

# **General Notes**

#### **About These Instructions**

These instructor notes provide information on using the JMP Introductory Lab Activities in the classroom. For each activity, the following information is provided:

- Core Standard the Common Core Standard addressed by the activity
- Objectives activity learning objectives
- Time Required a recommendation on how much time to allow for the activity
- Data Set the data used (from the Help menu > Sample Data in JMP)
- Prerequisites any prior knowledge, completed activity, or other requirements
- Notes notes related to the activity, the scenario, JMP, and/or the data

For some activities, additional discussion questions and statistical concept notes are also provided.

Additional JMP Introductory Lab Activity resources, such as an activity correlation with popular textbooks and an activity scoring guide, are available at <a href="mailto:imp.com/APStats">imp.com/APStats</a>.

# **Getting Started with JMP®**

If you're new to JMP, there are a number of ways to get started. Go to <a href="mailto:jmp.com/teach">jmp.com/teach</a> for a complete list:

- View a <u>Getting Started with JMP</u> video. These videos provide the basics for navigating and using JMP in the classroom.
- Visit the <u>Learning Library</u> (jmp.com/learn), which has a series of one-page guides, videos and tutorials. These can be viewed online, or the one-page guides can be downloaded as a complete set.

# Other JMP® Resources

A variety of resources have been developed for teaching and learning statistical concepts, including:

- A Case Study Library (jmp.com/cases).
- Interactive <u>concept discovery modules</u> (jmp.com/tools), for simulating and exploring core statistical concepts (sampling distributions, confidence intervals, hypothesis tests, probability distributions, regression and ANOVA).
- Additional JMP tutorials, available at http://web.utk.edu/~cwiek/201Tutorials/.



Activity 01: Working with Data

### **Core Standard:**

"Students understand that visual displays and summary statistics condense the information in data sets into usable knowledge."

### **Objectives:**

- Gain familiarity with basic navigation in JMP.
- Learn how to download data from the Internet and create a JMP data table.
- Learn how to create basic graphs in JMP and to copy and paste JMP output into a report.

Time Required: 45 minutes

### **Prerequisites:**

- Students should have basic knowledge of how to use a program for writing reports (Word, Google Docs, etc.).
- Students should be familiar with the school's login procedure and know how to access the Internet from the school's computers.
- All students should have access to JMP for all activities. The JMP SE CD can be used to install JMP on either Windows or Macintosh<sup>®</sup>. The two versions are almost identical in their commands.
- This activity can be done almost anywhere within the AP Statistics curriculum, as
  it requires little or no experience using JMP and little or no statistical knowledge.
  Its major objective is to learn how to download data.

- The JMP disk can be used to install JMP on either a Windows machine or a Macintosh. The two versions are almost identical in their commands.
- DASL (pronounced "dazzle") is an online library of data files and stories that illustrate the use of basic statistics methods. To copy and paste data into JMP with the column names, use Edit > Paste with Column Names. If this isn't used, the first row of data will be the column names.
- The selection tool on the toolbar is used for selecting JMP output. The toolbar is hidden by default on Windows. To always view the toolbar, go to File > Preferences > Windows Specific, and change Auto-hide menu and toolbars to Never.



Activity 02: Describing Categorical Data

### **Core Standard:**

"Students understand that visual displays and summary statistics condense the information in data sets into usable knowledge."

### **Objectives:**

- Gain more familiarity with basic navigation in JMP.
- Use JMP to display categorical data with a bar chart, mosaic plot, table of frequencies, and probabilities.

Time Required: 20 to 30 minutes

Data Set: Denim.jmp

### **Prerequisites:**

- Students should have experience producing numerical summaries and graphical displays by hand and/or with graphing calculators.
- Students should have basic knowledge of how to use a program for writing reports (Word, Google Docs, etc.).
- Students should have completed the Activity 01: Working with Data

- May be done after Activity 03: Describing Numerical Data, either as an activity or a homework assignment.
- Categorical variables (either ordinal or nominal) can show a mosaic plot in addition to the bar graph. A mosaic plot is essentially a stacked bar chart where each bar is proportional to the number of cases at each level of the variable. The probabilities in the bar have a sum of one (1).
- The default setting in JMP is for the table of frequencies and probabilities to be automatically displayed with the bar chart. Default settings can be changed in the JMP Preferences (under File > Preferences > Platforms > Distribution). Note: on the Mac, Preferences are under the JMP menu.
- Thread Wear Measured is a numeric variable. By converting Thread Wear
  Measured to an ordinal variable, Thread Wear, the graphical and numerical
  output we get from JMP are different.



# Activity 03: Describing Numerical Data

### **Core Standard:**

"Use appropriate statistics to summarize center (median, mean) and spread (interquartile range, standard deviation) of the data set; interpret changes in shape, center and spread in the context of the data set, accounting for possible effects of extreme data points."

## **Objectives:**

- Gain more familiarity with basic navigation in JMP.
- Use JMP to display a histogram, box plot, stem and leaf plot, and summary statistics.
- Create a report, incorporating graphs and tables from JMP.

Time Required: 45 minutes

Data Set: Big Class.jmp

### **Prerequisites:**

- Students should have experience producing numerical summaries and graphical displays by hand and/or with graphing calculators.
- Students should have basic knowledge of how to use a program for writing reports (Word, Google Docs, etc.).
- Students should have knowledge of opening, closing and saving files.

- JMP includes an extensive online help system. It contains a table of contents and
  can be used to search for a specific topic. In addition, JMP has context sensitive
  help. You can access it by selecting the help tool (?) from the tools toolbar and
  clicking inside a data table or display. JMP opens help specific to the clicked-on
  item.
- In general, you can undo a mistake by using the Edit > Undo command.
- **Gray triangles** are disclosure icons that reveal (or hide) additional information.
- Red triangles reveal additional options.
- To select data in the data table, click on the first item and drag to the final item to select contiguous items. Alternately, you can click on the first item, hold down the shift key, and click on the final item. Either method automatically converts to the

- selection tool while in the data table. Ctrl-click (Command-click on the Macintosh) to select noncontiguous rows or columns. To select items from the Distribution window, use the appropriate selection icon and the same selection methods.
- Other Macintosh equivalent key commands are: Ctrl = right-click, Option = Alt.
- To edit column names, double-click on them.
- The **Summary Statistics** table displays the mean, standard deviation, standard error of the mean, upper and lower 95% confidence limits for the mean, data set size and the median. To display additional statistics, click on the red triangle next to **Summary Statistics** and select **Customize Summary Statistics**.
- To change the default look and feel of JMP, or default output or statistics, use the **JMP Preferences** (under **File**, or under **JMP** on the Mac).



Activity 04: Graphically Assessing Normality

### Core Standard:

"The student understands that statistical methods take variability into account to support making informed decisions based on quantitative studies to answer specific questions."

### **Objectives:**

• Use the 68-95-99.7 rule and a Normal Quantile Plot to investigate normality

**Time required:** Activity - 45 minutes, Homework - 45 minutes

Data Set: Fitness.jmp

# **Prerequisites:**

 Students should have experience checking for normality by hand and/or with graphing calculators.

The activity should be printed for use as a worksheet.

### Notes:

To change a the layout of the histogram from vertical to horizontal, use the red triangle menu for weight, then select Display Option > Horizontal Layout. The default can be changed in JMP Preferences (File > Preferences > Platforms > Distribution, or JMP > Preferences > Platforms > Distribution on a Mac.)

# **Concept Notes:**

- It is important for the student to understand that these checks for normality do not guarantee that the population data are in fact normally distributed. These procedures can only show that normality is plausible and even though sample data approximately follow the 68-95-99.7-rule and show a linear pattern in the normal quantile plot, the population data still may not be normally distributed. These procedures can only show that you have no evidence to think the population data are not normal.
- Discuss with the students the danger in using only the 68-95-99.7-rule with small sets of data. For small data sets from a normal population, these checks may not be accurate even though the population is normal. For example the 68-95-99.7rule would be difficult to use with a sample size of 10. Each data point would

account for 10% of the data, and following the two or three standard deviation pattern would not be possible.



Activity 05: Least Squares Regression

### **Core Standard:**

"Summarize bivariate quantitative data by giving a regression line and a measure of goodness of fit."

### **Objectives:**

 Use JMP to display a scatter plot, least-squares regression line, and a scatter plot with least-squares regression line for grouped data.

Time Required: 60 minutes

Data Set: Big Class.jmp

### **Prerequisites:**

- Students should have basic knowledge of how to use JMP.
- Students should have experience producing scatter plots by hand and/or with graphing calculators.
- Students should have experience calculating least-squares regression lines by hand and/or with graphing calculators.
- Students should have experience analyzing scatter plots and least-squares regression lines for both grouped and ungrouped data.

- JMP includes an extensive online help system. It contains a table of contents and/or can be used to search for a specific topic. In addition, JMP has contextsensitive help. You can access it by selecting the help tool (?) from the toolbar and clicking inside a data table or report. JMP opens help specific to the clickedon item
- The **Analysis of Variance** computations would be useful for concepts generally developed late, if at all, in an introductory statistics course.
- As students work through the activity, they need to follow directions carefully.
   Anticipating steps in the activity will generally result in incorrect displays.
- Students will get different results based on which variable they chose as the explanatory variable.
- A simple simulation for exploring least squares regression lines, demoLeastSquares, is available under Help > Sample Data > Teaching Demonstrations.

# JMP<sup>®</sup> Introductory Lab Activities

# **Instructor Notes**



Activity 06: Geometric Probability

### **Core Standards:**

"The Law of Large Numbers provides a basis for estimating certain probabilities by use of empirical relative frequencies.

The probability of an outcome can be interpreted as an assertion about the long-run proportion of the time the outcome will occur if the random experiment is repeated a large number of times. "

# **Objective:**

- Simulate the geometric probability using JMP.
- Investigate how the size of the sample has an impact on how close we get to the true probability.

Time required: 45 minutes

# **Prerequisites:**

- Students understand how to find the true probability of a geometric distribution.
- Students are familiar with the Law of Large Numbers.
- It is important that before this distribution is simulated on JMP that the students have played the game in class. They need to understand the rules of the game and have an intuitive idea of the time required to win.

### Notes:

- The students that are not comfortable with the computer may have difficulty when they are entering a script formula.
- When completing the directions for #4 be sure they use the "+" from the program and not their keyboard.
- This can easily be changed to a Binomial Probability with a change in the directions from #4. Instead of choosing Geometric, choose Binomial. The directions can easily be adapted for a binomial distribution.

### **Concept Notes:**

The purpose of this activity is to simulate a large sample of a distribution quickly. The student can easily change the sample size and calculate the simulated probability. Be sure to use this opportunity to explain how, as the sample size becomes larger, the

student will have a clearer picture of the true probability—the **Law of Large Numbers**. Often in statistics we do not know the true probability of a situation. In this activity the student can easily calculate the true probability and compare it to samples of various sizes.

It is imperative that the student knows how to calculate the true probability using the geometric formula before attempting this activity.



Activity 07: Sampling Variation

### **Core Standards:**

"The Law of Large Numbers provides a basis for estimating certain probabilities by use of empirical relative frequencies.

Statistics is a process for making inferences about population parameters based on a sample from that population."

# **Objective:**

• Investigate sampling variability and how it relates to the size of the sample.

Time Required: 45 minutes

Data Set: Pollen.jmp

#### Notes:

- When the students are investigating the data and are concerned with negative values, explain that often data is measured using an estimated ideal value and then rescaled using zero as the center. All values that fall below this ideal become negative and data that falls above are positive.
- Problem #3: This activity may be done individually or as a group activity. This
  might be an ideal situation for students to see how their individual samples vary.
  There may be some students that have the same mean for their sample as in the
  population but by viewing all the samples together they will visually see the
  variability from sample to sample.
- Problem #4: The students should visualize that when the sample size is 10, the histogram may not necessarily look like a normal distribution. This should reinforce why checking for normality with a small set of data is often difficult.
- Problem #5: The students should see that when the sample size is increased, you are more likely to see that the sample appears more similar to the population. Expand on this by using the data from the entire class to visualize this variability.

# **Concept Notes:**

 This data set is used because it has over 3,000 observations and is in JMP, which is easily accessible. This is not a good data set to use to investigate the Central Limit Theorem because all of the variables have distributions that are

- approximately normal. It is ideal for visualizing sampling variability because of the large number of observations.
- We have described the data in Pollen as the entire population. In a realistic
  setting, we most often do not have access to the entire population. Explain this to
  the students because they may not see the relevance of taking a sample when
  the entire population is available. Cornelius does not have access to the entire
  population. He does not want to sample the entire corn field, only a random
  sample.
- It is important that this activity is done before investigating the sampling
  distribution of the mean. The sampling distribution of the mean will be
  approximately normal, regardless of the shape of the population distribution,
  provided the sample size is large enough. The sample of individual observations
  will generally follow the shape of the population, which may not be normal. In this
  activity we are looking at a random sample of individual observations.
- You can easily connect this activity to the Law of Large Numbers and explain how the sample size is related to the variability in the estimates of the population mean.



Activity 08: Confidence Interval for a Proportion

### Core standard:

"Use data from a sample survey to estimate a population parameter."

# **Objectives:**

- Use JMP to graphically display categorical data.
- Use JMP to calculate confidence intervals for a proportion using different confidence levels.

Time Required: 20 to 30 minutes

Data Set: Students.jmp

### **Prerequisites:**

• Students should have experience producing confidence intervals for a proportion by hand and/or with graphing calculators.

- You may wish to substitute a data set from DASL or another source.
- Confidence Interval for One Proportion calculators are available in the JMP
  Help under Sample Data > Calculators. These can be used with either data or
  summary statistics and can be used for exploring how the width of the confidence
  interval changes as the confidence level and sample size change. These
  calculators can calculate a confidence interval using either the Normal
  Approximation or the Binomial.



Activity 09: Confidence Interval for a Mean

### Core standard:

"Use data from a sample survey to estimate a population parameter."

### **Objectives:**

- Use JMP to graphically display and summarize continuous data.
- Use JMP to calculate confidence intervals for a mean using different confidence levels.
- Select, hide and exclude data to isolate the subjects of interest.

Time Required: 20 to 30 minutes

Data Set: Big Class.jmp

## **Prerequisites:**

• Students should have experience producing confidence intervals for a mean by hand and/or with graphing calculators.

- When a row is hidden, the observation will not appear in any graph. If a row is
  excluded, the observation will not be included in any future calculation or
  analysis (it will still be in any previously run analysis.)
- Confidence Interval for One Mean calculators are available in the JMP Help under Sample Data > Calculators. These can be used with either data or summary statistics, can calculate either a z- or a t- confidence interval, and can be used for exploring how the width of the confidence interval changes as the confidence level, sample size and standard deviation change.



Activity 10: Hypothesis Testing, the *z*-Test

### Core Standard:

"Understand statistics as a process for making inferences about population parameters based on a random sample from that population."

### **Objective:**

 Use JMP to conduct a hypothesis test for a mean when the standard deviation is known.

Time Required: 45 minutes

**Data Set:** Data is in a table in the activity

### **Prerequisites:**

- Students should have experience conducting tests of hypotheses by hand and/or with graphing calculators for a mean when the standard deviation is known.
- Students should have experience writing conclusions in the context of the problem for hypothesis tests.

- The z-test is rarely used in practice; most often the standard deviation of the population is unknown.
- As students enter data into the table, they need to be careful to enter the correct IQ for the student in each row.
- Students need to choose the correct *p*-value based on their alternative hypothesis. A brief discussion of the three *p*-values reported by JMP may be necessary.
- Excluding the row with Aaron allows Aaron to remain in the data table, but
  Aaron's value will not be used in any future analysis. If you wish to delete Aaron,
  select Aaron from the data table, and then choose Rows > Delete Rows from
  the menu bar (use caution; deleted data is not recoverable after closing the
  document. Excluding is the recommended method for removing data from an
  analysis).
- Hypothesis Test for One Mean calculators are available in the JMP Help under Sample Data > Calculators. These can be used with either data or summary statistics, can perform either a z-test or a t-test, and can be used for exploring

how test results (test statistics and $p$ -values) change as the confidence level, sample size and standard deviation change.



Activity 11: Hypothesis Testing, the *t*-Test

### **Core Standard:**

"Evaluate reports based on data."

# **Objective:**

 Use JMP to conduct a hypothesis test for a mean when the standard deviation is unknown.

Time Required: 30 minutes

**Data Set:** Data is in a table in the activity

# **Prerequisites:**

- Students should have experience conducting tests of hypotheses by hand and/or with graphing calculators for a mean when the standard deviation is unknown.
- Students should have experience writing conclusions in the context of the problem for hypothesis tests.
- Students should have had experience creating a data table in JMP.

- JMP includes an extensive online help system. It contains a table of contents and/or can be used to search for a specific topic. In addition, JMP has context sensitive help. You can access it by selecting the help tool (?) from the tools toolbar and clicking inside a data table or report. JMP opens help specific to the clicked-on item.
- Students need to choose the correct *p*-value based on their alternate hypothesis. A brief discussion of the three *p*-values reported by JMP may be necessary.
- The Normal Quantile Plot option adds a graph to the report that is useful for visualizing the extent to which the variable is normally distributed. If a variable is normal, the normal quantile plot approximates a diagonal straight line. This kind of plot is also called a quantile-quantile plot, or Q-Q plot.
- The Hypothesis Test for One Mean calculators (under Help > Sample Data > Calculators) can also be used here.



Activity 12: Hypothesis Testing, Paired *t*-Test

### **Core standard:**

"Evaluate reports based on data."

# **Objective:**

 Use JMP to conduct a hypothesis test for a comparison of two means for dependent samples (matched pairs).

Time Required: 30 minutes

Data Set: Therm.jmp

### **Prerequisites:**

- Students should have experience conducting tests of hypotheses by hand and/or with graphing calculators for a matched pairs design.
- Students should have experience writing conclusions in the context of the problem for hypothesis tests.

- In order to conduct a paired *t*-test the paired data must be in separate columns. For a two-sample *t*-test the data (the Y, Response) must be in one column and the labels for the X, Factor must be in a separate column.
- Students should understand that a paired *t*-test is the same as a one-sample *t*-test on the differences between the paired observations.
- The test can be conducted with either the **Distribution** platform, where the difference is the Y, **Column** and the hypothesized mean difference is zero, or with the **Matched Pairs** platform, where the paired responses are both used.
- This is a test of the null hypothesis that the mean difference is zero. JMP reports a confidence interval for the difference (in both Distribution and Matched Pairs).
- Discussing the confidence interval can help students interpret the test results (i.e., if the interval does not contain, zero there is evidence that the population difference is nonzero.)



Activity 13: Hypothesis Testing, Two-Sample *t*-Test

### **Core Standard:**

"Make inferences and justify conclusions from observational studies."

# **Objective:**

• Use JMP to conduct a hypothesis test for a comparison of two means.

Time Required: 45 minutes

Data Sets: HTWT12.jmp and HTWT15.jmp

### **Prerequisites:**

- Students should have experience conducting tests of hypotheses for a comparison of two means by hand and/or with graphing calculators.
- Students should have experience writing conclusions for hypothesis tests in the context of the problem.
- Students should have a good conceptual understanding of p-values.

- When conducting the significance test, students may wonder why they choose
  the test with unequal variances rather than conducting the Means/Anova/t-Test.
  Conducting a two-sample t-test for means assuming unequal variances is a more
  conservative approach. The Means/Anova/t-Test assumes equal variances and
  pools the data to estimate population variability. Many statisticians suggest using
  the unequal variance approach with an introductory statistics class.
- There will probably be a great deal of confusion in the calculation of the p-value for the 12-year-old comparison. The alternative hypothesis in this case is one-sided (Ha: the mean for males is > the mean for females). As a result, the correct p-value is found under prob > t.
- Students may question why a hypothesis test was even necessary, and in reality, a test was not required to conclude insufficient evidence to show 12-year-old males are taller than 12-year-old females.
- Reinforce the difference between a paired t-test and a two-sample t-test.
- In order to conduct a paired *t*-test, the paired data must be in separate columns. For a two-sample *t*-test, the data (the Y, Response) must be in one column and the labels for the X, Factor must be in a separate column.



Activity 14: Hypothesis Testing Proportions

### Core Standard:

"Summarize, represent, and interpret data on a single count variable."

# **Objectives:**

- Conduct a one-proportion test on JMP.
- Compare the results from the graphing calculator to the results in JMP.

Time Required: 30 Minutes

Data Set: Big Class.jmp

### Notes:

• JMP does not have an option for a one-proportion *z*-test. However, the chisquare goodness of fit test can be used to produce an equivalent result since the data are categorical.

### **Concept Notes:**

- It is important that the student understand that a proportion test must involve categorical data.
- The test of H<sub>o</sub>: p=0.5 vs. H<sub>A</sub>: p≠0.5 using a one sample z-test is equivalent to a chi-square goodness-of-fit test for one-way table with two categories. In the chi-square test, the null hypothesis should be that each of the two-category proportion was 0.5. Also note that the chi-square test is equivalent to the two-tailed test.
- A proportion test parallels a chi-square test. If the student squares the z-score
  they will see the chi-square value. To get the correct chi-square value, do not
  round the z-score.
- It is important that the students check the assumptions for this test.
- **Hypothesis Test for Proportion** calculators are available in the JMP Help under **Sample Data > Calculators**. These calculators conduct *z*-tests.



Activity 15: Exploring Categorical Data

### **Core Standard:**

"Summarize, represent and interpret data on two categorical variables."

# **Objective:**

Use JMP to display a mosaic plot and contingency table for categorical data.
 Row and column percentages will also be found.

Time Required: 45 minutes

Data Set: Denim.jmp

# **Prerequisite:**

• Students should have experience producing numerical summaries and graphical displays of categorical data by hand and/or with graphing calculators.

- The same command (Analyze > Fit Y by X) is used for both numeric and categorical data. JMP decides which analysis to use based on the modeling types for the variables selected.
- With two categorical variables (two-way tables), the mosaic plot that JMP displays includes a bar on the far right that represents all of the data.
- To better understand the mosaic plot, right click on the plot, and select Cell Labeling > Show Percents. These values correspond to the Row % values in the Contingency Table.



Activity 16:  $\chi^2$  Goodness-of-Fit Test

### **Core Standard:**

"Make inferences and justify conclusions from observational studies."

# **Objective:**

 Use JMP to calculate expected values in a contingency table and run a chisquare goodness-of-fit test.

Time Required: 45 minutes

Data Set: Denim.jmp

### **Prerequisites:**

- Students should have experience producing numerical summaries and graphical displays of categorical data by hand and/or with graphing calculators.
- Students should have completed Lab Activity Exploring Categorical Data.

### Notes:

• The chi-square statistic that is calculated by the formula shown in the student directions and found in introductory textbooks is the Pearson test statistic.



Activity 17:  $\chi^2$  Test of Independence

### **Core Standard:**

"Use data from a randomized experiment to compare two treatments."

# **Objective:**

 Use JMP to calculate expected values in a contingency table and run a chisquare test of independence.

Time Required: 45 minutes

Data Set: Denim.jmp

# **Prerequisites:**

- Students should have experience producing numerical summaries and graphical displays of categorical data by hand and/or with graphing calculators.
- Students should have completed Exploring Categorical Data (X<sup>2</sup> Goodness-of-Fit Test can be done independently of X<sup>2</sup> Test of Independence.)

# Notes:

• The chi-square statistic that is calculated by the formula shown in the student directions and found in introductory textbooks is the Pearson test statistic.



Activity 18: Inference for Regression

### **Core Standard:**

Summarize bivariate quantitative data by giving a regression line and a measure of goodness of fit.

# **Objective:**

Use JMP to conduct a hypothesis test for linear regression.

Time Required: 45 minutes

Data Set: Body Measurements.jmp

### **Prerequisites:**

- Students should have experience conducting tests of hypotheses by hand and/or with graphing calculators for linear regression.
- Students should have experience writing conclusions for hypothesis tests in the context of the problem.

- Students can convert measurements to the English system and use those values for the regression analysis (using the Formula Editor in JMP).
- A simple simulation for exploring least squares regression lines, demoLeastSquares, is available under Help > Sample Data > Teaching Demonstrations.
- A more comprehensive least-squares regression simulation, Demonstrate
  Regression, can be found at jmp.com/tools. A JMP add-in, including this
  demonstration and other statistical concept modules, can be downloaded from
  this site (Interactive Teaching Modules Add-in, on the far right of the page).
  Downloading this add-in requires that you create a SAS profile. When you open
  the downloaded add-in in JMP, it will automatically install a new menu that
  provides access to all of the teaching modules.