

JMP Notes

(A Quick Reference Supplement to the JMP User's Guide)

These notes outline the JMP commands for various statistical methods we will discuss in class (and then some). Don't worry if you don't understand technical terms like "variance inflation factors." That's what class is for – this document is intended to give you quick pointers to various JMP commands and features, not explain statistical concepts or give detailed descriptions of JMP output.

1. Get familiar with JMP.

You should familiarize yourself with the basic features of JMP by reading the first three chapters of the *JMP Introductory Guide* (JMP Help menu > Books) and doing the Beginner's Tutorial in JMP (JMP Help menu > Tutorials). You don't have to get every detail, especially on all the Formula Editor functions, but you should get the main idea about how stuff works. Some of the things you should be able to do are:

- Open a JMP data table.
- Understand the basic structure of data tables. Rows are observations, or sample units. Columns are variables, or pieces of information about each observation.
- Identify the modeling type (continuous, nominal, ordinal) of each variable. Be able to change the modeling type if needed.
- Make a new variable in the data table using JMP's formula editor. For example, take the log of a variable.
- Use JMP's tools to copy a table or graph and paste it into Word (or your favorite word processor).
- Use JMP's online help system to answer questions for you.

2. Generally Neat Tricks

(a) Dynamic Graphics

JMP graphics are *dynamic*, which means you can select an item or set of items in one graph and they will be selected in all other graphs and in the data table.

(b) Including and Excluding Points

Sometimes you will want to determine the impact of a small number of points (maybe just a single point) on your analysis. You can exclude a point from an analysis by selecting the point in any graph or in the data table and choosing Rows \Rightarrow Exclude/Unexclude. Note that excluding a point will remove it from all future numerical calculations, but the point may still appear in graphs. To eliminate the selected point from graphs, select Rows \Rightarrow Hide/Unhide. You can readmit excluded and/or hidden points by selecting them and choosing Rows \Rightarrow Exclude/Unexclude a second time. An easy way to select all excluded points is by double clicking the "Excluded" line in the lower left portion of the data table. You can also choose Rows \Rightarrow Row Selection \Rightarrow Select Excluded from the menu.

(c) Taking a Subset of the Data

The easiest way to take a subset of the data is by selecting the observations you want to include in the subset and choosing Tables \Rightarrow Subset from the menu. For example, suppose you are working with a data set of CEO salaries and you only want to investigate CEO's in the finance industry. Make a histogram of the industry variable (a categorical variable in the data set), and click on the "Finance" histogram bar. Then choose Tables \Rightarrow Subset from the menu and a new data table will be created with just the finance CEO's.

(d) Marking Points for Further Investigation

Sometimes you notice an unusual point in one graph and you want to see if the same point is unusual in other graphs as well. An easy way to do this is to select the point in the graph, and then right click on it. Choose "Markers" and select the plotting character you want for the point. The point will appear with the same plotting character in all other graphs.

(e) Changing Preferences

There are many ways you can customize JMP. You can select different toolbars, set different defaults for the analysis platforms (distribution of Y, fit Y by X, etc.), choose a default location to search for data files, and several other things. Choose File \Rightarrow Preferences from the menu to explore all the things you can change.

JMP Notes

(A Quick Reference Supplement to the JMP User's Guide)

(f) Shift Clicking and Control Clicking

Sometimes you may want to select several variables from a list or select several points on a graph. You can accomplish this by holding down the shift or control keys as you make your selections. Note that shift and control clicking in JMP works just like it does in other Windows applications.

3. The Distribution of Y

All the following instructions assume you have launched the “Distribution of Y” analysis platform.

(a) Continuous Data

By default you will see a histogram, boxplot, a list of quantiles, and a list of moments (mean, standard deviation, etc.) including a 95% confidence interval for the mean.

i) One Sample t-Test

Click the little red triangle on the gray bar over the variable whose mean you want to do the t-test for. Select “Test Mean.” A dialog box pops up. Enter the mean you wish to use in the null hypothesis in the first field. Leave other fields blank. Click OK.

ii) Normal Quantile Plot (or Q-Q Plot)

Click the little red triangle on the gray bar over the variable you want the Q-Q plot for. Select “Normal Quantile Plot.”

(b) Categorical Data

Sometimes categorical data are presented in terms of counts, instead of a big data set containing categorical information for each observation. To enter this type of data into JMP you will need to make two columns. The first is a categorical variable that lists the levels appearing in the data set. For example, the variable *Race* may contain the levels “Black, White, Asian, Hispanic, etc.” The second is a numerical list revealing how many times each level appeared in the data set. Suppose you name this variable *counts*. When you launch the Distribution of Y analysis platform, select *Race* as the Y variable and enter *counts* in the “Frequency” field.

4. Fit Y by X

All the following instructions assume you have launched the “Fit Y by X” analysis platform. The appropriate analysis will be determined for you by the modeling type (continuous, ordinal, or nominal) of the variables you select as Y and X.

(a) The Two Sample t-Test (or One Way ANOVA). [[Continuous Y & categorical X]]

In the Fit Y by X dialog box, select the variable whose means you wish to test as “Y.” Select the categorical variable identifying group membership as “X.” For example, to test whether a significant salary difference exists between men and women, select “Salary” as the Y variable and “Sex” as the X variable. A dotplot of the data will appear.

i) How to do a T-Test

Click on the little red triangle on the gray bar over the dotplot. Select “Means/Anova / T-test.”

ii) Manipulating the Data

Sometimes the data in the data table must be manipulated into the form that JMP expects in order to do the two sample t-test. The “Tables” menu contains two options that are sometimes useful. “Stack” will take two or more columns of data and stack them into a single column. “Split” will take a single column of data and split it into several columns.

iii) Display Options

The “Display Options” sub-menu under the little red triangle controls the features of the dotplot. Use this menu to add side-by-side boxplots, means diamonds, to connect the means of each subgroup, or to limit over-plotting by adding random jitter to each point's X value.

(b) Contingency Tables/Mosaic Plots. [[Categorical Y & categorical X]]

JMP Notes

(A Quick Reference Supplement to the JMP User's Guide)

In the Fit Y by X dialog box select one categorical variable as Y and another as X. As far as the contingency table is concerned, it doesn't matter which is which. The mosaic plot will put the X variable along the horizontal axis and the Y variable on the vertical axis (as you would expect).

i) Entering Tables Directly Into JMP

Sometimes categorical data are presented in terms of counts, instead of a big data set containing categorical information for each observation. To enter this type of data into JMP you will need to make three columns. The first is a categorical variable that lists the levels appearing in the X variable. For example, the variable *Race* may contain the levels "Black, White, Asian, Hispanic, etc." The second is a list of the levels in the Y variable. For example, the variable *Sex* may contain the levels "Male, Female." Each level of the X variable must be paired with each level of the Y variable. For example, the first row in the data table might be "Black, Female." The second row might be "Black, Male." The final column is a numerical list revealing how many times each combination of levels appeared in the data set. Suppose you name this variable *counts*. When you launch the Fit Y by X analysis platform, select *Race* as the X variable, *Sex* as the Y variable, and enter *counts* in the "Frequency" field.

ii) Display Options for a Contingency Table

By default, contingency tables show counts, total %, row %, and column %. You can add or remove listings from cells of the contingency table by using the little red triangle in the gray bar above the table.

(c) Simple Regression [[One Y and one X, both continuous]]

In the Fit Y by X dialog box select the continuous variable you want to explain as Y and the continuous variable you want to use to do the explaining as X. A scatterplot will appear. The commands listed below all begin by choosing an option from the little red triangle on the gray bar above the scatterplot.

i) Fitting Regression Lines

Select "Fit Line" from the little red triangle.

ii) Fitting Non-Linear Regressions

(1) Transformations

Choose "Fit Special" from the little red triangle. A dialog box appears. Select the transformation you want to use for Y, and for X. Click Okay. If the transformation you want to use does not appear on the menu you will have to do it "by hand" using JMP's formula editor. Simply create a new column in the data set and fill the new column with the transformed data. Then use this new column as the appropriate X or Y in a linear regression.

(2) Polynomials

Choose "Fit Polynomial" from the little red triangle. You should only use degree greater than two if you have a strong theoretical reason to do so.

(3) Splines

Choose "Fit Spline" from the little red triangle. You will have to specify how wiggly a spline you want to see. Trial and error is the best way to do this.

iii) The Regression Manipulation Menu

You may want to ask JMP for more details about your regression after it has been fit. Each regression you fit causes an additional little red triangle to appear below the scatterplot. Use this little red triangle to:

- Save residuals and predicted values
- Plot confidence and prediction curves
- Plot residuals

(d) Logistic Regression [[Categorical Y & one or more Xs]]

In the Fit Y by X dialog box choose a nominal variable as Y and a continuous variable as X. There are no special options for you to manipulate. You have more options when you fit a logistic regression using the Fit Model platform.

JMP Notes

(A Quick Reference Supplement to the JMP User's Guide)

5. Matched Pairs

Use the “Matched Pairs” analysis platform to do the paired t-test. The data should be organized into two columns, so that each row gives two values for each observation in the data set. In the matched pairs dialog box select both these columns as Y.

(a) Manipulating Data

Sometimes the data in the data table must be manipulated into the form that JMP expects in order to do the paired t-test. The “Tables” menu contains two options that are sometimes useful. “Stack” will take two or more columns of data and stack them into a single column. “Split” will take a single column of data and split it into several columns.

(b) What you See

The graph you see when you do matched pairs plots the average value for each observation on the X axis and the difference between Y values for each observation on the Y axis. No information from the X axis is used by the paired t-test, only the Y axis (the differences). The average difference for the sample is plotted as a solid horizontal line. Another solid horizontal line is plotted at zero, for reference. Dotted lines represent the 95% confidence interval for the mean difference in the data set. If the observations are spread out enough you will see a diamond plotted. If you tilt your head so that the diamond looks like a square then you will see a scatterplot of one Y variable against the other.

6. Multivariate [[Two or more continuous variables]]

Launch the multivariate platform and select the continuous variables you want to examine.

(a) Correlation Matrix

You should see this by default.

(b) Covariance Matrix

You have to ask for this using the little red triangle.

(c) Scatterplot Matrix

You may or may not see this by default. You can add or remove the scatterplot matrix using the little red triangle. You read each plot in the scatterplot matrix as you would any other scatterplot. To determine the axes of the scatterplot matrix you must examine the diagonal of the matrix. The column the plot is in determines the X axis, while the plot's row in the matrix determines the Y axis. The ellipses in each plot would contain about 95% of the data if both X and Y were normally distributed. Skinny, tilted ellipses are a graphical depiction of a strong correlation. Ellipses that are almost circles are a graphical depiction of weak correlation.

7. Fit Model [[Continuous Y and one or more Xs]]

The Fit Model platform is what you use to run sophisticated regression models with several X variables.

(a) Running a Regression

Choose the Y variable and the X variables you want to consider in the Fit Model dialog box. Then click “Run Model.”

(b) Once the Regression is Run

i) Variance Inflation Factors

Right click on the Parameter Estimates box and choose Columns \Rightarrow VIF in the menu that appears.

ii) Confidence Intervals for Individual Coefficients

Right click on the Parameter Estimates box and choose Columns \Rightarrow Lower 95% in the menu that appears. Repeat to get the Upper 95%, completing the interval.

iii) Save Columns

This menu lives under the big gray bar governing the whole regression. Options under this menu will save a new column to your data table. Use this menu to obtain:

JMP Notes

(A Quick Reference Supplement to the JMP User's Guide)

- Cook's Distance
- Leverage (or "Hats")
- Saving Residuals
- Saving Predicted (or "Fitted") Values
- Confidence and Prediction Intervals

This menu is especially useful for making new predictions from your regression. Before you fit your model, add a row of data including the X variables you want to use in your prediction. Leave the Y variable blank. When you save predicted values and intervals JMP should save them to the row of "fake data" as well.

iv) Row Diagnostics

This menu lives under the big gray bar governing the whole regression. Options under this menu will add new tables or graphs to your regression output. Use this menu to obtain:

- Residual Plot (Residuals by Predicted Values)
- The Durbin Watson Statistic
- Plotting Residuals by Row

v) Expanded Estimates

The expanded estimates box reveals the same information as the parameter estimates box, but categorical variables are expanded to include the default categories. To obtain the expanded estimates box select Estimates \Rightarrow Expanded Estimates from the little red triangle on the big gray bar governing the whole regression.

(c) Including Interactions and Quadratic Terms

To include a variable as a quadratic, go to the Fit Model dialog box. Select the variable you want to include as a quadratic. Then click on the "Macros" menu and select "Polynomial to Degree." The "Degree" field under the Macros button controls the degree of the polynomial. The degree field shows "2" by default, so the Polynomial to Degree macro will create a quadratic.

There are two basic ways to include an interaction term. The first is to select the two variables you want to include as an interaction (perhaps by holding down the Shift or Control key) and hitting the "Cross" button. The second is to select two or more variables that you want to use in an interaction (using the Shift or Control key) and select Macros \Rightarrow Factorial to Degree.

(d) To Run a Stepwise Regression

In the Fit Model Dialog box, change the "Personality" to "Stepwise." Enter all the X variables you wish to consider (including any interactions—see the instructions on interactions given above). Then click "Run Model."

The Stepwise Regression dialog box appears. Change "Direction" to "Mixed" and make the probability to enter and the probability to leave small numbers. The probability to enter should be less than or equal to the probability to leave. Then click "Go." When JMP settles on a model, click "Make Model" to get to the familiar regression dialog box. You may want to change "Personality" to "Effect Leverage" (if necessary) so that you will get the leverage plots.

One of the odd things about JMP's stepwise regression procedure is that it creates an unusual coding scheme for dummy variables. Suppose you have a categorical variable called *color*, with levels Red, Blue, Green, and Yellow. JMP's stepwise procedure may create a variable named something like *color [Red&Yellow-Blue&Green]*. This variable assumes the value 1 if the color is red or yellow. It assumes the value -1 if the color is blue or green. This type of dummy variable compares colors that are either red or yellow to colors that are either blue or green.

(e) Logistic Regression [[Categorical (binary) Y and one or more Xs]]

Logistic regression with several X variables works just like regression with several X variables. Just choose a binary (categorical) variable as Y in the Fit Model dialog box. Check the little red triangle on the big gray bar to see the options you have for logistic regression. You can save the following for each item in your data set: the probability an observation with the observed X would fall in each level of Y, the value of the linear predictor, and the most likely level of Y for an observation with those X values.