provide some appropriate descriptive statistics and possibly hypothesis testing results. If the user wants output for a specific procedure then that can be requested in a second phase.

As can be seen, this approach to analysis focuses first on relationships between variables and the measurement scales of the variables. The type of variable for both X and Y defines or determines how one looks at the relationships relative to the type of graphical pictures used to describe the relationship and the mathematical models built to evaluate the relationship. The testing procedures also depend on assumptions associated with the particular type of mathematical model. Using this focus on the types of variables as a first step in analysis helps guide students in the selection of appropriate graphical representation and analysis procedures.

Webster West in a keynote talk for the 2012 Electronic Conference on Teaching Statistics emphasized that a primary goal in a statistics class should be to "get students excited about data." His point agrees with our point that the statistics class should start with data. We need to get students to think about potential or anticipated relationships between variables. Then we can teach them techniques for painting pictures of the relationships between the variables and the type of picture will depend on the type of variables. Next we can move to measuring the strength or magnitude of the relationship as indicated by the data. As with painting the picture, the type of procedure for obtaining strength of relationship measurements depends on the type of variables and how the data are recorded. To obtain the proper analysis the software user must know how the data have been recorded and how the software procedure requires that the data be recorded.

Table 1 contains a guide created to assist the data analyst (student) in selecting the proper testing procedure based on variable type and how data were gathered and recorded. This guide is laid out to be similar to and used like the matrix JMP has in the Fit Y by X menu. The user of the guide can find the appropriate located in the table corresponding to the types of the variables the user wants to use and how the data were obtained and recorded. Students have found this guide to be useful when they were given a scenario and asked to select the appropriate testing procedure for the situation described in the scenario. This type of problem is often found to be very challenging when students have only focused on the individual procedures in the typical text. Merely spending time on each of the individual procedures does not provide the student with the big picture and fails to truly enable the student to be able to use the learned individual procedures when faced with a real situation requiring analysis. It should be noted that Table 1 does include logistic regression, which is not covered in many business statistics texts. However, some do, such as *Business Statistics* by Sharpe, De Veaux and Velleman.

The classic statistics course focuses on determining statistical significance of an appropriate test that essentially tests H_0: There is no relationship between the variables versus H_A: there is a relationship between the variables. In the business arena the amount of data that may be available for analysis may make determining statistical significance of no real practical value because an extremely large sample size may identify statistically significant effects for effects that would not provide any real improvement on the business side of the operation. Hence

students need to learn to think about assessing the practical significance of a result in addition to the statistical significance.

Table 1: Guide for Selecting the Appropriate Testing Procedure

<table>
<tr><td colspan="5">Relationship Testing Procedures Based on Type of Variables [JMP procedure]</td></tr>
<tr><td rowspan="2">Type of variable for Y {Response}</td><td colspan="4">Type of variable(s) for X {Predictor(s)}</td></tr>
<tr><td>Categorical (2 categories)</td><td>Categorical (3 or more categories)</td><td>Quantitative (1 variable)</td><td>Quantitative (2 or more variables)</td></tr>
<tr><td rowspan="3">Quantitative (1 variable)</td><td>Difference in 2 Means
2 sample Pooled t
(= variances assumption)
(2 Independent Samples)
[Fit Y by X; Means/Anova/Pooled t]</td><td rowspan="3">Equality of k Means
ANOVA
(k Independent Samples)
[Fit Y by X; means/anova]</td><td rowspan="3">Linear relation between Y and X
Simple Linear Regression
(1 Y and 1 X for each data case)
[Fit Y by X; Fit Line]
or [Fit Model]</td><td rowspan="3">Linear relation between Y and Xs
Multiple Linear Regression
(1 Y and k Xs for each data case)
[Fit Model]</td></tr>
<tr><td>Difference in 2 Means
2 sample Welch's t
(≠ variances assumption)
(2 Independent Samples)
[Fit Y by X; t Test]</td></tr>
<tr><td>Difference in 2 Means
(Paired Sampling)
[Matched Pairs]</td></tr>
<tr><td>Categorical (3 or more categories)</td><td>Independence of 2 Categorical Variables
Chi-square test of Independence
[Fit Y by X]</td><td>Independence of 2 Categorical Variables
Chi-square test of Independence
[Fit Y by X]</td><td>Multinomial Logistic Regression
Not Covered in this Class</td><td>Multinomial Logistic Regression
Not Covered in this Class</td></tr>
<tr><td rowspan="2">Categorical - Dichotomous (2 categories)</td><td>Equality of 2 Proportions
(2 Independent Samples)
[No JMP procedure]</td><td rowspan="2">Independence of 2 Categorical Variables
Chi-square test of Independence
[Fit Y by X]</td><td rowspan="2">Prob(category) = f(X)
Logistic Regression
(Dichotomous Y and 1 X for each data case)
[Fit Y by X] or {Fit Model]</td><td rowspan="2">Prob(category) = f(X)
Logistic Regression
(Dichotomous Y and k Xs for each data case)
[Fit Model]</td></tr>
<tr><td>Independence of 2 Categorical Variables
Chi-square test of Independence
[Fit Y by X]</td></tr>
<tr><td></td><td>Categorical (2 categories)</td><td>Categorical (3 or more categories)</td><td>Quantitative (1 variable)</td><td>Quantitative (2 or more variables)</td></tr>
<tr><td></td><td colspan="4">Type of X or [predictor variable(s)]</td></tr>
<tr><td colspan="4">Cells with no shading denote that Data are in Independent Samples for each category</td><td></td></tr>
<tr><td colspan="3">Shaded Cells denote that Data are in matched pairs/blocks</td><td></td><td></td></tr>
</table>

SUMMARY

The panelists, with varied levels of experience and from business schools at four different universities, will present their views on these articulated issues relative to teaching statistics in a business school. In addition to their presentations, time will be provided for those attending the session to also enter into the dialogue.

The main focus on the session will be on what is being taught in a class and how it is taught. But the session title also allows for discussion of how establishing a relationship with the students in the class can help them be more engaged in the learning activities. The panelists will have the opportunity to share what they consider to be the type of relationship with their students that they try to develop and why.

REFERENCES

[1] Davenport, Thomas H. and Harris, Jeanne G., *Competing on Analytics: The Science of Winning*, Harvard Business School Press, Boston, MA, 2007

[2] IBM Institute for Business Value, *The New Value Integrator: Insights from the Global Chief Financial Officer Study*, IBM Global Business Services, Somers, NY, March 2010

[3] Libertore, Matthew J. and Luo, Wenhong, "The Analytics Movement: Implications for Operations Research," *Interfaces*, Vol. 40, No. 4 (July-August, 2010), pp. 313-324.

[4] Sharpe, Norean R., De Veaux, Richard D. and Velleman, Paul F., *Business Statistics*, Addison Wesley, Boston, MA, 2012

[5] West, Webster, "The Impact of Technology on the Teaching of Statistics," Keynote address, Friday May 18, 2012 the Electronics Conference on Teaching Statistics.