2017–18 OCTAE Technical Assistance to States
Final Summary Report for the College of Menominee Nation

Prepared under contract to
U.S. Department of Education

RTI International
1618 SW First Avenue, Suite 300
Portland, OR 97201

Contact
Sandra Staklis
sstaklis@rti.org
503-428-5676

June 2018
Contents

Introduction ......................................................................................................................... 1

Data Analysis Resources .................................................................................................. 2
Data Conversion Template ................................................................................................. 2
  Figure 1. Data Conversion Template .............................................................................. 3
Data Analysis Instructions Workbook .............................................................................. 3
  Figure 2. Data Analysis Instructions Workbook .......................................................... 4
Data Analysis Examples Workbook ................................................................................. 4
  Figure 3a. Data Analysis Examples Workbook: Instructions ........................................ 5
  Figure 3b. Data Analysis Examples Workbook: Calculations ...................................... 5
  Figure 3c. Data Analysis Examples Workbook: Analysis and Interpretation ............... 6
Statistics Cheat Sheet ....................................................................................................... 6
Webinar ............................................................................................................................. 7
Conclusion ........................................................................................................................ 7

Appendix A. Data Analysis Instructions ............................................................................. A-1

Appendix B. Statistics Cheat Sheet ................................................................................... B-1
  Table 1. Descriptive Statistics .................................................................................. B-1
  Table 2. Inferential Statistics (Tests Comparing Group Values) ................................... B-1
Using Visuals With Data ................................................................................................. B-1
  Example 1. Box and Whisker Plot (Also Called a Box Plot) .................................. B-1
  Example 2. Bar Graph (Also Called a Bar Chart) ....................................................... B-2
  Example 3. Scatter Plot ............................................................................................... B-3

Appendix C. Webinar for College of Menominee Nation Data Team ......................... C-1
Introduction

The U.S. Department of Education’s Office of Career, Technical, and Adult Education (OCTAE), Division of Adult and Technical Education, sponsors an annual program to provide customized technical assistance (TA) to states. The TA projects are intended to improve states’ career and technical education (CTE) data collection and reporting capacity and strengthen the analysis of data contained within states’ accountability systems. Emphasis is placed on data related to the Carl D. Perkins Career and Technical Education Improvement Act of 2006 (Perkins IV). RTI International serves as the TA provider on behalf of OCTAE.

College of Menominee Nation (CMN) requested TA to increase its data analysis capacity. CMN staff possessed several years of student survey data on CTE participants but lacked the statistical background to analyze the data to support program review. CMN requested materials to develop instructor and administrator understanding of statistical analysis, provide guidance on answering questions related to program improvement, and function as templates for future analysis. CMN requested the use of Microsoft Excel for analyses and templates because it is the software most familiar and accessible to CMN staff.

RTI’s approach involved developing a series of complementary resources to outline a step-by-step process for performing data analyses. After consulting with CMN staff, RTI created a package of the following five resources:

1. **Data conversion template**: An Excel workbook to transform student data exports from CMN into a worksheet suitable for analysis.
2. **Data analysis instructions**: An Excel workbook containing guidance on steps for forming research questions and conducting student-level survey data analysis.
3. **Data analysis examples**: An Excel workbook containing example analyses following the approach outlined in the data analysis instructions workbook. The examples answer research questions that were proposed by CMN staff and RTI. The data used were welding student data from 2014–17.
4. **Statistics cheat sheet**: A Word document describing statistical terminology and tests, interpretation of statistical tests and ideas for visual displays of data.
5. **Webinar**: An interactive video presentation to CMN faculty and staff on how to use the resources and interpret results.

This report summarizes the materials provided to CMN as part of RTI’s TA to CMN during 2017–18. Select materials are provided in the appendices.
Data Analysis Resources

During 2017–18, the RTI research team met with CMN staff via teleconference to discuss site data and materials. The research team intermittently provided draft materials to CMN and received regular feedback on which components were most useful as well as what additional components could be included. The set of resources is described below.

Data Conversion Template

CMN indicated that student CTE data is output from its data system into an Excel worksheet, with multiple rows per student where each represents one term of enrollment. However, this format presents a problem for analyzing data because most basic statistical analysis tests in Excel require one row per student, with data about all terms of enrollment in the same row. While RTI could transform the provided data manually from multiline student data to single line student data for this project, CMN would be left with the onerous task of transforming the data every time the staff wanted to look at a new group of students or a different group of student measures.

To solve this problem, RTI created a Data Conversion Template in the form of an Excel workbook that uses macros to transform multiline student data into a single row per student. Figure 1 displays text from the Data Conversion Template. The worksheet pictured, labeled “Main page,” includes instructions and a button to transform the dataset. The macros sort the data chronologically by term and keep the most recent data for each student when there are multiple values across lines (e.g., if GPA is reported in each term, GPA from the last term will be kept). If there is only one value for a measure (e.g., grade from a course taken once), then that one value is kept, even if it was not recorded in the most recent term. The data to be transformed are entered into the “Data input” worksheet. The “Background” worksheet is locked for editing because the transformation calculation takes place there. The workbook outputs a new Excel worksheet with the transformed data.

This resource is valuable because it allows CMN staff to export data and transform it with a click of a button, saving its team the time it would take to convert the data manually or to create a second data output from the system. The workbook can be reused each time CMN creates a new output of student data with the same or different numbers of students and student measures.

The Data Conversion Template is included as a resource on the Perkins Collaborative Resource Network.
Data Analysis Instructions Workbook

CMN requested assistance with how to begin student data analysis in Excel. Many CMN staff members have little to no statistical background. RTI provided an Excel workbook containing the following:

- **Resources**: Suggested texts and websites for reference beyond the materials provided for this TA project.
- **General Instructions**: Guidance for moving from a data export to data analysis.
- **Question Instructions**: Guidelines for research question development.
- **Analysis Instructions**: A step-by-step process for answering research questions using survey data in Excel.
Figure 2 displays the first three steps of the analysis process from the Analysis Instructions worksheet. The complete analysis instructions are included in Appendix A. CMN staff reviewed and provided feedback on this workbook during the development process.

**Figure 2. Data Analysis Instructions Workbook**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question:</strong> Clearly state a concise research question to guide the analysis. See &quot;Question instructions&quot; tab for more details on constructing a research question.</td>
<td><strong>1</strong></td>
<td><strong>Determine what data is needed to answer the question</strong></td>
<td><strong>What information do you need to answer this question? How many pieces of information do you need? What years of data do you need? Do you need student-level data?</strong></td>
<td></td>
</tr>
<tr>
<td><strong>2</strong></td>
<td><strong>Define how the concepts will be measured</strong></td>
<td><strong>How the concepts will you use the data to answer the question? Are there different ways to define concepts or ideas in your research question? What are the concrete definitions you will use? Are there administrative definitions that can be used to guide the concrete definitions?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3</strong></td>
<td><strong>Select the sample for analysis</strong></td>
<td><strong>What individuals will be analyzed? Are there specific, identifiable individuals who should or should not be in the analysis? RTI suggests omitting any individuals who do not have usable information on all pieces of data to be used in the analysis.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4</strong></td>
<td><strong>Transform the data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Data Analysis Examples Workbook**

CMN provided a list of four questions that its staff was interested in exploring for program improvement. The research team used those questions to display how the data analysis process described in the Data Analysis Instructions Workbook could be used to answer questions related to program improvements. The research team also included two additional research questions based on CMN’s description of future questions of interest. CMN provided data on welding program participants from Fall 2014 to Fall 2017 for the example analyses. This workbook is not publicly available because it contains private student data.

For each research question, the Data Analysis Examples Workbook includes two worksheets: instructions and data. The instructions worksheet follows the step-by-step process described in the Data Analysis Instructions Workbook, and the data worksheet includes the relevant student data, calculations, and interpretation. Together, the pair of worksheets for each research question describes the process for answering each question (Figure 3a), shows the calculations to answer that question (Figure 3b), and provides example interpretations of the results (Figure 3c). CMN staff reviewed and provided feedback on the instructions, interpretations, and calculations for the workbook throughout the development process.
### Figure 3a. Data Analysis Examples Workbook: Instructions

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>CMN Question:</strong> Does a welding student’s math placement affect subsequent student success in welding courses?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1. Determine what data is needed to answer the question.</td>
<td>There are 2 items to measure: math placement course and welding course success. We will need data on both items to look at the relationship between math placement course and student success in welding. The question does not specify a time frame, so we will use data on students from all semesters and years.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2. Define how the concepts will be measured.</td>
<td>In this step we define how the concepts will be measured using the data. Specifically, we provide precise definitions for the concepts of “math placement course” and “student success.” Math placement course is well defined from the start since there are only a few options for what course students can be assigned to. We measure math placement course as a 3 category variable, where each value is one math placement course option: Foundation, SOAR, or College Ready. Student success in welding courses, however, is a fuzzier concept that can be defined in a number of ways: Did the student complete their welding course? How many courses were completed? Did the student pass his/her welding courses? What is the students’ GPA in welding courses? In this data, we have information on the students’ grade in each welding course. We will define student “success” in two ways using this grade data. First, how many welding courses did the student pass with a grade B or above? Second, what percentage of the welding courses taken did the student pass with a grade B or above?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Figure 3b. Data Analysis Examples Workbook: Calculations

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td><strong>Average number (rounded) of welding courses completed by students</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>College Ready</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Foundation</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>SOAR</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>All students</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td><strong>Interpretation:</strong> Students take an average of 6 welding courses. The number of courses taken is vary similar across placement groups.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td><strong>Average number (rounded) of welding courses passed by placement course</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>College Ready</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Foundation</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>SOAR</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>All students</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td><strong>Interpretation:</strong> Here, passing a class is defined as receiving a 2.5 or higher. Students, on average, pass 4 welding courses. Students placed in the College Ready course pass the most classes on average, while students placed in the SOAR course are pass the fewest classes on average.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td><strong>Percentage of welding courses passed by placement course</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>College Ready</td>
<td>80.06%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Foundation</td>
<td>58.10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>SOAR</td>
<td>44.11%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>All students</td>
<td>51.15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td><strong>Interpretation:</strong> Students, on average, pass 55% of the welding courses that they take. Students placed in the College Ready course pass the highest proportion of their classes (80%), while students placed in the SOAR course pass the lowest proportion of their classes (44%).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td><strong>ANOVA test for differences between groups on number of passed courses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>p-value</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Box and Plot (with link to data)*
CMN requested information on the uses and limitations of basic statistical tests and terms. Examples include mean, proportion, t-test, and regression. CMN also requested details on how to use visual aids when presenting data and how to interpret both regression and ANOVA test output. The research team provided a document containing guidance on statistical terminology, the uses and limitations of statistical tests, the use of visual aids with numeric data, and the interpretation of regression and ANOVA test output. The document references examples in the Data Analysis Examples file where applicable.

A version of the Statistics Cheat Sheet is in Appendix B. Alterations include changing examples from CMN-specific items to general items applicable to most colleges, as well as removing the output from the student data analysis.
**Webinar**

CMN requested a webinar for its data team, consisting of staff and faculty, for instruction on the data analysis process and resources provided by the research team. The research team hosted an interactive one-hour webinar with the data team in May 2018. The data team was presented with the data analysis instructions and examples ahead of time, and the webinar included time to ask questions about the resources. After the webinar, the research team provided the webinar slides to CMN as an additional resource. Slides from the webinar are included in Appendix C.

**Conclusion**

This TA to States project for CMN resulted in a package of resources to help staff analyze student data for program improvement purposes and translate questions for data analysis. Moving forward, CMN will be able to use these resources to transform student data, answer questions about student and program outcomes, and understand the interpretation of statistical tests. Other states and colleges may also benefit from the resources provided to CMN that are included in this report (Data Analysis Instructions, Statistics Cheat Sheet, webinar slides) and through the Perkins Collaborative Resource Network (Data Conversion Template).
# Appendix A. Data Analysis Instructions

**Question:** Clearly state a concise research question to guide the analysis. See "Question Instructions" tab for more details on constructing a research question.

1. **Determine what data is needed to answer the question**
   What information do you need to answer this question? How many pieces of information do you need? What years of data do you need? Do you need student-level data?

2. **Define how the concepts will be measured**
   How will you use the data to answer the question? Are there different ways to define concepts or ideas in your research question? What are the concrete definitions you will use? Are there administrative definitions that can be used to guide the concrete definitions?

3. **Select the sample for analysis**
   What individuals will be analyzed? Are there specific, identifiable individuals who should or should not be in the analysis? RTI suggests omitting any individuals who do not have usable information on all pieces of data to be used in the analysis.

4. **Transform the data**
   How will you manipulate the data to match your concrete definitions? Will you make categories (e.g. students above or below a threshold; students who do or do not take a class; students in who take disparate types of classes)? Will you combine existing categories? How many categories will you compare?

5. **Examine the data with visualizations (Visual analysis)**
   Visualizations are useful for understanding the context for numbers that result from calculations. If you are only working with one piece of data, you can examine the data using pie charts, bar graphs, and box and whisker plots (NOTE: Box and whisker plots are only chart options in Excel 2016). If you are working with more than one piece of data, you can use visualizations that incorporate both pieces of data. Do you have two categorical pieces of data? Try using a bar chart. Do you have one categorical piece of data and one continuous piece of data? Try using a bar chart or a line graph. Do you have two continuous pieces of data? Try using a scatter plot.

6. **Examine the data with calculations (Numeric analysis)**
   Types of calculations and statistics will depend on your question and on the types of data you are using. Is your question about describing just one group? Descriptive statistics, such as mean or percentiles, would be a good starting place. Is your question asking about a comparison between two or more groups? Inferential statistics comparing groups, such as t-tests and ANOVA tests would be a good starting place. Refer to the Statistics Cheat Sheet and the examples in the companion workbook for example tests and information about how to interpret output.

7. **Display the findings with visualizations**
   Revisit the initial visualizations. Do your initial visualizations reflect what your calculations show as the answer to your question? Do they skew or misrepresent the data? Do they show meaningful differences between groups that arose out of the calculations? Are there other ways you could display the findings that would tell a story about what the calculations show?
## Appendix B. Statistics Cheat Sheet

**Table 1. Descriptive Statistics**

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Purpose</th>
<th>Measures</th>
<th>Application/Example</th>
</tr>
</thead>
</table>
| Categorical variable (Also called: nominal variable or binary variable) | A measure or metric where individuals are assigned one value out of a fixed set of values. Each value represents a characteristic that the group shares. The characteristics can be conceptual (e.g., male, female) or numeric (e.g., 0, 1-5, 6-10, 11+) The values may be ordered, such that some values are higher or lower than each other (e.g. grades), or the values may be independent of each other and un-ordered (e.g., sex, program of study). | • Letter grades  
• Class-level  
• Program of study  
• Math placement course | Visually, categorical variables can be displayed using bar charts, histograms, or pie charts. Categorical variables can be shown in visual aids in relation to categorical and continuous variables using bar charts. |
| Continuous variable | A measure or metric where individuals take on one numeric value of an infinite number of possibilities (e.g., 1, 2, 3, 4, ... or .3, .5, .8, .9,... or 1000, 2500, 10000, ...). The measure or metric may or may not have a defined range (e.g., GPA has a minimum of 0). | • Placement test score  
• GPA  
• Number of courses passed | Visually, continuous variables can be explored using box and whisker plots or histograms. Continuous variables can be shown in visual aids in relation to other categorical variables with bar charts and in relation to continuous variables with scatter plots. |
### Table 1. Descriptive Statistics—Continued

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Purpose</th>
<th>Measures</th>
<th>Application/Example</th>
</tr>
</thead>
</table>
| Percentage/Proportion | Indicator of how many individuals in a group meet a criterion. These values are measured at one point in time in a defined group but can be compared over time or across groups. A percentage and a proportion can express the same value in different terms. A proportion can be transformed into a percentage by multiplying it by 100 and adding a percentage sign. For example, a 0.50 proportion could be expressed as a percentage by writing 50%. To be used with categorical or continuous variables. | • Percentage of courses passed  
• Percentage of students with an industry-recognized credential. | On average, students pass 55% of the courses that they take. Said differently, on average, the proportion of courses taken that students pass is 0.55. |
| Mean (Also called: average) | Use with continuous variables. Most common measure of central tendency for continuous variables. Better measure of central tendency than the median when there are no outliers (i.e. no extremely high or low values). The mean is the basis for many other statistical tests (e.g., t-test). | • Number of courses passed  
• Number of courses taken  
• SAT score  
• GPA | The mean, or average, number of courses taken students is 6 courses. |
| Median (Also called: 50th percentile) | Use with continuous variables. Common measure of central tendency for continuous variables. Better measure of central tendency than the mean when there are outliers (i.e. a few extreme values). For example, the mean value of GPA would be skewed downward if a few students have values of 0 for GPA, while the median GPA value would not be skewed downward. | • Number of courses passed  
• Number of courses taken  
• SAT score  
• GPA | The median 1st term GPA for students is 3.25, which means that 50% of students had a GPA higher than 3.25 and 50% of students had a GPA lower than 3.25. |
### Table 1. Descriptive Statistics—Continued

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Purpose</th>
<th>Measures</th>
<th>Application/Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviation</td>
<td>Use with continuous variables and interpret alongside the Mean.</td>
<td>• SAT score</td>
<td>“Small” deviation – indicates that most values are in a range closer to the mean (e.g. a mean GPA of 2.5 and a standard deviation of 0.1 indicates that most students receive between a 2.4 and 2.6 GPA.)</td>
</tr>
<tr>
<td></td>
<td>Average distance from the mean. The value judgment of whether a standard deviation is “small” or “large” depends on the range of values for a measure. For example, a standard deviation of 4 would be large for a measure that ranges from 1 to 10, but small for a measure that ranges from 1 to 100. This number is only useful if the measure is normally distributed. Not recommended with small sample sizes.</td>
<td>• GPA</td>
<td>“Large” deviation – indicates that most values are in a range that is father than the mean. (e.g. a mean GPA of 2.5 and a standard deviation of 1 indicates that most students receive between a 1.5 and 3.5 GPA.)</td>
</tr>
<tr>
<td>Percentiles/Quartiles</td>
<td>Use with continuous variables. Related to the Median. Provides information on distance from the median in terms of percentiles. Quartiles refer to the 25th, 50th, and 7th percentile. This statistic does not require an assumption of a normal distribution.</td>
<td>• SAT score</td>
<td>The 25th percentile of 1st term GPA for students is 2.50, which means that 75% of students had a GPA higher than 2.50 and 25% of students had a GPA lower than 2.50. Box and whisker plots are useful visualizations of quartiles.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• GPA</td>
<td></td>
</tr>
</tbody>
</table>
## Appendix B. Statistics Cheat Sheet

### Table 2. Inferential Statistics (Tests Comparing Group Values)

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Purpose</th>
<th>Application/Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parametric tests</strong></td>
<td>Compare group means and assume a normal distribution.</td>
<td>Examples explained below: T-tests, ANOVA</td>
</tr>
<tr>
<td><strong>Non-parametric tests</strong></td>
<td>Compare group medians and make no assumptions about the distribution.</td>
<td>Examples explained below: Wilcoxon (alternative to dependent sample t-test), Mann-Whitney (alternative to independent sample t-test), Kruskal Willis (alternative to ANOVA test)</td>
</tr>
</tbody>
</table>

**t-test**

Compares means or proportions across **two groups**. The groups may be related or unrelated. For use with one categorical variable (with two categories or groups) and one continuous variable.

The “p-value” is used it determine whether the groups are significantly different from one another. A p-value of less than 0.05 indicates that we are 95% confident that if we selected other groups, then we would still see group differences in average values.

- **One-sample t-test**
  - Compare data on a subgroup to existing data on the entire group (the “population”).
  - The measure that is being compared across the groups must be a continuous variable.
  - Test whether student GPAs in a specific major are similar or different from the average GPA for all students at the college.

- **Independent samples t-test**
  - Compare data from two separate (non-overlapping) samples (i.e., different people provided data for each sample).
  - The measure that is being compared across the groups must be a continuous variable.
  - Compare GPA’s across students who participate versus do not participate in extracurricular activities.
  - Compare number of courses passed across students who are male and students who are female.

- **Dependent samples t-test** *(Also called: Paired Samples t-test)*
  - Compare data from same students across two time points (e.g., pre-/post-tests).
  - The measure that is being compared across the groups must be a continuous variable.
  - Compare student GPAs before and after participating in a mentoring program.
## Appendix B. Statistics Cheat Sheet

### Table 2. Inferential Statistics (Tests Comparing Group Values)—Continued

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Purpose</th>
<th>Application/Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analysis of Variance (ANOVA)</strong>&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Compares means or proportions across <strong>two or more groups</strong>. For use with one categorical variable (with two categories for comparison) and one continuous variable. The “p-value” is used to determine whether the groups are significantly different from one another. A p-value of less than 0.05 indicates that we are 95% confident that if we selected other groups, then we would still see group differences in average values.</td>
<td></td>
</tr>
</tbody>
</table>
| **One-way ANOVA**<sup>1</sup>  | Compare two or more separate, non-overlapping groups or datasets. This is similar to independent samples t-test but can be conducted with more than 2 groups. The measure that is being compared across the groups must be a continuous variable. | • GPA across Freshmen, Sophomores, Juniors, and Seniors.  
• GPA across 3 different math placement groups.  
• SAT score across groups of students enrolling in 4 different semesters. |
| **Within Groups ANOVA**<sup>1</sup>  | Compare data from two or more related groups or same people over time. The measure that is being compared across the groups must be a continuous variable. This would be used for longitudinal data tracking groups across 2 or more time periods. | • GPA when the same student is followed over time. |
| **Factorial ANOVA**<sup>1</sup>  | Compare the interaction between two or more variables, factors, or dimensions (e.g., female AND freshman) on another variable. The measure that is being compared across the groups must be a continuous variable. Similar to a one-way ANOVA but incorporates one or more additional factor. | • GPA across Freshmen, Sophomores, Juniors, and Seniors who take different math placement courses. |
Table 2. Inferential Statistics (Tests Comparing Group Values)—Continued

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Purpose</th>
<th>Application/Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression(^1)</td>
<td>The type of regression described here is an “ordinary least-squares” regression.</td>
<td></td>
</tr>
</tbody>
</table>
| Ordinary least-squares regression | Tests whether there is a relationship between two variables under the framework that a student’s value on one variable tells us what value a student is likely to have on another variable.  
Allows for prediction of an outcome (or criterion) based on knowledge of some predictor variable (i.e., we can make a prediction of a value based on what we know about another factor).  
Though this test allows you to “predict” an outcome, it is not a causal test of whether one variable caused an outcome. This test simply examines whether there is a relationship between two variables.  
The outcome variable must be a continuous measure and cannot be a categorical measure.  
The predictor variable can be either a continuous measure or a category measure with 2 values (e.g. yes/no, true/false, Welding/Other POS) | • What factors (e.g. 1\(^{st}\) term GPA) predict student retention?  
• Is SAT score a predictor of 1\(^{st}\) term GPA? That is, is SAT score related to 1\(^{st}\) term GPA?  
• Is the number of courses passed related to the likelihood of finding full-time work? |
Table 2. Inferential Statistics (Tests Comparing Group Values)—Continued

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Purpose</th>
<th>Application/Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-parametric tests</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best used with very small samples. However, difficult to calculate in Excel because commands for these tests in the basic Excel package. Calculations must be done by hand with a complex set of commands, using advances statistical software, or with the use of user-created Excel add-on packages.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chi-Square</strong></td>
<td>Tests for the strength of the association between two categorical variables, each with two or more categories. This test compares the actual total of individuals in each category to the expected totals if there were no differences between individuals across groups. The analysis uses the size of the difference between the actual and expected totals to determine if the group allocations are significantly different from what is expected.</td>
<td>• Does the proportion of STEM vs non-STEM majors depend on the gender of the student? Is the proportion of females higher among STEM majors than it is among non-STEM majors? • Does the proportion of industry certificate attainers depend on the math placement course of the student? That is, is the proportion of industry certificate attainers higher among students who took specific placement courses?</td>
</tr>
<tr>
<td><strong>Mann-Whitney</strong></td>
<td>Non-parametric alternative to the independent sample t-test using one categorical variable (with two categories) and one continuous variable. Instead of comparing the mean values, this test compares the mean rank of values between two independent groups. The process involves ordering individuals by value, ranking individuals by value, and finally comparing the mean rank of all individuals in each group. Best used with very small samples. However, it is difficult to calculate in Excel because commands for these tests in the basic Excel package. Calculations must be done by hand with a complex set of commands, using advances statistical software, or with the use of user-created Excel add-on packages.</td>
<td>• -</td>
</tr>
</tbody>
</table>
### Appendix B. Statistics Cheat Sheet

#### Table 2. Inferential Statistics (Tests Comparing Group Values)—Continued

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Purpose</th>
<th>Application/Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wilcoxon signed-ranks</strong></td>
<td>Non-parametric alternative to the dependent sample t-test using one categorical variable (with two categories) and one continuous variable. Instead of comparing the mean values, this test compares the mean rank of values for one group at two points in time. The analysis involves ordering individuals by value, ranking individuals by value, and finally examining the difference between the first rank and the second rank for each individual. Best used with very small samples. However, it is difficult to calculate in Excel because commands for these tests in the basic Excel package. Calculations must be done by hand with a complex set of commands, using advances statistical software, or with the use of user-created Excel add-on packages.</td>
<td>• -</td>
</tr>
<tr>
<td><strong>Kruskal-Willis</strong></td>
<td>Non-parametric alternative to the ANOVA using one categorical variable (with two or more categories) and one continuous variable. Instead of comparing the mean values, this test compares the mean rank of values between two or more independent groups. The analysis involves ordering individuals by value, ranking individuals by value, and finally comparing the mean rank of individuals in each group. Best used with very small samples. However, it is difficult to calculate in Excel because commands for these tests in the basic Excel package. Calculations must be done by hand with a complex set of commands, using advances statistical software, or with the use of user-created Excel add-on packages.</td>
<td>• -</td>
</tr>
</tbody>
</table>
Using Visuals With Data

Why use visual aids with numeric data? Visual aids can help you understand your data and can help your readers understand your findings.

When choosing a visual aid, consider the following:

1) What type(s) of variables would you like to display? Are they categorical or continuous?
2) How many variables would you like to display? Is it possible to use the same visual aid to chart all variables of interest? Do you need to create a visual aid that incorporates all variables of interest?
3) Do the axes show the full range of values for your variable(s)? Sometimes Excel creates visuals that restrict the range of values more than what would make sense for your data. For example, a visual aid displaying GPA should generally have a range that goes up to 4.0, not up to 3.0 or 10.0, because most GPA's are between 0 and 4.
4) Does the resulting visual aid accurately display the key takeaways from your analysis?

Example 1. Box and Whisker Plot (Also Called a Box Plot)

Why use it: The box plot shows the characteristics of a single continuous variable ("Dependent variable"). The vertical axis shows the scale of the variable (e.g., for GPA the scale might go from 0 to 4; for percentage of courses passed the scale might go from 0% to 100%). There are several points for interpretation about the range and the quartiles of the variable. First, the minimum and maximum values are noted by the end of the whiskers, or the end of the lines above and below the box. Second, the line through the box is the median, which indicates the 50th percentile. Third, 50% of individuals fall into the values covered by the box. Fourth, 25% of individuals have values between the value of the 1st quartile (bottom of the box) and the median (line through the box) and 25% of individuals have values between the median (line through the box) and the value of the 3rd quartile (top of the box). Some box and whisker plots also show an “x” to mark the mean, or average, value.
Example 2. Bar Graph (Also Called a Bar Chart)

One variable:

![Bar Graph Example](image)

Two variables:

![Bar Graph Example](image)

**Why use it:** The bar graph shows the characteristics of a single categorical variable, or the relationship between a categorical variable and any other type of variable. Bar graphs can have up to 2 axes, allowing you to display information on 2 variables with different scales (e.g. you could include average GPA and average financial aid on the same graph, but each with its own axis). The interpretation of a bar graph will depend on how it is set up. To interpret the bar graph, compare the heights of bars to one another and consider the heights of the bars relative to the range of the variable(s).
Example 3. Scatter Plot

Why use it: The scatter plot shows the relationship between two continuous variables. Scatter plots can also incorporate information on a third categorical variable by using different colored dots for each category or group. Each dot represents one individual in the dataset. To interpret a scatter plot, look at where the dots cluster and how the placement of dots on the plot change as you move from left to right. Excel also allows the addition of a “line of best fit,” or a line that indicates the trend of dots from left to right. The incline of the line indicates whether the two variables are related (a steeper incline indicates a stronger relationship) and, if so, whether they increase with each other or if one increases as the other decreases.
Unlocking the Meaning of Student Data
Technical Assistance Webinar for the
College of Menominee Nation
Natassia Rodriguez Ott, PhD

What can data do for you?

Only 0.5% of data collected is ever used
What can data do for you?

- Show how well students are preforming
- Identify which students might need additional support
- Understand how well a program or curriculum is working
- Reveal patterns across students with different characteristics
- Verify that our perceptions are true (or false)

Choosing and transforming data

- Determine what data is needed to answer the question
- Define how the concepts will be measured
- Select the sample for analysis
- Transform the data

Think of the process like a maze. Directional choices along the way affect where you end up. Some paths lead to dead ends, while others lead to an exit.
Choosing and transforming data

- **Determine what data is needed to answer the question**
- Define how the concepts will be measured
- Select the sample for analysis
- Transform the data

<table>
<thead>
<tr>
<th>What information do you need to answer this question?</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many pieces of information do you need?</td>
</tr>
<tr>
<td>Do you need student-level data?</td>
</tr>
<tr>
<td>What time period(s) of data do you need?</td>
</tr>
</tbody>
</table>

**Is a welding student’s math placement course related to subsequent success in welding courses among students who take welding courses in 2014-2017?**

- **2 pieces of data: math placement course, student success**
- **Student-level data**
- **Data on students in Welding at CMN**
- **2014-2017**
Choosing and transforming data

- Determine what data is needed to answer the question
- **Define how the concepts will be measured**
  - Select the sample for analysis
  - Transform the data

Is a welding student’s math placement course related to subsequent success in welding courses among students who take welding courses in 2014-2017?
- “Math placement course” (SOAR, Foundation, College Ready)
- “Student success”
  - How many Welding courses did a student pass?
  - What percentage of the Welding courses taken did the student pass?
Choosing and transforming data

- Determine what data is needed to answer the question
- Define how the concepts will be measured
- **Select the sample for analysis**
- Transform the data

Is a welding student’s math placement course related to subsequent success in welding courses among students who take welding courses in 2014-2017?

- **Welding students in 2014-2017**
- **Students with math placement course data and grade data for at least one welding course**
- Other ideas: **Students who graduated**
Choosing and transforming data

- Determine what data is needed to answer the question
- Define how the concepts will be measured
- Select the sample for analysis
- **Transform the data**

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will you manipulate the data to match your concrete definitions?</td>
<td></td>
</tr>
<tr>
<td>Will you make categories or groups?</td>
<td></td>
</tr>
<tr>
<td>Will you combine existing categories?</td>
<td></td>
</tr>
<tr>
<td>How many categories will you compare?</td>
<td></td>
</tr>
</tbody>
</table>

Is a welding student’s math placement course related to subsequent success in welding courses among students who take welding courses in 2014-2017?

- Math placement course: Foundation, SOAR, College Ready
- **How many Welding courses did a student pass?**
  - Count the number of courses passed
- **What percentage of the Welding courses taken did the student pass?**
  - Count the number of courses taken, count the number of courses passed, and then divide the number of courses passed by the number of courses taken
Selecting statistical tests

- Consider what types of variables are involved: categorical vs continuous
- If there is a comparison involved, start by describing one group at a time before comparing the two groups.

- Resources:
  - Statistics Cheat Sheet, developed by RTI International
  - “There’s a Stat for that! What to do and When to do it”, by Bruce B. Frey

Is welding student’s math placement course related to subsequent success in welding courses among students who take welding courses in 2014-2017?

- Count the number of students who took each placement course
- Average and median success for all students and by group
- Range of success (minimum and maximum) for all students and by group
- ANOVA test for the significance of the differences in success across placement courses (i.e. are we likely to find this pattern in other groups of welding students?)
Interpreting Analysis

Understand the data
- Which numbers are highest? Lowest?
- Do the numbers seem higher or lower than what you expected?
- How do the numbers compare to each other for different groups?
- Are the between-group patterns what you expected?
- What visualizations can be used to display the data?

Put the data into context
- Do the data present any problematic inequalities?
- What do these patterns mean for these students? For this program? For this course sequence?
- What types of program changes could be made based on this data?
- What other information is needed to understand the pattern? To understand possible interventions?
Interpreting Analysis

Barriers to Returning to College
How likely is it that the following issues would cause you to withdraw from class or from this college?

- Working full-time: 43%
- Caring for dependents: 36%
- Academically unprepared: 18%
- Lack of finances: 53%

Source: Center for Community College Student Engagement 2014 CMN Report
Interpreting Analysis

Aspects of Highest Student Engagement

Source: Center for Community College Student Engagement 2014 CMN Report

Resources

- Excel Data Analysis Instructions Workbook
  - General instructions
  - Research question tips
  - Data analysis steps

- Excel Data Analysis Examples Workbook
  - 6 example questions
    - Today's examples came from Q1
  - Step by step instructions on approaching data analysis
  - Output and interpretation from analysis

- Statistics Cheat Sheet
  - Overview of statistical tests with use cases and limitations
  - Example output and interpretation
Other Resources

- Stanford Data Lab recommended tools for data analysis and data visualization

- For quick data on higher education institutions, try the online tools provided by the National Center for Education Statistics