

# SAS/INSIGHT<sup>®</sup> Software or JMP<sup>®</sup> Software: A Comparison

Andy Mauromoustakos, University of Arkansas, AR

## ABSTRACT

At the University of Arkansas in the Dale Bumpers College of Agricultural, Food and Life Sciences the Agricultural Statistics Laboratory is responsible for teaching every year “service courses” on Statistical Methods, Experimental Designs, and Applied Regression. In the past couple of years, we have experimented using SAS/INSIGHT<sup>®</sup> and JMP<sup>®</sup> as the tool kits for carrying out the analysis covered in these courses. In this paper we are going to concentrate on the advantages and disadvantages of each in carrying out analysis, gaining insight in terms of the models being fitted and understanding better the results of the fit. We have found that both SAS/INSIGHT and JMP were excellent exploratory and analytical tools. They both helped students discover more, interact more and understand more as compared to previous years when students only used SAS<sup>®</sup>, SAS/STAT<sup>®</sup> and SAS/GRAPH<sup>®</sup> to perform these same analyses. In addition, we have found that both were better for checking the assumptions of the various models fitted and easier in the graphical presentation of the final results. We will also exploit the differences in terms of capabilities between the two and we will attempt to suggest the kinds of additional features and improvements that could enhance both products as tools for carrying “ALL” the analyses typically covered in these type of courses.

## INTRODUCTION

At the University of Arkansas in the Dale Bumpers College of Agricultural, Food and Life Sciences the Agricultural Statistics Laboratory is responsible for teaching every year service courses on *Statistical Methods* (i.e. “First Methods”, Experimental Designs, and Applied Regression courses). In the past couple of years we have experimented using both SAS/INSIGHT and JMP. Both products can serve as excellent tools for carrying out analyses usually covered in these courses, and for aiding students in better understanding the models assumptions and results. Both provide routinely graphics in support of the analysis for the purpose of checking models assumptions and displaying the results. One of the main differences between the two is that SAS/INSIGHT is not meant to be used as a stand alone product without SAS. Because it is part of the SAS System, you can use SAS/INSIGHT software to explore results from any SAS procedure. You could also use any SAS procedure to analyze results from SAS/INSIGHT software. In contrast JMP and JMP IN<sup>®</sup> (the student version of JMP) are standalone products. Another difference that is critically

related to the first one is the amount of space (and memory) required to run the two products is vastly different. SAS/INSIGHT is the clear looser in that comparison. This is due to the fact that it SAS/INSIGHT requires big portions of other SAS products before it can be used effectively. As far as capabilities SAS/INSIGHT has the advantage in fitting better all generalized linear models where as JMP better fits the general linear model, and nonlinear models. Also, JMP allows minimal handling for random effects and analytical tools for ordinal responses. We give JMP the edge in reading and writing data sets, and found its printed output to be of higher quality. We did noticed recently improvements in SAS/INSIGHT printing capabilities since version 6.10. We are going first to address advantages and disadvantages of each when examined by themselves. Last we going to attempt and suggest additional analytical features and potential improvements that could enhance in our view both products. Keep in mind that our objective is to teach practitioners (graduate researchers) how to use an interactive easy to learn product to carry out complex analysis. In the same time we are hoping to aid the practitioner in exploring, understanding and displaying the results of these analyses without requiring major knowledge of programming in the SAS system.

## SAS/INSIGHT

SAS/INSIGHT is an excellent tool that provides a comprehensive set of exploratory and analytical tools for the general linear model, and the generalized linear model analyses. Both SAS/INSIGHT and JMP are designed as dynamic data analytic and exploratory tools. Data stay always in front of you in a direct view in a spreadsheet fashion. Analysis and graphics are ready to respond to you interest and everything is alive until closed. Rows (observations) in the spreadsheet are cross linked in all the windows. For example, when you click on a point in a plot, the point highlights and the label of the point appears in the plot. There are many similarities in both products in the available operations of rows and columns (variables), in their attributes and in the availability of other general tools and operations (such as brushing, zooming etc.). Unfortunately SAS/INSIGHT as the name suggest requires major portions of SAS and also portions of SAS/FSP<sup>®</sup>, and SAS/GRAPH before it can be used effectively. We estimated the minimal amount of hard disk space needed to handle the typical analytical tools covered in the above courses to be close to 60 MB for a “PC/Windows” microcomputer environment.

SAS/INSIGHT is slightly more effective than JMP in exploration of distributions (parametric and nonparametric) and the analysis of

response data from the entire class of exponential family distributions. One of its major advantage over JMP is that it allows the record and submit SAS/INSIGHT statements to create graphs and analyses automatically. This feature can save practitioners valuable time when dealing with repetitive analysis or when you work with large data sets. It also includes Quasi-likelihood estimation that is particularly useful in dealing with overdispersion that often occurs with Binomial and Poisson data. We found SAS/INSIGHT most useful in our Regression course.

Some of its weakness include: lack of multivariate routines, inability to fit nonlinear models, lack of handling random effects in context of the general linear mixed model theory, and lack of nonparametric analysis (other than in the form of relationships). When dealing with classification effects in the context of the general linear model analysis the usual printout lacks important information that usually is provided by the LSMEANS, ESTIMATE, and CONTRAST statements in the usual GLM messy data analysis. But by far the most limiting factor for our purposes was that it cannot handle random effects or models with multiple error terms (i.e. Split-Plot and repeated measures analysis). Also it doesn't include any correlated error regressions (autoregression) that is typically covered in a standard Regression course.

Other limitations (omissions) that we did not encounter in the context of the above courses include the lack of survival routines and time series routines. Also most multivariate routines other than principal component analysis are absent (i.e. factor, discriminate, cluster and other multidimensional scaling techniques). Of course all of these limitations can be performed by SAS/STAT, and SAS/ETS<sup>®</sup> procedures that could be used to produce results that SAS/INSIGHT can further explore. The above nice interface between SAS/INSIGHT and other SAS products could also be counted as a big advantage. Printing has improved considerably in version 6.11.

### **JMP**

JMP on the other hand is a stand alone statistical package that takes about a one tenth of the combination of SAS/INSIGHT and other SAS<sup>®</sup> products/procedures needed to perform the analyses for these courses. It has two basic fitting machines refer to as springs, machine for fitting continuous responses and pressure volumes for categorical responses. It can read data other than SAS data sets (only ASCII at this time) and can provide a nice journal facility for storing the important parts of the output from the analysis that can later be incorporated into a word-processing document (as rtf files). JMP offers six statistical analysis platforms that include capabilities for performing univariate statistics, analysis of variance and multiple regression, nonlinear fitting, multivariate analysis, and nonparametric tests. It also features integrated capabilities for quality improvement and design of experiments, offering five type of classical designs for estimating the effects of one or more factors on a dependent variable. JMP

also provides a variety of graphical tools designed for quality control. About ninety percent of its analytical (other than multivariate and survival tools) and graphical tools (other than contour and ternary graphs) can be found in JMP IN.

Important features especially in the teaching of Experimental Designs found in JMP (and JMP IN) but not found in SAS/INSIGHT include is the Design Experiment command that offers a design choice for almost every situation. Since experimentation is usually expensive good software could facilitate good experimental design especially if you are not an expert. Most all of these capabilities for experimental design can also be explored by the FACTEX and OPTEX procedures of SAS/QC<sup>®</sup>. Some of the analytical tools found in JMP but missing from SAS/INSIGHT include the fitting of nonlinear models, survival analysis, cluster analysis and repeated measures analysis. JMP also includes methods for handling ordinal types of data in addition to the continuous and nominal types included in SAS/INSIGHT. JMP introduced the first implementation of the general leverage plot and pioneered comparison circles, the first means-comparison method that works for unequal samples. But is its unified approach to statistical methodology, its integration of statistical analysis and graphics and its ease of operation combined with its comprehensiveness that our students found exceptionally rewarding.

Even though both products have lots of similarities in touch and feel since everything is direct-view we still found major improvements in JMP as compared to SAS/INSIGHT. These include: a more comprehensive tools (nine versus three) for specialty clicks, an ability to rework data tables in seven ways to form new data tables and a built in calculator to manipulate and create new variables. Also we found most helpful the option in JMP for providing plots of the means in the FIT MODEL platform very useful for better understanding tests of effects (main effects and two-way interaction effects). We will like to see that particular feature extended past the two-way interactions for higher order factorial treatment designs. For example a three factor interaction could be represented in terms of several two factor mean plots. The design goal of JMP is: to choose family of methods that can be unified, use analysis depending on modeling type, and organize methods into platforms that handle general situations. JMP anticipates "better" (especially true in the one way layout situations) and routinely provides lack of fit tests in regression situations and tests against base models in categorical analyses situations. It includes capabilities for more custom tests in the general linear model analysis, and provides more details including sample size calculations, power, and least significant number and value for each of the factors in the model. All these additions help answer better the following questions in the context of the general linear model analysis.

- How sensitive is my data design?
- Do I need more data and how much?
- In a future similarly conducted experiment how likely am I to detect a significant effect?

Another major advantage is JMP ability to provide automatic test synthesis for mixed models (similar to the one provided by PROC GLM). Thus unlike SAS/INSIGHT it can handle random effects and provides expected mean squares, variance component estimates and test denominator synthesis, performs MANOVA, including repeated measures and effective hypothesis tests when dealing with missing data cells. JMP also includes the three most popular multiple comparisons methods unfortunately only for the one-way layout situation. Both products include BY processing (SAS/INSIGHT in a more straightforward way) but neither allows for simultaneous printing of all the by processing analysis windows.

Unfortunately its approach in performing mixed model analysis is only "like" PROC GLM (least squares) and not as in PROC MIXED (maximum likelihood approach). Thus JMP provides only an approach that is expedient for dealing with mixed models. Again some other weakness (not necessarily encountered in the typical teaching of the above courses) include: lack of more multivariate routines, time series routines and full blown generalized linear theory that includes quasi likelihood estimation. JMP inability to run in a batch mode environment couple with its inability to store sequences of steps for analysis into programming statements (that could be submitted in a batch mode) makes it hard to use for repetitive analyses of large data.

## CONCLUSION

JMP provided practitioners (graduate students in applied fields) without any significant expertise on data management, preparation, manipulation, analysis and presentation of the results with an easy to learn agile instrument that is amazing capable. We learned based on written course evaluations that both allowed students to learn more by discovering more things about their data, allow them to interact more with their data and understand better the results of the analyses. Needless to say that they found JMP the clear winner between the two. Based on our experience we will not hesitate to highly recommend JMP to others as the primary tool for teaching students enrolled in services courses (such as a first Statistical Methods, Experimental Design and Applied Regression). On the other hand we will only recommend SAS/INSIGHT as an excellent tool for only the first Statistical Methods and maybe the Applied Regression course. Both take

less time to learn than the combinations of SAS with the SAS/STAT procedures required to teach the same material. It provides students (and the instructors) with more time to concentrate on the data at hand, the model and its assumptions and helps them to better communicate the results and their significance.

Still JMP the most complete capable instrument of the two was not as complete tool for our purposes as we will like it to be. Important omissions included: maximum likelihood and other types of estimation popular in mixed model analysis and a larger variety of multivariate analyses. Hopefully in the future JMP IN will include the multivariate capabilities (repeated measures analysis) found in JMP. Also absent from both (JMP and SAS/INSIGHT) in the general linear model framework was the availability of TYPE II and TYPE IV Sums of Squares that could be useful in some messy data situations. Both lack availability of multiple comparisons for general multifactor linear model. JMP provides quite a lot more in this area but only for the one-way layout situation. We will also welcome additions to both products of most of the options and adjustments available for further comparisons performed by the MEANS and LSMEANS statements that exist in PROC GLM. We especially appreciated the WARNINGS found in JMP in cases where Satterwaite approximation was in order in mixed models situations. Last we would like to see both having the ability to save output in an HTML format.

Expert SAS users can always produce analysis and graphs far better than those produced by JMP or SAS/INSIGHT but with a lot more work. Overall, student response to both products far exceeded the marks gotten previously when we used SAS with selected SAS/STAT procedures to carry similar analyses.

## REFERENCES

*JMP® Statistics and Graphics Guide, Version 3.1*, Cary, NC: SAS Institute Inc. 1995.

*SAS/INSIGHT® User's Guide, Version 6, Third Edition*, Cary, NC: SAS Institute Inc. 1995.

JMP, JMPIN, SAS, SAS/ETS, SAS/FSP, SAS/GRAPH, SAS/INSIGHT, SAS/QC, and SAS/STAT are registered trademarks of SAS Institute Inc. in the USA and other countries. ®indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their registered companies.