

# JMP Density Ellipses for Pharmaceutical Analysis

By Charles E. Shipp and Kirk P. Lafler

## Abstract

JMP is a data discovery tool used in both (1) the exploratory phase and (2) the analysis phase, of experimental and theoretical work. The density ellipse shows the extent of data, the center of mass, the least squares fit line, and the degree of correlation between a selected dependent variable and a selected independent variable. When superimposed, the ellipses of similar variables for different classes show and contrast their sameness and differences. The example used is pharmaceutical dummy data in the area of cancer research.

## Background

The history of SAS analysis is well known. There are latest developments and leading edge analysis R/D in the JMP suite of statistical analysis platforms. Part of the JMP direction is to make the use direct and intuitive, so the user can explore the nature and relationships of the data unencumbered.

The techniques of density ellipse go far back also, but are not understood and used to their fullest potential.

Pharmaceutical endeavors, in the preliminary stages and subsequent stages of business, are involved in looking for relationships in the dependent and independent variables. In some studies, determining the causal effects is the goal, with their magnitudes.

## Introduction

The method used here is to use JMP to run a matrix of correlation graphs and use the results for preliminary analysis. Other exploratory methods can also be used at this point. Decisions need to be made as to which are independent variables and which are dependent.

The second phase is to select the proper dependent and independent variables to run for different classes, such as clinical sites or drug-chemo treatment. These density ellipses are then superimposed to show consistency or differences in regions, treatment, etc.

## The Importance of Statistical Graphics

Statistical graphics finds its strongest advantages for the following work:

- Seeing subtle differences, trends, and other important measures for R/D
- Communicating in a visible and quick manner, to FDA and management
- Discovering and realizing new areas for detailed analysis.

Tables are nice and important, but graphics communicate differently. What if stoplights had the words 'stop' and 'go' instead of red, yellow and green in different positions? It would not work as well. Managers/reviewers often do not get all information from tables—they absorb more from good graphics.

## Strategy for Improving Cancer Treatment

There is a lot to do in understanding and improving the methods of cancer treatment. In this fast-moving and competitive field, there is a trade-off between being recognized for good treatment results and proprietary R/D methodologies in discovering what works. Every tool available, especially the best SAS tools, further the research and benefit patients.

## Composite Variables in Pharmaceutical Work

Composite variables are new variables that are made from combining observable variables. The three categories are before, during, and after cancer treatment and thus focus on severity, treatment, and benefits categories. Of the variables in each category, there is a science of prediction that can use ways of combining the variables within the three categories. For example, an overall Benefit Measure, Y1, could be the percent recovery times the quality of life measure.

$$\{X\} \rightarrow \{T\} \rightarrow \{Y\}$$

## SEVERITY MEASURES, {X}

These measurement variables consist of cancer patient categories that show the degree of health degradation and status of the patient before treatment. Measures of the patient include age, sex, race, heredity, fitness, diet, smoking, stress, and body mass index. In addition, there are the status measures of number of tumors, size of tumors, stage (I, II, III, IV), morbidity, and complications. Composite variables for severity can be defined as: X1, X2, X3, ... . and are evaluated for each patient by different prescribed formulas.

## TREATMENT MEASURES, {T}

Treatment variables (measures and classes) include the regimen (drug/chemo) given and the percent completion of treatment. Other classes could be different clinics, cities, or methodologies. Composite variables for treatment can be defined as: T1, T2, T3, ... .

## **BENEFIT MEASURES, {Y}**

The measure of treatment benefit to cancer patients (in this data) is QL (quality of life), R% (percent recovery), survivability, longevity, and composite variables calculated by combining them. Composite variables for benefit can be defined as: Y1, Y2, Y3, ...

## **Future Work to Define Composite Variables**

Choosing the proper variables for R/D and presentations is a good challenge. Composite variables combine the measures in three categories so that a single variable can be isolated and analyzed. The three categories are before, during, and after cancer treatment, in this example.

## **The Data**

In this paper real data collections are not used for several obvious reasons, and also for the reason that the examples can be generalized for your own purposes. Independent variables are chosen to be {X1, X2, and X3} and the dependent variables are chosen to be {Y1, Y2}. Think of the three class variables {T1, T2, T3} as three competing chemotherapy treatments, and the two dependent variables as measures of success.

## **Phase I**

You can enter your columns of data directly into JMP, or have them in a flat file that is imported. Once the data resides in a JMP table, click on the following command options:

- JMP Starter
- Multivariate (twice)
- Identify Y as all numeric variables
- Click OK (shows the Matrix of Correlation, with visual density ellipses.)

The results can be stored as a Journal Entry, and cut and pasted to your formal reports.

The results of doing these steps in JMP are shown in Graph 1 at the end of the paper. Picture of the results show an n by n correlation matrix graph of variables, showing density ellipses.

The analyst often knows which dependent variable is the object of most interest, to maximize. Looking at the results, we see which variables have strong correlation.

## **Phase II**

Choosing the most interesting variables for further analysis, we could run the data by clinical classes consisting of treatment sites, {S1, S2, S3, S4, S5, S6, S7}. We would look more closely by choosing the independent variable and dependent variable match-up of most interest, and look more closely at how they correlate at the different sites.

We now run JMP and include the regimen of drugs as the CLASS variable. Click on the following command options:

- Basic Statistics
- Bivariate
- Identify X as stress (or an independent variable of your choice)
- Identify Y as R% (or a dependent variable of your choice)
- Click OK
- Group By: Regimen (or a CLASS variable of your choosing)
- 50% Density Ellipses
- Click on OK

The results of doing these steps in JMP are shown in Graph 2 at the end of the paper. Picture of the results shows example overlay showing three classes of regimen overlaid. The analyst can see the similarity and differences with this excellent visual graphic.

## **Additional Class Example**

Combining all patient data for the most interesting variables for further analysis, we can run the data by classes consisting of resident cities, {C1, C2, C3, C4, C5, C6, C7}. We would be looking more closely by choosing the independent variable and dependent variable match-up of most interest, and looking more closely at how they correlate at all different cities.

You follow the same steps as given above, but include the city as the CLASS. This shows the correlation between Y1 vs. X3, for example, at each of the cities. The analyst can see the similarity and differences with this excellent visual graphic.

The final statistical graphics that you use can consist of individual variables or combined variables for the three categories described (dependent, class, and independent.)

## **Summary**

For studies that do not have definitive independent and dependent variables, their relationships can be explored using correlation graphs within JMP. Once the variables of most interest are selected, a closer look among them versus the treatment is possible.

## **Conclusions**

JMP is not only intuitive and powerful, but it also contains, built in, the suite of statistical approaches that assist in exploration and analysis. This paper illustrates just one approach in using JMP for research and communication. The message is given via statistical graphs so that the message can be observed quickly, and subtle details can be seen. JMP has many other statistical graphics capabilities in addition to the ones shown in this paper.

## Future Work

There continues to be much to do in the competitive area of Cancer Research. The use of statistical graphics is a good area to be explored. The company and research team that can make their selling points graphically will go farther than those that rely on tables of obscure numbers alone. A picture augments the value of words and tables, and can be the big difference in discovery, communication, and acceptance.

## References

Sall, John, et. al., JMP Start Statistics, 2<sup>nd</sup> Ed., Duxbury, Pacific Grove, CA, 2001.  
JMP Discovery Software Version 4 set of manuals.

## Acknowledgments

We would like to thank Carolyn Bahler (SESUG Posters Section Chair) for assistance with this paper, and also Heidi Markovitz, Program Chair, and the SESUG Leadership for their support of a great Conference.

## Trademark Citations

SAS, SAS Alliance Partner, and SAS Certified Professional are registered trademarks of SAS Institute Inc. in the USA and other countries. The ® symbol indicates USA registration.

## About the Authors

Charles E. Shipp is a programmer and consultant with over 25 years experience working with the SAS System. He has authored numerous articles and co-authored the popular Books by Users (BBU) book, Quick Results with SAS/GRAPH Software. His expertise includes application design and development, training, and programming using base-SAS, JMP, SQL, ODS, SAS/FSP, SAS/AF, SCL, FRAME, and SAS/GRAPH software. He has also been involved with web design, content creation, and development.

Kirk P. Lafler is a SAS Alliance Partner® and SAS Certified Professional® with over 25 years of SAS software experience. He has written over one hundred articles for professional journals and SAS User Group proceedings. His popular SAS Tips column appears regularly in the SANDS and SESUG Newsletters. His expertise includes application design and development, training, and programming using base-SAS, SQL, ODS, SAS/FSP, SAS/AF, SCL, FRAME, and SAS/EIS software.

< <http://www.Software-Intelligence.com> >



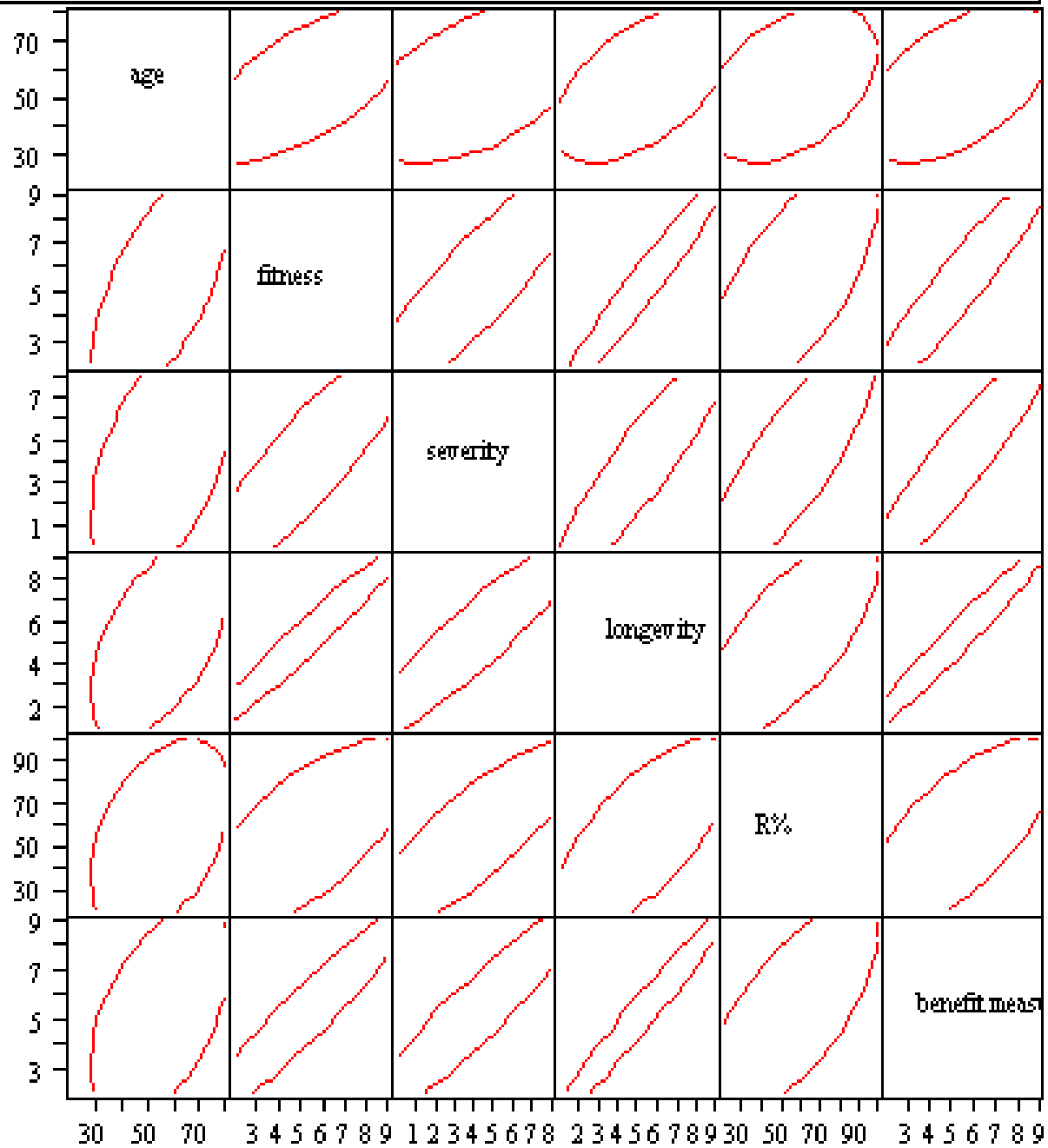
Comments and suggestions can be sent to:

Charles E. Shipp  
CONSULTANT  
1650 Barrywood Avenue  
San Pedro, California 90731-1254  
E-mail: CharlieShipp@AOL.com

# JMP Graph Example 1

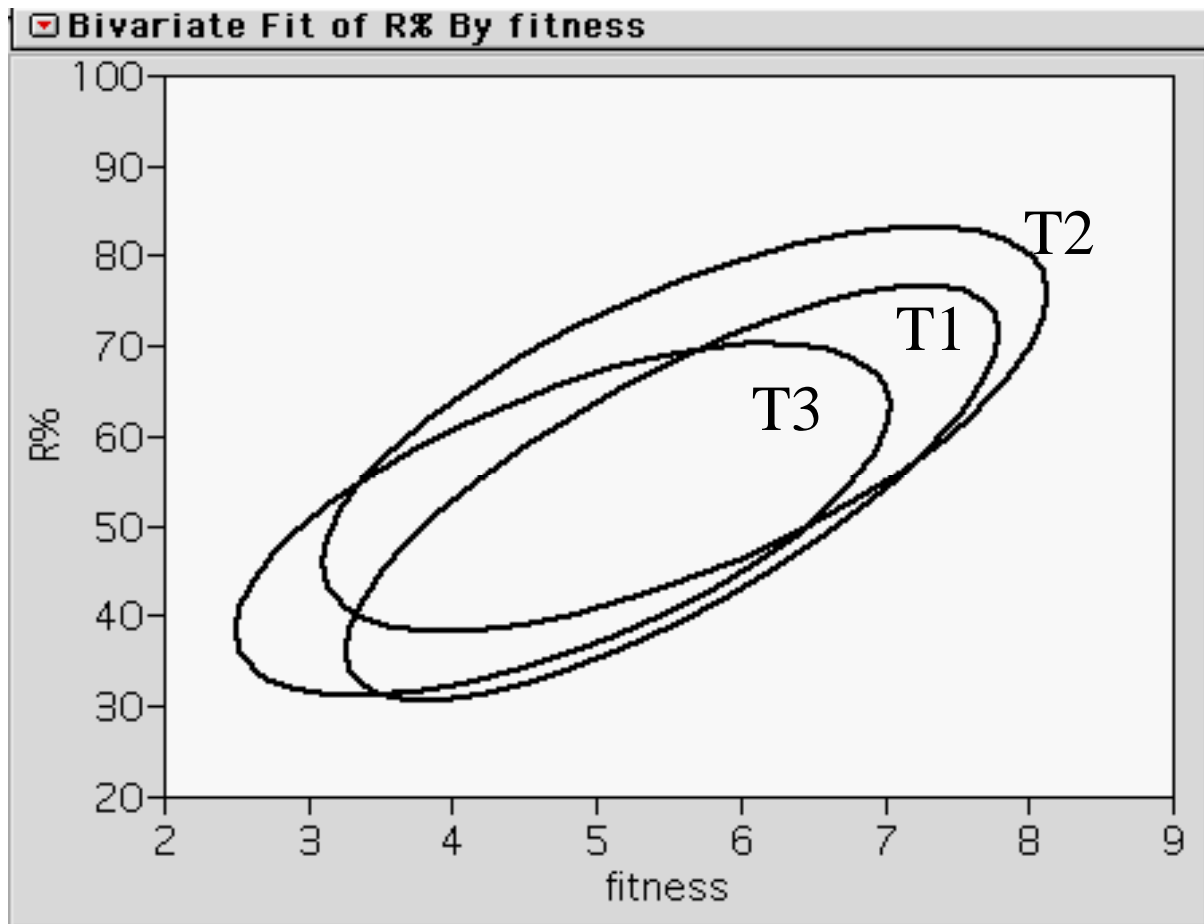
Multivariate

Scatterplot Matrix



This N x N matrix of density ellipse graphs shows how each of the continuous variables correlates to all of the other continuous variables. The researcher then selects from these the most important match between one dependent and one independent variable.

# JMP Graph Example 2: Case Comparison



Correlation drug regimen=="T1"					
Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
fitness	5.51	1.924953	0.769403	0.0093	10
R%	54	19.5505			

Correlation drug regimen=="T2"					
Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
fitness	5.6	2.135936	0.66659	0.0353	10
R%	61	19.11951			

Correlation drug regimen=="T3"					
Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
fitness	4.75	1.928298	0.635684	0.0482	10
R%	51	16.6333			

Notice that in this study that all three regimen classes have positive correlation between fitness and percent recovery, however regimen treatment #2 (T2) has the greatest expected positive result, while regimen treatment #1 (T1) has the greatest correlation, (narrowest ellipse.)