JMP Guide

(Compiled from STAT 301 students-Spring 2017 and updated by Dr. Ziegler)

<u>Analysis</u>	Page
Descriptive Statistics and Graphs for Quantitative Variable (One Mean)	2
Descriptive Statistics and Graphs for Categorical Variable (One Proportion)	2
Descriptive Statistics and Graphs for Comparing Multiple Means using Distribution	2-3
Descriptive Statistics and Graphs for Comparing Multiple Means using Fit Y by X	3
Confidence Interval for One Mean	3
Hypothesis Test for One Mean	3
Confidence Interval for Difference in Two Independent Means	4
Hypothesis Test for Difference in Two Independent Means	4
Confidence Interval for Difference in Means with Paired Data	4
Hypothesis Test for Difference in Means with Paired Data	5
Scatterplot using Graph Builder	5
Scatterplot Matrix and Correlation Matrix	5
Simple Linear Regression using Fit Y by X	6
Linear Regression using <i>Fit Model</i>	6
<u>VIF</u>	6
Creating Indicator Variables	6
Multiple Regression including Interaction using Fit Model	7
Multiple Regression with Quadratic Term using Fit Model	7
Normal Quantile Plot and Histogram for Residuals using Fit Model	7
Transformation: Create New Variable that is the Natural Log of a Different Variable.	8
Leverage/Influence Plot	8
Prediction Intervals and Confidence Intervals for a Mean at Specific Values of X* usin	g Fit
<u>Model</u>	9
Model Selection-RMSE, R ² , Adjusted R ² , and Mallow's C _p	9
Model Selection-Backward Elimination	9
Partial F-test: Hypothesis Test for Categorical Variable with More Than 2 Levels in	
Regression using Fit Model	10
One-way ANOVA using Fit Y by X	10
One-way ANOVA using Fit Model	11
Multiple Comparisons with One-way ANOVA using Fit Y by X	11
Scatterplot with Manually Added Simple Linear Regression Lines using Graph Builder	<u>-</u> 11
Contingency Tables and Chi-Square Test for Independence using Fit Y by X	12
Logistic Regression with Quantitative Explanatory Variable(s)	12

Descriptive Statistics and Graphs for Quantitative Variable (One Mean):

- Go to *Analyze*, *Distribution*.
- Put your quantitative variable in the *Y*, *Columns* box.
- Click *OK*.
- Click the red triangle next to your variable name and select *Display Options, Horizontal Layout*.
- Click the red triangle next to your variable name and select *Histogram Options*, *Prob Axis*.
- Right click on the horizontal axis labels and select *Add axis label*... Type in the name of the variable in the box and click *OK*.
- Click the red triangle next to *Summary Statistics* and select *Customize Summary Statistics*.
- Add extra summary statistics values as needed.

Descriptive Statistics and Graphs for Categorical Variable (One Proportion):

- Go to *Analyze*, *Distribution*.
- Put your variable of interest in the *Y*, *Columns* box.
- Click OK.
- Click the red triangle next to your variable name and select *Histogram Options* and *Separate Bars*.
- Click the red triangle next to your variable name and select *Display Options, Horizontal Layout*.
- Click the red triangle next to your variable name and select *Histogram Options*, *Prob*
- Right click on the horizontal axis labels and select *Add axis label*... Type in the name of the variable in the box and click *OK*.
- Click the red triangle next to your variable name and select *Histogram Options, Separate Bars*.

<u>Descriptive Statistics and Graphs for Comparing Multiple Means using Distribution:</u>

- Go to *Analyze*, *Distribution*.
- Put your quantitative variable in the Y, Columns box.
- Put your categorical variable in the By box.
- Click *OK*.
- Click the red triangle next to your variable name for each group and select *Display Options*, *Horizontal Layout*.
- Click the red triangle next to your variable name for each group and select *Histogram Options* and *Prob Axis*.
- Change the limits and scale of each histogram so they are on the same scale. Double-click on a number on the x-axis of each histogram, one at a time. For each, change the *Minimum* to a value equal to or below the overall minimum and change the *Maximum* to a value equal to or above the overall maximum. Use the same values for all histograms. Edit the *Increment* and # *Minor Ticks* to be the same for all histograms.

- Click the red triangle next to *Summary Statistics* for each group and select *Customize Summary Statistics*.
- Add extra summary statistic values as needed.

Descriptive Statistics and Graphs for Comparing Multiple Means using Fit Y by X:

- Go to Analyze, Fit Y by X.
- Put the *quantitative* (response) variable in the Y, response box and categorical (explanatory) variable in the X, factor box.
- Click *OK*.
- To see the means, standard deviations, and sample sizes for each group, click on the red arrow next to *One-way Analysis* and select *Means and Std Dev*.
- To see the histograms for each group, click on the red arrow next to *One-way Analysis* and select *Display Options, Histograms*.
- To see the boxplots for each group, click on the red arrow next to *One-way Analysis* and select *Display Options, Boxplots*.

Confidence Interval for One Mean:

- Go to *Analyze*, *Distribution*.
- Select your variable of interest and place in in the *Y*, *Columns* box.
- Click OK.
- Click the red triangle next to the name of the variable you are analyzing.
- Hover over "Confidence Interval" and select the probability you wish to use for your confidence level.
- Your computed confidence interval will appear at the bottom of your data window under "Mean: Lower Cl and Upper Cl."

Hypothesis Test for One Mean:

- Go to *Analyze*, *Distribution*.
- Select your variable of interest and place in in the Y, Columns box.
- Click *OK*.
- Click the red triangle next to the name of the variable you are analyzing.
- Select *Test Mean*.
- Enter the hypothesized mean into the first box. Leave the second box empty.
 - The second box would give a *z*-distribution instead of a *t*-distribution.
- Click *OK*.
- Under the heading *Test Mean*, a new set of output will appear.
- Identify the test statistic in the output.
- Below the test statistic will be the *p*-values. Match the sign from the H_a to select the *p*-value of interest.
 - \circ Prob > |t| is the *p*-value for a two-tailed test.
 - \circ Prob > t is the *p*-value for a right-tailed test.
 - \circ Prob < t is the *p*-value for a left-tailed test.

Confidence Interval for Difference in Two Independent Means:

- Go to Analyze, Fit Y by X.
- Put your quantitative variable in the *Y*, *Response* box and your categorical group variable in the *X*. *Factor* box.
- Click *OK*.
- For un-equal variances: Click the red triangle next to *One-way Analysis* and select *t test*.
- For equal variances: Click the red triangle next to *One-way Analysis* and select *means/ANOVA/Pooled t*.
- The *Upper CL Dif* and *Lower CL Dif* is the 95% confidence interval.
 - To change the confidence interval, for example 99%, click the red down arrow next to *Oneway Analysis*.
 - Select Set α Level and choose the significance level.
- To see the histograms for each group, click on the red arrow next to *One-way Analysis* and select *Display Options, Histograms*

Hypothesis Test for Difference in Two Independent Means:

- Go to *Analyze*, *Fit Y by X*.
- Put your quantitative variable in the *Y*, *Response* box and your categorical group variable in the *X*, *Factor* box.
- Click OK.
- For un-equal variances: Click the red triangle next to *One-way Analysis* and select *t test*.
- For equal variances: Click the red triangle next to *One-way Analysis* and select *means/Anova/Pooled t*.
- The test statistic and p-values show up on the bottom right.
- *P*-value:
 - \circ Prob > |t| is the *p*-value for a two-tailed test.
 - Prob > t is the p-value for a right-tailed test.
 - Prob < t is the *p*-value for a left-tailed test.

Confidence Interval for Difference in Means with Paired Data:

- Create a column of differences.
 - Make sure your data is in two columns.
 - Go to Cols, New Columns....
 - Give your new column a name such as Differences.
 - Select Column Properties, Formula.
 - Click on the name of one of your columns.
 - \circ Click on the sign at the top in the middle.
 - Click on the name of your other column.
 - Click *OK* and *OK*.
- Go to *Analyze*, *Distribution*.
- Select your difference variable and place in in the *Y*, *Columns* box.
- Click *OK*.
- Click the red triangle next to the name of the difference variable.
- Hover over *Confidence Interval* and select the probability you wish to use for your confidence level.
- Your computed confidence interval will appear at the bottom of your data window under "Mean: Lower Cl and Upper Cl."

Hypothesis Test for Difference in Means with Paired Data:

- Create a column of differences.
 - Make sure your data is in two columns.
 - Go to Cols, New Columns....
 - Give your new column a name such as *Differences*.
 - o Select Column Properties, Formula.
 - Click on the name of one of your columns.
 - \circ Click on the sign at the top in the middle.
 - Click on the name of your other column.
 - \circ Click OK and OK.
- Go to *Analyze*, *Distribution*.
- Select your difference variable and place in in the Y, Columns box.
- Click OK.
- Click the red triangle next to the name of the difference variable and select *Test Mean*.
- Enter the hypothesized mean 0 into the first box. Leave the second box empty.
- Click *OK*.
- Under the heading *Test Mean*, a new set of output will appear.
- Identify the test statistic in the output.
- Below the test statistic will be the *p*-values. Match the sign from the H_a to select the *p*-value of interest.
 - \circ Prob > |t| is the *p*-value for a two-tailed test.
 - \circ Prob > t is the *p*-value for a right-tailed test.
 - \circ Prob < t is the *p*-value for a left-tailed test.

Scatterplot using *Graph Builder*:

- Click on *Graph*, *Graph Builder*.
- Drag your quantitative explanatory variable to the *x*-axis and the quantitative response variable to the *y*-axis.
- Click the small icon at the top that shows a small scatter plot with a line on it.
- Right click on the scatterplot. Go to *Line of Fit* and select *Equation*.
- If you would like to split your points by categories, drag your categorical grouping variable to the *Overlay* box on the top right.
- If we would like to make a Graphical Representation of the Prediction Intervals, on the bottom left next to *Confidence*, click on the *Prediction* box. The graph should now show a lighter shaded blue band. This gives the prediction intervals.
- Click *Done* on the top left.

Scatterplot Matrix and Correlation Matrix:

- Go to *Analyze, Multivariate methods, Multivariate.*
- Place all the variables in the *Y*, *Column* box.
- Click OK.

Simple Linear Regression using Fit Y by X:

- Go to *Analyze*, *Fit Y by X*.
- Put the response variable in the *Y*, *Response* box and the explanatory variable in the *X*, *Factor* box.
- Click OK.
- Click on the red triangle next to *Bivariate Fit* and select *Fit Line*.
- To add the residual plots, click on the red triangle next to *Linear Fit* in the middle of the output, and select *Plot Residuals*. You will want to look at the first plot (*Residual by Predicted Plot*) and last plot (*Residual Normal Quantile Plot*).
- To get confidence intervals for the estimates, right click on any number in the *Parameter Estimates* section, go to *Columns*, and check *Lower 95%* and *Upper 95%*.

Linear Regression using Fit Model:

- Go to Analyze, Fit Model.
- Put the response variable in the *Y* box.
- Select the explanatory variable(s) and click the *Add* button to put the variables in the *Construct Model Effects* box.
- Click Run.
- To show the regression equation, click on the red triangle next to *Response* and select *Estimates, Show Prediction Expression*.
- To get confidence intervals for the estimates, click on the red triangle next to *Response* and select *Regression Reports, Show all Confidence Intervals*.
- To compute predictions, click on the red triangle next to *Response* and select *Estimates*, *Multiple Comparisons*. Enter the values of x you would like to make predictions for.

VIF:

- In the *Fit Model* output:
- Hold the cursor over the *t Ratio* in the *Parameter Estimates* part of the output.
- Next right click and select Columns, VIF.

Creating Indicator Variables:

- Option 1: Recoding
 - o Highlight the Column of the categorical variable.
 - Click on *Cols*, *Recode*.
 - Replace the categories with numbers in the *New Values* column.
 - O Click *Recode*. JMP will label this new column as 'categorical variable name 2'.
 - Right click on your new column, 'categorical variable name 2' and click *Column Info*...
 - Change *Data Type* to *Numeric* and *Modeling Type* to *Continuous*.
 - Click OK.
- Option 2: Indicator Columns (DO NOT USE IF YOU HAVE MISSING DATA)
 - Highlight the Column of the categorical variable.
 - Click on Cols, Utilities, Make Indicator Columns.
 - Click OK.

Multiple Regression including Interaction using *Fit Model***:**

- If you have any categorical variables to include in the interaction term(s), create indicator variables in JMP.
- Go to *Analyze*, *Fit Model*.
- Put the response variable in the *Y* box.
- Select the explanatory variables and click the *Add* button to put the variables in the *Construct Model Effects* box.
- Highlight the two explanatory variables you would like to include in the interaction in the *Select Columns* box on the left.
- Click *Cross* to add the interaction into the *Construct Model Effects* box.
- Click on the red triangle next to *Model Specification* and uncheck *Center Polynomials*.
- Click Run.
- To show the regression equation, click on the red triangle next to *Response* and select *Estimates, Show Prediction Expression*.

Multiple Regression with Quadratic Term using Fit Model:

- If you have any categorical variables to include in the interaction term(s), create indicator variables in JMP.
- Go to *Analyze*, *Fit Model*.
- Put the response variable in the *Y* box.
- Select the explanatory variable(s) and click the *Add* button to put the variables in the *Construct Model Effects* box.
- Highlight the explanatory variable you would like to include in the quadratic term in the *Select Columns* box on the left.
- Click *Macros* and select *Polynomial to Degree* to add the interaction into the *Construct Model Effects* box.
- Click on the red triangle next to *Model Specification* and uncheck *Center Polynomials*.
- Click Run.
- To show the regression equation, click on the red triangle next to *Response* and select *Estimates*, *Show Prediction Expression*.

Normal Quantile Plot and Histogram for Residuals using Fit Model:

- Go to the *Fit Model* output.
- Normal Quantile Plot:
 - Click on the red triangle next to *Response*.
 - o Go to Row Diagnostics, then click on Plot Residual by Normal Quantiles.
- Histogram:
 - Click on the red triangle next to *Response*.
 - o Go to Save Columns, the click on Residuals.
 - Go to *Analyze*, *Distribution*.
 - Put *Residual* in the *Y*, *Columns* box.
 - Click *OK*.

Transformation: Create New Variable that is the Natural Log of a Different Variable:

- Go to *Cols, New column*.
- Name the new column with natural log in it (e.g., ln(variable)).
- Select Column Properties, Formula.
- On the left, click on *Transcendental*, *ln*.
- Click on the variable you would like to transform in the *Columns*. Click *OK* and *OK*.

Leverage/Influence Plot:

- Fit your model using *Fit Model*.
- In your *Fit Model* output, click on the red triangle next to *Response* and select *Save Columns*, *Residuals*.
- Go back to the original JMP data set.
- Compute the standardized residuals:
 - Go to *cols*, *New Column*.
 - o Give the column a name of "Standardized Residuals."
 - Select Column Properties, Formula.
 - Click on the column *Residual*, click on the ÷ button in the middle, then type the value of your root mean square error in the denominator. You can find the value of your root mean square error in the *Fit Model* output.
 - Click *OK* and *OK*.
- Store your leverage values (hats):
 - Go back to your *Fit Model output*. Click on the red triangle next to *Response* and select *Save Columns*, *Hats*.
- Store your Cook's D values:
 - Go back to your *Fit Model output*. Click on the red triangle next to *Response* and select *Save Columns*, *Cook's D Influence*.
- Create the Leverage and Influence Plot:
 - Click on *Graph, Graph Builder*.
 - Drag the *Standardized Residuals* to the Y-axis and the *Leverage Values* (labeled as h) to the *x*-axis.
 - Drag the *Cook's D* to the *Size* box on the right.
 - To get a legend for *Cook's D*, drag *Cook's D* to the *Color* box on the right.
 - Change the *x*-axis label to "Leverage Values" and the title to "Leverage and Influence Plot."
 - Unclick the small icon at the top that shows a small scatterplot with a curve on it.
 - Right click on the y-axis and select Axis Settings....
 - In the Reference Lines section, type 0 in the Value box.
 - Click *Add* and *OK*.
 - Add additional horizontal reference lines at -2 and +2, and possibly at -3 and +3.
 - Right click on the x-axis and select Axis Settings....
 - o In the *Reference Lines* section, type the cutoff values for high and very high leverage in the *Value* box. For the high value, type in *High Leverage* in the *Label* box. For the very high value, type in *Very High Leverage* in the *Label* box.
 - Click *Add* and *OK*.
 - Click Done.

<u>Prediction Intervals and Confidence Intervals for a Mean at Specific Values of X* using Fit Model:</u>

- Go to the original JMP data set and scroll down to the end. We need to create a row for the X* value we want to compute the intervals for. Type the value of X* in the last row and appropriate column. Do not type any other values in the data table.
- Fit your model using *Fit Model*.
- Go to your *Fit Model* output. Click on the red triangle next to *Response* and select *Save Columns*, *Mean Confidence Interval*.
- Click on the red triangle next to *Response* again and select *Save Columns*, *Indiv Confidence Interval*.
- Go to your data table and look at the last four columns in the last row where you typed the value of X*.
- The columns for 95% Mean gives the confidence interval and the columns for 95% Indiv gives the prediction interval. Note that this also gives the intervals for every observation in the dataset.

Model Selection-RMSE, R², Adjusted R², and Mallow's C_p:

- Go to Analyze, Fit Model.
- Put the response variable in the *Y* box.
- Select the explanatory variable(s) and click the *Add* button to put the variables in the *Construct Model Effects* box.
- For *Personality* on the top right, select *Stepwise* from the dropdown menu.
- Click Run.
- Click on the red triangle next to Stepwise Fit and select All Possible Models.
- Change the *Number of best models to see*: to be a large number such as 100.
- Click *OK*.
- Each model will be shown at the bottom. To see the RMSE, R², Adjusted R², Mallow's C_p for each model, select the radio button in the last column for each model separately. The values are shown towards the top.

Model Selection-Backward Elimination:

- Go to *Analyze*, *Fit Model*.
- Put the response variable in the Y box.
- Select the explanatory variable(s) and click the *Add* button to put the variables in the *Construct Model Effects* box.
- For *Personality* on the top right, select *Stepwise* from the dropdown menu.
- Click Run.
- Change the *Stopping Rule* to *P-value threshold*.
- Change the *Prob to Leave* to 0.05.
- Change the *Direction* to *Backward*.
- Select *Enter All* and click *Go*.
- The steps are shown at the bottom in the *Step History* and the chosen variables are checked in the *Current Estimates*.
- Click Run Model.

<u>Partial F-test: Hypothesis Test for Categorical Variable with More Than 2 Levels in Regression using Fit Model:</u>

- Fit your model using *Fit Model*.
- Click on the red triangle next to *Response* and select *Estimates, Custom Test*.
- At the bottom of your output, a list of boxes will appear. At the very bottom, click *Add Column* until you have the same number of columns as the number of predictors you plan to test (the number of categories for your categorical variable minus one).
- In the first column, replace the 0 with a 1 for the first category.
- In the second column, replace the 0 with a 1 for the next category.
- Repeat for all categories.
- Click Done.
- The output shown in the box is the results for the test of your categorical variable.

One-way ANOVA using Fit Y by X:

- Go to *Analyze*, *Fit Y by X*.
- Put the *quantitative* (*response*) variable in the *Y*, *response* box and *categorical* (*explanatory*) variable in the *X*, *factor* box.
- Click *OK*.
- Click on the red arrow next to *One-way Analysis* and select *Means/Anova*.
- To see the standard deviations and sample sizes for each group, click on the red arrow next to *One-way Analysis* and select *Means and Std Dev*.
- To see the histograms for each group, click on the red arrow next to *One-way Analysis* and select *Display Options, Histograms*.
- To see the boxplots for each group, click on the red arrow next to *One-way Analysis* and select *Display Options, Boxplots*.
- To create a normal quantile plot:
 - Click on the red arrow next to *One-way Analysis* and select *Save, Save Residuals*.
 - Go to your data table; you should see a column labeled *Y centered by X*. These are the residuals.
 - Go to *Analyze*, *Distribution*.
 - Put your residual variable in the *Y*, *Columns box*.
 - Click *OK*.
 - Click the red triangle next to your variable name and select *Normal Quantile Plot*.
- If you want to create a residual plot:
 - Click on the red arrow next to *One-way Analysis* and select *Save, Save Residuals* if you have not already.
 - Click on the red arrow next to *One-way Analysis* and select *Save, Save Predicted*.
 - Go to the data table.
 - Select *Graph*, *Graph Builder*, and put your residuals on the y-axis and predicted values on the x-axis. JMP will label the column of residuals as *Y centered by X* and label the column of predicted values as *Y mean by X*.
 - Put appropriate labels on your axis.

One-way ANOVA using Fit Model:

- Go to *Analyze*, *Fit Model*.
- Put the *quantitative* (*response*) variable in the *Y* box and *categorical* (*explanatory*) variable in the *Construct Model Effects* box.
- Click Run.
- To create a normal quantile plot, click on the red arrow next to *Response* and select *Row Diagnostics, Plot Residual by Normal Quantiles*.

Multiple Comparisons with One-way ANOVA using Fit Y by X:

- Go to Analyze, *Fit Y by X*.
- Put quantitative variable in the *Y*, *Response* box and your categorical variable in the *X*, *Factor* box.
- Click *OK*.
- Click on the red triangle next to *Oneway Analysis* and select *Means/ANOVA*.
- To change the confidence level, click on red triangle next to *Oneway Analysis* and select *Set* α *level* and enter your value.
- Click on the red triangle next to *Oneway Analysis* and select *Compare Means*, *All Pairs*, *Tukey HSD*.

Scatterplot with Manually Added Simple Linear Regression Lines using Graph Builder:

- Click on *Graph*, *Graph Builder*.
- Drag your quantitative explanatory variable to the *x*-axis and the quantitative response variable to the *y*-axis.
- Unclick the small icon at the top that shows a small scatter plot with a curve on it.
- Right click on the scatterplot. Go to *Customize*....
- Click on the + on the top left.
- Select *Templates*, *Y Function*.
- Replace $_function_of_x_$ with the value of your y-intercept + value of the slope * x. For example, if your simple linear regression equation was $\hat{y} = 10 + 2 * x$, you would type in 10 + 2 * x.
- If you would like to change the color of the line, select *Templates*, *Pen Color*. Replace "blue" with any standard color (e.g., "red").
- Click on *Apply*.
- If you would like to add more simple linear regression lines, click on the + again and add in the new equation.
- If you would like to add text describing what the lines refer to, click on the + again.
- Select *Templates*, *Text*.
- Replace _x_ with the x coordinate you would like the text to show up with and _y_ with the y coordinate you would like the text to show up.
- Replace _*text*_ with your text (e.g., "Average X2").
- Once you have finished adding regression lines and text, click on OK.
- Click *Done* on the top left.

Contingency Tables and Chi-Square Test for Independence using Fit Y by X:

- Go to *Analyze*, *Fit Y by X*.
- Put the *categorical* (*response*) variable in the *Y*, *response* box and *categorical* (*explanatory*) variable in the *X*, *factor* box.
- Click *OK*.
- To see the expected counts, click on the red arrow next to *Contingency Table* and select *Expected*.

<u>Logistic Regression with Quantitative Explanatory Variable(s):</u>

- Go to *Analyze*, *Fit Model*.
- Put the *categorical* (*response*) variable in the *Y* box and *quantitative* (*explanatory*) variable in the *Construct Model Effects* box.
- Click Run.
- To remove the points for the graph, click on the red arrow next to *Nominal Logistic Fit*, select *Plot Options*, and uncheck *Show Points*.