IBM SPSS Statistics 23
Part 2: Test of Significance
Winter 2016, Version 1

Table of Contents

Introduction ...................................................................................................................................... 2
  Downloading the Data Files........................................................................................................... 2
Null Hypothesis .................................................................................................................................. 2
Statistical Tests .................................................................................................................................. 3
  Tests of Significance ........................................................................................................................... 3
    Correlations ............................................................................................................................................ 3
    Paired-Samples T Test ............................................................................................................................ 4
    Independent-Samples T Test .................................................................................................................. 5
Multiple Response Sets ...................................................................................................................... 7
  Multiple Response Frequencies ............................................................................................................. 8
  Multiple Response Crosstabs ............................................................................................................... 9
Data Manipulation ........................................................................................................................... 11
  Copying and Pasting Variable Properties ............................................................................................. 12
  Inserting Variables and Cases .............................................................................................................. 13
  Deleting Variables and Cases ............................................................................................................ 14
Merging Data Files ........................................................................................................................... 14
  Creating the Data Files for Merging ..................................................................................................... 14
  Merging the Data Files ....................................................................................................................... 16
Appendix ......................................................................................................................................... 19

For additional training resources, visit www.calstatela.edu/training.
Introduction

SPSS stands for Statistical Package for the Social Sciences. This program can be used to analyze data collected from surveys, tests, observations, etc. It can perform a variety of data analyses and presentation functions, including statistical analysis and graphical presentation of data. Among its features are modules for statistical data analysis. These include (1) descriptive statistics such as frequencies, central tendency, plots, charts, and lists; and (2) sophisticated inferential and multivariate statistical procedures such as analysis of variance (ANOVA), factor analysis, cluster analysis, and categorical data analysis. IBM SPSS Statistics 23 is well-suited for survey research, though by no means is it limited to just this topic of exploration.

This handout (1) introduces several data entry and data manipulation techniques that help you save time, (2) covers the basic skills necessary to perform tests of significance such as correlations and t tests, and (3) provides an introduction to multiple response sets. The step-by-step instructions will help you understand how to interpret the output of your tests from data supplied by your research question(s). Please follow the steps carefully to get appropriate results. Please note that a slightly different process might yield unexpected and complicated results.

Downloading the Data Files

This handout includes sample data files that can be used to follow along the steps. If you plan to use the data files, download the following ZIP file to your computer and extract the files. It is recommended to save the data files on your desktop for easy access.

- IBM SPSS Statistics 23 Part 2 Data Files

Null Hypothesis

The null hypothesis \( (H_0) \) represents a theory that has been presented, either because it is believed to be true or because it is to be used as a basis for an argument. It is a statement that has not been proven. It is also important to realize that the null hypothesis is the statement of no difference. For example, in a clinical trial of a new drug, the null hypothesis might state that the new drug is no better, on average, than the current drug (in other words, the new drug exhibits the same behavior as the old drug). The null hypothesis \( (H_0) \) and the alternative hypothesis \( (H_1) \) can be stated as:

\[
H_0: \text{There is no difference between the two drugs.} \\
H_1: \text{There is a significant difference between the two drugs.}
\]

Special consideration is given to the null hypothesis. This is due to the fact that the null hypothesis relates to the statement being tested, whereas the alternative hypothesis relates to the statement to be accepted if and when the null hypothesis is rejected.

After testing is complete, the final conclusion is given in terms of the null hypothesis. The result is either "Reject \( H_0 \) in favor of \( H_1 \)" or "Do not reject \( H_0 \)"; the conclusion is never "Reject \( H_1 \)" or "Accept \( H_1 \)." If the conclusion is "Do not reject \( H_0 \)," this does not necessarily mean that the null hypothesis is true. It only suggests that there is no sufficient evidence against \( H_0 \) in favor of \( H_1 \). Rejecting the null hypothesis then suggests that the alternative hypothesis may be true.

NOTE: The null hypothesis essentially states that the given cases or items under consideration are statistically the same or exhibit the same behavior without any significant difference. The alternative hypothesis states that the given cases exhibit different behavior or that they have a statistically significant difference.
Statistical Tests

Statistics is a set of mathematical techniques used to summarize research data and determine whether the data supports a proposed hypothesis. SPSS Statistics includes tools that can be used to analyze variables and determine the strength and nature of the relationship between two variables, and whether the means (averages) of two data sets (samples) are statistically the same or different.

Tests of Significance

The following examples are sample research questions that can be answered using SPSS Statistics’ analytical methods.

Correlations

A correlation is a statistical device that measures the strength or degree of a supposed linear association between two or more variables. One of the more common measures used is the Pearson correlation which estimates a relationship between two interval variables.

Research Question # 1

Is there a relationship between academic performance and Internet access?

To run a correlation analysis:

1. Start IBM SPSS Statistics 23, and then open the Part2.sav file.
2. Click the Analyze menu, point to Correlate, and then click Bivariate.
3. In the Bivariate Correlations dialog box, select the posttest, gpa, and active variables in the left box, and then click the transfer arrow button to move them to the Variables box (see Figure 1).

   NOTE: You can select multiple variables by clicking the first variable, holding down the Ctrl key, and then clicking each of the other variables.

4. Make sure that the Pearson check box and the Two-tailed option button are selected.

   Figure 1 – Bivariate Correlations Dialog Box
5. Click the OK button. The Output Viewer window opens with a Correlations table (see Figure 2).

NOTE: Pay attention to the values with asterisks. They indicate statistical significance.

<table>
<thead>
<tr>
<th></th>
<th>posttest</th>
<th>gpa</th>
<th>active</th>
</tr>
</thead>
<tbody>
<tr>
<td>posttest</td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.383*</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>gpa</td>
<td>Pearson Correlation</td>
<td>.398*</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>active</td>
<td>Pearson Correlation</td>
<td>.514**</td>
<td>.502**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
</tbody>
</table>

*: Correlation is significant at the 0.05 level (2-tailed).
**: Correlation is significant at the 0.01 level (2-tailed).

Pay attention to the values with asterisks. They indicate statistical significance.

There are moderate and positive relationships between (1) the frequency of internet access and posttest scores, and (2) the frequency of internet access and GPA.

Figure 2 – Bivariate Correlations Output

Answer to Research Question # 1
Is there a relationship between academic performance and Internet access?

Answer: Yes

Explanation: As shown in the Correlations table above, the correlation index for the relationship between active and posttest is 0.514, which is between 0.4-0.7. The correlation index for the relationship between active and gpa is 0.502, which is between 0.4-0.7. The results from these analyses indicate that there is a moderate, positive relationship between academic performance and Internet access.

Paired-Samples T Test

For any type of t test, the following conditions must be met: the data (1) has normal distribution, (2) is a large data set, and (3) has no outliers. If any of these conditions is not met, then a nonparametric test should be used. Researchers use different types of t tests depending on the design of their study. One such t test is a Paired-Samples T Test which is used to test if an observed difference between the two means of a paired samples set is statistically significant.

Research Question # 2
Does Internet access in the computer class have any instructional effect on student academic achievement?

H₀: There is no instructional effect on academic achievement from having access to the Internet in the computer class.

H₁: There is an instructional effect on academic achievement from having access to the Internet in the computer class.

The null hypothesis states that there is no instructional effect influencing the academic achievement from having access to the Internet in the computer class. The variables that reflect academic achievement are pretest and posttest.
To run a Paired-Samples T Test:
1. Click the Analyze menu, point to Compare Means, and then click Paired-Samples T Test.
2. In the Paired-Samples T Test dialog box, select the pretest and posttest variables in the left box, and then click the transfer arrow button to move them to the Paired Variables box (see Figure 3).
3. Click the OK button. The Output Viewer window opens with several tables, including a Paired Samples Test table (see Figure 4).

**Figure 3 – Paired-Samples T Test Dialog Box**

**Paired Samples Test**

<table>
<thead>
<tr>
<th>Paired Variables:</th>
<th>Pair</th>
<th>Variable1</th>
<th>Variable2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>[pretest]</td>
<td>[posttest]</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4 – Paired-Samples T Test Output**

**Answer to Research Question # 2**
Does Internet access in the computer class have any instructional effect on student academic achievement?

**Answer:** Yes

**Explanation:** The observed mean difference is -4.51724. Since the value of t is -3.820 at p < .001, the mean difference (-4.51724) between pretest and posttest is statistically significant. According to the Sig. of 0.001 (which is less than 0.05), the hypothesis is rejected. Therefore, it can be inferred that there was an instructional effect on academic achievement from having access to the Internet in the computer class.

**Independent-Samples T Test**

An Independent-Samples T Test is used to determine the likelihood that two independent data samples came from populations with identical means. If this were true, then the difference between the means should equal zero. In this case, the null hypothesis would indicate that the two means are equal.
Two variables are required in the data set. One variable is the measured parameter. Examples include weight, height, or frequency. The second variable divides the data set into two groups. The means of the light and dark groups will be compared.

**Research Question # 3**

*Is there a difference in the average number of seedlings grown in the light and those grown in the dark?*

In this example, 20 Petri dishes, each with 10 celery seeds, were observed for one week. Ten of the dishes were kept in the dark; the other 10 were placed under a grow light. At the end of the week, observers counted the number of sprouted seeds in each dish.

**$H_0$:** variance (light) = variance (dark)

**$H_1$:** variance (light) ≠ variance (dark)

**$H_0$:** There is no difference between seedlings under the light and in the dark ($\mu_{\text{light}} = \mu_{\text{dark}}$).

**$H_1$:** There is sig. difference between seedlings under the light and in the dark ($\mu_{\text{light}} \neq \mu_{\text{dark}}$).

**NOTE:** The first set of hypotheses is testing the variance while the proceeding set is testing the mean. Variance, the arithmetic mean of the squared deviations from the mean, indicates how far the single samples are from the mean. The variances have to be equal before we can determine if the means are equal. If the variances are equal, the T Test can be used. If the variances are not equal, more testing is required.

**To run an Independent-Samples T Test:**

1. Open the Seedlings.sav file.
2. In Data View, click the Analyze menu, point to Compare Means, and then click Independent-Samples T Test.
3. In the Independent-Samples T Test dialog box, select the Seedlings variable in the left box, and then click the transfer arrow button to move it to the Test Variable(s) box (see Figure 5).
4. Select the Treatment variable in the left box, and then click the transfer arrow button to move it to the Grouping Variable box.
5. Click the Define Groups button.

![Figure 5 – Independent-Samples T Test Dialog Box](image)
6. In the Define Groups dialog box, type 0 in the Group 1 box, type 1 in the Group 2 box, and then click the Continue button (see Figure 6).
7. In the Independent-Samples T Test dialog box, click the OK button. The Output Viewer window opens with several tables, including an Independent Samples Test table (see Figure 7).

![Figure 6 – Define Groups Dialog Box](image)

<table>
<thead>
<tr>
<th>Independent Samples Test</th>
<th>Levene’s Test for Equality of Variances</th>
<th>Test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig</td>
</tr>
<tr>
<td>Seedlings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equivariances assumed</td>
<td>1.998</td>
<td>.194</td>
</tr>
<tr>
<td>Equivariances not</td>
<td>-3.179</td>
<td>13.834</td>
</tr>
</tbody>
</table>

![Figure 7 – Independent-Samples T Test Output](image)

**Answer to Research Question # 3**

Is there a difference in the average number of seedlings grown in the light and those grown in the dark?

**Answer:** Yes

**Explanation:** The mean difference in seedlings sprouted between the two treatments (light and dark) was -2.900. The value of t, which is -3.179, was statistically significant (p=0.005). Therefore, the null hypothesis is rejected.

**Multiple Response Sets**

Surveys often contain questions that allow respondents to select more than one answer. Managing such questions in SPSS Statistics can produce some difficulty. Each response in a multiple response question should be coded as a separate variable and then grouped under a multiple response set of variables. The multiple response set can then be analyzed using frequency counts or crosstabs.

**To define a multiple response set of variables:**

1. Open the Airlines.sav file.
2. In Data View, click the Analyze menu, point to Multiple Response, and then click Define Variable Sets.
3. In the Define Multiple Response Sets dialog box, select the American, Delta, United, Southwest and Other airline variables and move them to the Variables in Set box (see Figure 8).
4. Make sure that the Dichotomies option button is selected, and then type 1 in the Counted value box.
5. Type **Airlines** in the **Name** box.
6. Type **Airline frequency of response** in the **Label** box.
7. Click the **Add** button. The set is created as $Airlines and listed in the **Multiple Response Sets** box.
8. Click the **Close** button.

![Figure 8 – Define Multiple Response Sets Dialog Box](image)

**Multiple Response Frequencies**

The next research questions can be answered by running a frequency analysis for each airline variable. The result of this analysis will provide an overall raw frequency for each response and will not allow percentage comparisons between the different airlines. A frequency analysis using a multiple response set will provide an appropriate response with a concise output.

**Research Question # 4**

In a survey of airline passengers, which airline was selected as having been flown most often in the previous six months?

To analyze the frequency of response for each variable in a multiple response set:

1. Click the **Analyze** menu, point to **Multiple Response**, and then click **Frequencies**.
2. In the **Multiple Response Frequencies** dialog box, select the multiple response set labeled $Airlines and move it to the **Table(s) for** box (see Figure 9).
3. Click the **OK** button. The **Output Viewer** window opens with the frequency analysis (see Figure 10).
In a survey of airline passengers, which airline was selected as having been flown most often in the previous six months?

**Answer:** United

**Explanation:** As shown in the Output Viewer window, there were 18 people surveyed and 44 total responses generated. Of the 44 total responses, **United** was selected most often with 12 responses (representing 27.3% – the largest portion of the total responses).

**Multiple Response Crosstabs**

Without using a multiple response set, each airline would have to be analyzed against the variable passengers used to identify themselves as being afraid of flying. This would require the use of a crosstab analysis. However, the overall results would not allow for easy comparison between each of the airlines. To best answer the question, include the multiple response set in a crosstab analysis.
Research Question # 5
In a survey of airline passengers, which airline was selected most often by those passengers who identified themselves as afraid to fly?

To incorporate a multiple response set into a crosstab analysis:
1. Click the **Analyze** menu, point to **Multiple Response**, and then click **Crosstabs**.
2. In the **Multiple Response Crosstabs** dialog box, select the **FearFactor** variable and move it to the **Row(s)** box (see Figure 11).
3. Select the **$Airlines** multiple response set and move it to the **Column(s)** box.
4. Select the **FearFactor** variable in the **Row(s)** box. The **Define Ranges** button becomes active.
5. Click the **Define Ranges** button.

6. In the **Multiple Response Crosstabs: Define Variable Ranges** dialog box, type 0 in the **Minimum** box and 1 in the **Maximum** box for the **FearFactor** variable, and then click the **Continue** button (see Figure 12).

7. In the **Multiple Response Crosstabs** dialog box, click the **Options** button.
8. In the **Multiple Response Crosstabs: Options** dialog box, in the **Percentages Based on** section, select the **Cases** option button, and then click the **Continue** button (see Figure 13).

![Multiple Response Crosstabs: Options Dialog Box](image)

**Figure 13 – Multiple Response Crosstabs: Options Dialog Box**

9. In the **Multiple Response Crosstabs** dialog box, click the **OK** button. The **Output Viewer** window opens with the crosstab results (see Figure 14).

![FearFactor * Airlines Crosstabulation](image)

**Table:**

<table>
<thead>
<tr>
<th>FearFactor * Airlines Crosstabulation</th>
<th>Airlines frequency of response</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>American</td>
<td>Delta</td>
</tr>
<tr>
<td>Not Afraid to Fly</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Afraid to Fly</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8</strong></td>
<td><strong>8</strong></td>
</tr>
</tbody>
</table>

Percentages and totals are based on respondents.

a. Dichotomy group tabulated at value 1.

**Figure 14 – Multiple Response Crosstabs Output**

**Answer to Research Question # 5**

In a survey of airline passengers, which airline was selected most often by those passengers who identified themselves as afraid to fly?

**Answer:** Southwest

**Explanation:** Of the 18 people surveyed, 10 identified themselves as being afraid to fly. Within that group of survey respondents, **Southwest** was the airline selected most often (7 times).

**Data Manipulation**

SPSS Statistics provides tools to simplify data manipulation which save users time and effort. Users can copy and paste variables as well as insert or delete cases and variables.
Copying and Pasting Variable Properties

Copying and pasting is very useful when the same properties are assigned to multiple variables.

To copy and paste variable properties:

1. Click the File menu, point to New, and then click Data.
2. Click the Variable View tab in the lower-left corner of the Data Editor window (see Figure 15).

3. Type active in the first cell under the Name column, and then press the Enter key.
4. Click in the first cell under the Decimals column, and then change the entry to 0 using the spin box.
5. Click in the first cell under the Values column, and then click the Ellipses button that appears in the cell.
6. In the Value Labels dialog box, type 1 in the Value box, type Strongly Disagree in the Label box, and then click the Add button.
7. Assign 2, 3, 4, and 5 for Disagree, Neutral, Agree, and Strongly Agree, respectively, by repeating step 6 for each value added. See Figure 16 for the results.
   
   NOTE: This is an example of a typical 5-point Likert scale which is a common psychometric rating scale used in questionnaires.

8. Click the OK button.

9. Click the row 1 header to select the entire row (see Figure 17).
10. Click the Edit menu, and then click Copy to copy the properties of the active variable.
11. Select the rows you want to apply the same properties to by dragging across the row headers (see Figure 18).

![Figure 17 – Selecting a Variable](image1)

![Figure 18 – Selecting Target Variables](image2)

12. Click the **Edit** menu, and then click **Paste**. The copied properties of the **active** variable are applied to the target variables, and the **Variable View** and **Data View** change (see Figure 19 and Figure 20).

![Figure 19 – Variable View Showing the New Variables](image3)

![Figure 20 – Data View Showing the New Variables](image4)

## Inserting Variables and Cases

By using **Insert Variable** and **Insert Cases**, variables and cases can be added to any location of the data file in a simple, straightforward manner. Assume that one wants to insert a new variable named **midterm** between **pretest** and **posttest** and use it for test score data. The following instructions describe how to insert a new variable and assign it the **Numeric** data type.

**To insert a variable:**

1. Open the **Part2.sav** file.
2. Switch to **Data View**.
3. Click the **posttest** variable header to select the column.
4. Click the **Edit** menu, and then click **Insert Variable**. A new variable is inserted to the left of the selected variable (**posttest**).

   **NOTE:** The new variable is created with a default name **VAR00001** which can be changed later.

5. To define the properties of the new variable, double-click the variable header. **Variable View** is activated for the new variable.
6. Type midterm in the Name column of the new variable, and then press the Enter key.
7. Change the variable type if desired.

Likewise, it is possible to insert cases in a particular location in Data View. For instance, assume that a particular student’s record should be inserted between case 10 and 11. By following the instructions below, one case will be inserted after the 10th case.

To insert a case:
1. Switch to Data View.
2. Click the row 11 header to select the case.
3. Click the Edit menu, and then click Insert Cases. A new case is inserted above case 11.

Deleting Variables and Cases
Variables and cases can be deleted by using the Clear command.

To delete a variable or a case:
1. In Data View, click the variable header or case number to select it.
2. Click the Edit menu, and then click Clear. The selected variable or case is deleted.

Merging Data Files
The Merge Files function is useful when records are stored in separate files and need to be combined later. This allows users to import data from one file into another as long as both sets of data (from each file) contain a common identifier for each of the cases that the user wishes to combine.

An identifier has no meaning other than to distinguish each case from one another and to identify the correlating cases from the additional data files. This identifier can be a unique value, number, or letter combination to be applied to each case.

NOTE: The variables do not have to be the same across data files.

Creating the Data Files for Merging
Scenario: A campus research team needs to create a file for a longitudinal study of 10 students. Each file will have the same students, but four different focal points of study pertaining to each question. Over the study’s 5 year span, the 10 students will be asked 12 questions each year (1 a month) and the same questions will be asked each year. At the end of the year, the three files will be combined in an annual questionnaire file to be analyzed. The Merge Files function can be used to satisfy this requirement.

The files must be created before they can be merged.

To create a data file for merging:
1. Click the File menu, point to New, and then click Data to create a new file.
2. Switch to Variable View.
3. Click in the first variable cell, type ID (to be used as the identifier variable), and then press the Enter key.
4. Click in the Type column, and then click the Ellipses button.
5. In the Variable Type dialog box, select the String option button, type 30 in the Characters box, and then click the OK button.
6. Click in the second variable cell, type **January**, and then press the **Enter** key.
7. Change the **Type** attribute to **String** and the **Width** attribute to **30**.
8. Type **What pet would you like to own?** under the **Label** column, and then press the **Enter** key.
9. Repeat steps 6 through 8 to enter the data in Table 1. See Figure 21 for the results.

### Table 1 – Variables for Case Study

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Width</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>February</td>
<td>String</td>
<td>30</td>
<td>What is your favorite shape?</td>
</tr>
<tr>
<td>March</td>
<td>String</td>
<td>36</td>
<td>It is 12:30 p.m., what are you eating?</td>
</tr>
<tr>
<td>April</td>
<td>String</td>
<td>36</td>
<td>What is your preferred beverage?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Width</th>
<th>Decimals</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ID</td>
<td>30</td>
<td>0</td>
<td>What pet would you like to own?</td>
</tr>
<tr>
<td>2</td>
<td>January</td>
<td>30</td>
<td>0</td>
<td>What is your favorite shape?</td>
</tr>
<tr>
<td>3</td>
<td>February</td>
<td>30</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>March</td>
<td>36</td>
<td>0</td>
<td>It is 12:30 p.m., what are you eating?</td>
</tr>
<tr>
<td>5</td>
<td>April</td>
<td>36</td>
<td>0</td>
<td>What is your preferred beverage?</td>
</tr>
</tbody>
</table>

### Figure 21 – Defining Variables in Variable View

10. Once the variables have been defined in **Variable View**, switch to **Data View** to enter the corresponding case data.
11. Enter **Alfred** in case 1 of the **ID** variable, **Bethel** in case 2 of the **ID** variable, and so on. Enter the corresponding data according to Table 2. See Figure 22 for the results.

### Table 2 – Case Data

<table>
<thead>
<tr>
<th>Case</th>
<th>ID</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alfred</td>
<td>Dog</td>
<td>Star</td>
<td>Pizza</td>
<td>Water</td>
</tr>
<tr>
<td>2</td>
<td>Bethel</td>
<td>Cat</td>
<td>Square</td>
<td>Fruit</td>
<td>Soda Pop</td>
</tr>
<tr>
<td>3</td>
<td>Chris</td>
<td>Cat</td>
<td>Triangle</td>
<td>Veggies</td>
<td>Grape Juice</td>
</tr>
<tr>
<td>4</td>
<td>Dante</td>
<td>Dog</td>
<td>Rectangle</td>
<td>Sandwich</td>
<td>Orange Juice</td>
</tr>
<tr>
<td>5</td>
<td>Erica</td>
<td>Tiger</td>
<td>Oval</td>
<td>Chips</td>
<td>Aloe Water</td>
</tr>
<tr>
<td>6</td>
<td>Fernando</td>
<td>Tarantula</td>
<td>Circle</td>
<td>Calzone</td>
<td>Beer</td>
</tr>
<tr>
<td>7</td>
<td>Grenadine</td>
<td>Dog</td>
<td>Octagon</td>
<td>Salad</td>
<td>White Wine</td>
</tr>
<tr>
<td>8</td>
<td>Harold</td>
<td>Bees</td>
<td>Polygon</td>
<td>Soup</td>
<td>Naked Juices</td>
</tr>
<tr>
<td>9</td>
<td>Isadora</td>
<td>Turtle</td>
<td>Rhombus</td>
<td>Panda Express</td>
<td>V8 Juice</td>
</tr>
<tr>
<td>10</td>
<td>Jessie</td>
<td>Hamster</td>
<td>Oval</td>
<td>Egg Salad</td>
<td>Lemonade</td>
</tr>
</tbody>
</table>
To save the file, click the **File** menu, and then click **Save**.

In the **Save Data As** dialog box, select the **Desktop** as the destination, type **Merge1** in the **File name** box, and then click the **Save** button.

Close the **Output Viewer** window.

### Merging the Data Files

To merge data files, all files must have a common variable. The common variable in this case is ID.

**To merge data files:**

1. Open the **Merge2.sav** and **Merge3.sav** files and compare them to the **Merge1.sav** file to make sure that all three files have the same IDs.
2. Close the **Merge2.sav** and **Merge3.sav** files.
3. With the **Merge1.sav** file displayed, click the **Data** menu, point to **Merge Files**, and then click **Add Variables**.
4. In the **Add Variables to Merge1.sav** dialog box, select the **An external SPSS Statistics data file** option button, and then click the **Browse** button (see Figure 23).

5. In the **Add Variables: Read File** dialog box, navigate to the location where you saved the data files, select the **Merge2.sav** file, and then click the **Open** button (see Figure 24).
6. In the Add Variables to Merge1.sav dialog box, click the Continue button.
7. In the Add Variables from Merge2.sav dialog box, select the Match cases on key variables check box, the Cases are sorted in order of key variables in both datasets check box, and the Both files provide cases option (see Figure 25).
8. Select the ID>(+) variable in the Excluded Variables box and move it to the Key Variables box.
9. Click the OK button.
10. A warning message dialog box opens. Click the OK button (see Figure 26).

![Figure 26 – Warning Message Dialog Box](image)

11. Close the **Output Viewer** window. The results should look like Figure 27.

![Figure 27 – Merged Data](image)

12. Repeat steps 3 through 10 for the **Merge3.sav** file.

**NOTE**: SPSS also provides the option to select a file to merge from any currently open datasheets in the program. Make sure to save the file that is to be merged to ensure that the most recent data is inputted into the receiving file.
QUESTIONNAIRE

This survey is designed to investigate relationships between Internet access and academic success. It consists of three parts: questions related to the background information of the respondent, questions about Internet use patterns, and an open-ended question. Please select appropriate answers that best describe your activities on the Internet as truthfully as possible. The results of this study will be used anonymously for the IBM SPSS Statistics 23 Part 2: Test of Significance workshop.

Background Information

1. Age: ______________________________________________
2. Major: ____________________________________________
3. GPA: ______________________________________________
4. Monthly Income: _________________________________

Internet Access

5. Do you have a device capable of accessing the Internet at home?
   1. Yes  2. No
6. Where do you surf on the Internet? (You can circle more than one option for this question.)
   1. At school  2. At home  3. At work  4. Other ________________
7. How long do you stay online per day?
   1. Less than 1 hour  2. 2-3 hours  3. 3-4 hours  4. More than 5 hours

Questions 8 through 19 are designed to investigate the frequency and types of activities on the Internet. These questions have a 5-point Likert scale ranging from strongly disagree to strongly agree. Please circle the option that best describes your activities on the Internet.

1 = Strongly Disagree
2 = Disagree
3 = Neutral
4 = Agree
5 = Strongly Agree

8. I am a very active Internet surfer. 1 2 3 4 5
9. I surf the Internet to look for articles for research papers. 1 2 3 4 5
10. I surf the Internet to read current news. 1 2 3 4 5
11. I use the Internet only to email my friends, family, and professors. 1 2 3 4 5
12. I use the Internet to keep up with friends and family, or for social networking (e.g., Twitter, Facebook).  

13. I surf the Internet to look for business information (e.g., Yellow Pages, Yelp).  

14. I surf the Internet to look for job openings.  

15. I use the Internet to play games.  

16. I use the Internet to download forms and files (e.g., income tax forms).  

17. I surf the Internet to improve my computer skills.  

18. I surf the Internet to purchase books.  

19. I surf the Internet to purchase other merchandise (e.g., clothes, computers).  

Question 20 is an open-ended question.  

20. Are there any other Internet activities that are not included in this survey? If so, please describe them below.

_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________