IBM SPSS Statistics 20
Part 4: Chi-Square and ANOVA

Summer 2013, Version 2.0

Table of Contents

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 20 is well-suited for survey research, though by no means is it limited to just this topic of exploration.

This handout introduces basic skills for performing hypothesis tests utilizing Chi-Square test for Goodness-of-Fit and generalized pooled t tests, such as ANOVA. The step-by-step instructions will guide users in performing "tests of significance" using SPSS Statistics and will help users understand how to interpret the output for research questions.

Downloading the Data Files

This handout includes sample data files that can be used for hands-on practice. The data files are stored in a self-extracting archive. The archive must be downloaded and executed in order to extract the data files.

- The data files used with this handout are available for download at http://www.calstatela.edu/its/training/datafiles/spss20p4.exe.
- Instructions on how to download and extract the data files are available at http://www.calstatela.edu/its/training/pdf/download.pdf.

Chi-Square

The *Chi-Square ($\chi 2$)* test is a statistical tool used to examine differences between nominal or categorical variables. The Chi-Square test is used in two similar but distinct circumstances:

- To estimate how closely an observed distribution matches an expected distribution – also known as the Goodness-of-Fit test.
- To determine whether two random variables are independent.

Chi-Square Test for Goodness-of-Fit

This procedure can be used to perform a hypothesis test about the distribution of a qualitative (categorical) variable or a discrete quantitative variable having only finite possible values. It analyzes whether the observed frequency distribution of a categorical or nominal variable is consistent with the expected frequency distribution.

With Fixed Expected Values

Research Question # 1

Can the hospital schedule discharge support staff evenly throughout the week?

A large hospital schedules discharge support staff assuming that patients leave the hospital at a fairly constant rate throughout the week. However, because of increasing complaints of staff shortages, the hospital administration wants to determine whether the number of discharges varies by the day of the week.

H_0: Patients leave the hospital at a constant rate (there is no difference between the discharge rates for each day of the week).
H_1: Patients do not leave the hospital at a constant rate.

To perform the analysis:

1. Start **IBM SPSS Statistics 20**.
2. Click the **Open** button  on the **Data Editor** toolbar. The **Open Data** dialog box opens.
3. Navigate to the **Data Files** folder, select the **Chi-hospital.sav** file, and then click the **Open** button.

The observed values must be declared before running the Chi-Square test.

To declare the observed values:

1. Click the **Data** menu, and then click **Weight Cases**. The **Weight Cases** dialog box opens (see Figure 1).
2. Select the **Weight cases by** option.
3. Select the **Average Daily Discharges [discharge]** variable in the box on the left, and then click the transfer arrow button  to move it to the **Frequency Variable** box.
4. Click the **OK** button.

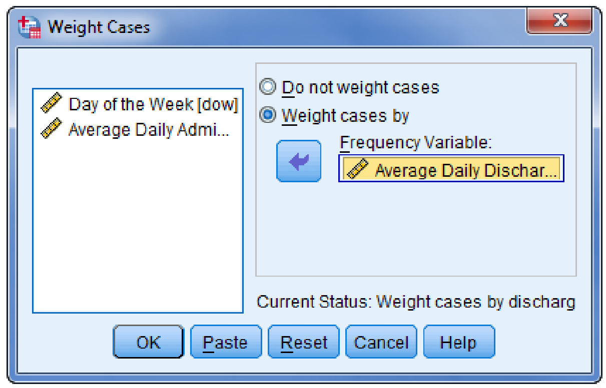

Figure 1 – Weight Cases Dialog Box

To perform the analysis:

1. Click the **Analyze** menu, point to **Nonparametric Tests**, point to **Legacy Dialogs**, and then click **Chi-square**. The **Chi-square Test** dialog box opens.
2. Select the **Day of the Week [dow]** variable and move it to the **Test Variable List** box (see Figure 2).
3. Click the **OK** button. The **Output Viewer** window opens (see Figure 3).

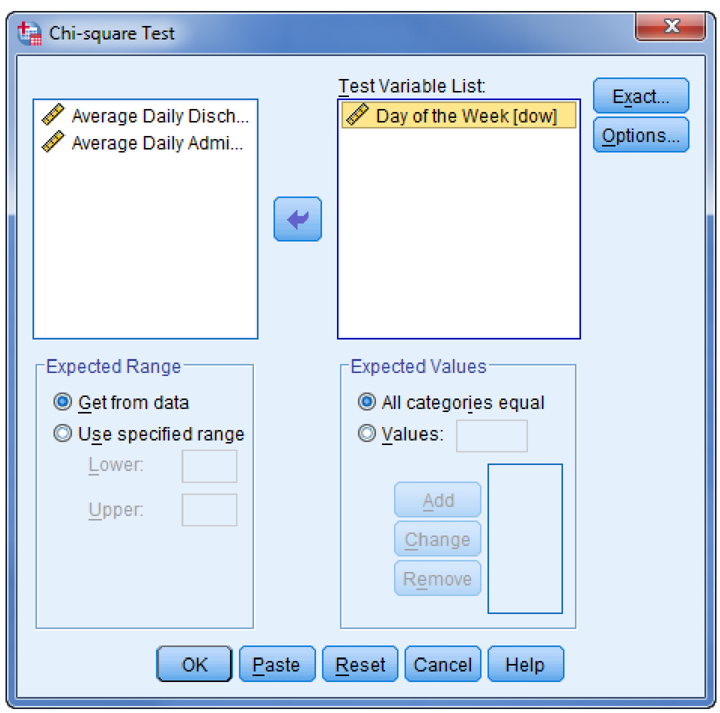

Figure 2 – Chi-square Test Dialog Box

Frequencies

Day of the Week

	Observed N	Expected N	Residual
Sunday	44	84.1	-40.1
Monday	78	84.1	-6.1
Tuesday	90	84.1	5.9
Wednesday	94	84.1	9.9
Thursday	89	84.1	4.9
Friday	110	84.1	25.9
Saturday	84	84.1	-.1
Total	589		

The observed frequency for each row is the actual number of patients discharged per day.

The expected value for each row is equal to the sum of the observed frequencies divided by the number of rows in the table.

The *Residual* is equal to the observed frequency minus the expected value.

Figure 3 – Chi-Square Frequencies Output Table

	Day of the Week
Chi-Square	29.389[a]
df	6
Asymp. Sig.	.000

Figure 4 – Chi-Square Test Statistics Output Table

<u>Reporting the analysis results</u>:
H_0: Rejected in favor of **H_1**.
H_1: Patients do not leave the hospital at a constant rate.

Explanation: Figure 4 indicates that the calculated χ^2 statistic, for 6 degrees of freedom, is 29.389. Additionally, it indicates that the significance value (0.000) is less than the usual threshold value of 0.05. This suggests that the null hypothesis, **H_0** (patients leave the hospital at a constant rate), can be rejected in favor of the alternate hypothesis, **H_1** (patients leave the hospital at different rates during the week).

With Fixed Expected Values and within a Contiguous Subset of Values

By default, the Chi-Square test procedure builds frequencies and calculates an expected value based on all valid values of the test variable in the data file. However, it may be desirable to restrict the test's range to a contiguous subset of the available values, such as weekdays only (Monday through Friday).

<u>Research Question # 2</u>

The hospital requests a follow-up analysis. Can staff be scheduled assuming that patients discharged on weekdays only (Monday through Friday) leave at a constant daily rate?

H_0: Patients discharged on weekdays only (Monday through Friday) leave at a constant daily rate.
H_1: Patients discharged on weekdays only (Monday through Friday) do not leave at a constant daily rate.

To run the analysis:

1. Click the **Analyze** menu, point to **Nonparametric Tests**, point to **Legacy Dialogs**, and then click **Chi-square**. The **Chi-square Test** dialog box opens.
2. Select the **Use specified range** option in the **Expected Range** section (see Figure 2).
3. Type **2** in the **Lower** box and **6** in the **Upper** box.
4. Click the **OK** button. The **Output Viewer** window opens (see Figure 5 and Figure 6). Notice that the test range is restricted to Monday through Friday.

	Day of the Week			
	Category	Observed N	Expected N	Residual
1	Monday	78	92.2	-14.2
2	Tuesday	90	92.2	-2.2
3	Wednesday	94	92.2	1.8
4	Thursday	89	92.2	-3.2
5	Friday	110	92.2	17.8
Total		461		

Figure 5 – Chi-Square (Subset) Frequencies Output Table

	Day of the Week
Chi-square	5.822[a]
df	4
Asymp. Sig.	.213

Figure 6 – Test Statistics Output Table

NOTE: The expected values are equal to the sum of the observed values divided by the number of rows, while the observed values are the actual number of patients discharged.

Reporting the analysis results:
H_0: Do not reject. Patients discharged on weekdays only (Monday through Friday) leave at a constant daily rate.

Explanation: Figure 5 indicates that, on average, 92 patients were discharged from the hospital each weekday. The rate for Mondays was below average and the rate for Fridays was above average. Figure 6 indicates that the calculated value of the Chi-Square statistic was 5.822 at 4 degrees of freedom. Because the significance level (0.213) is greater than the rejection threshold of 0.05, **H_0** (patients were discharged at a constant rate on weekdays) could not be rejected.

Using the Chi-Square test procedure, the hospital determined that the patient discharge rate was not constant over the course of an average week. This was primarily due to more discharges on Fridays and fewer discharges on Sundays. When the test's range was restricted to weekdays, the discharge rates appeared to be more uniform. Staff shortages could be corrected by adopting separate weekday and weekend staff schedules.

With Customized Expected Values

Research Question # 3
Does first-class mailing provide quicker response time than bulk mail?

A manufacturing company tries first-class postage for direct mailings, hoping for faster responses than with bulk mail. Order takers record how many weeks each order takes after mailing.

H_0: First-class and bulk mailings do not result in different customer response times.
H_1: First-class and bulk mailings result in different customer response times.

Before the Chi-Square test is run, the cases must be weighted. Because this example compares two different methods, one method must be selected to provide the expected values for the test and the other will provide the observed values.

To weight the cases:

1. Open the **Chi-mail.sav** file.
2. Click the **Data** menu, and then click **Weight Cases**. The **Weight Cases** dialog box opens.
3. Select the **Weight cases by** option.
4. Select the **First Class Mail [fcmail]** variable and move it to the **Frequency Variable** box.
5. Click the **OK** button.

To run the analysis:

1. Click the **Analyze** menu, point to **Nonparametric Tests**, point to **Legacy Dialogs**, and then click **Chi-square**. The **Chi-square Test** dialog box opens.
2. Select the **Week of Response [week]** variable and move it to the **Test Variable List** box.
3. Select the **Values** option in the **Expected Values** section.
4. Type **6** in the **Values** box, and then click the **Add** button.
5. Repeat step 4, adding the values **15.1**, **18**, **12**, **11.5**, **9.8**, **7**, **6.1**, **5.5**, **3.9**, **2.1**, and **2** (in that order).
6. Click the **OK** button. The **Output Viewer** window opens.

   NOTE: The expected frequencies in this example are the response percentages that the firm has historically obtained with bulk mail.

	Observed N	Expected N	Residual
1	10	6.0	4.0
2	22	15.1	6.9
3	14	18.0	-4.0
4	10	12.0	-2.0
5	9	11.5	-2.5
6	8	9.8	-1.8
7	7	7.0	.0
8	10	6.1	3.9
9	4	5.5	-1.5
10	2	3.9	-1.9
11	2	2.1	-.1
12	1	2.0	-1.0
Total	99		

Figure 7 – First-Class/Bulk Mail Week of Response

	Week of Response
Chi-square	12.249[a]
df	11
Asymp. Sig.	.345

Figure 8 – Week of Response Test Statistics

Reporting the analysis results:

H_0: Do not reject. There was no statistical difference between customer response times using first-class mailing and customer response times using bulk mailing.

Explanation: The manufacturing company hoped that first-class mail would result in quicker customer response. As indicated in Figure 7, the first two weeks indicated different response times of four and seven percentage points, respectively. The question was whether the overall differences between the two distributions were statistically significant.

The Chi-Square statistic was calculated to be 12.249 at 11 degrees of freedom (see Figure 8). The significance value (p) associated with the data was 0.345, which was greater than the threshold value of 0.05. Hence, $\mathbf{H_0}$ was not rejected because there was no significant difference between first-class and bulk mailings. The first-class mail promotion did not result in response times that were statistically different from standard bulk mail. Therefore, bulk postage was more economical for direct mailings.

One-Way Analysis of Variance

One-way analysis of variance (One-Way ANOVA) procedures produce an analysis for a quantitative dependent variable affected by a single factor (independent variable). Analysis of variance is used to test the hypothesis that several means are equal. This technique is an extension of the two-sample t test. Think of it as a generalization of the pooled t test. Instead of two populations (as in the case of a t test), there are more than two populations or treatments.

Research Question # 4

Which of the alloys tested would be appropriate for creating an underwater sensor array?

To find the best alloy for an underwater sensor array, four different types are tested for resistance to corrosion. Five plates of each alloy are submerged for 60 days after which the number of corrosive pits on each plate is measured.

$\mathbf{H_0}$**:** The four alloys exhibit the same kind of behavior and are not different from one another.
$\mathbf{H_1}$**:** The four alloys exhibit different kind of behaviors and are different from one another.

To run One-Way ANOVA:

1. Open the **Alloy.sav** file.

   NOTE: Each case within the One-Way ANOVA data file represents one of the 20 metal plates (5 plates of 4 different alloys) and is characterized by 2 variables. One variable assigns a numeric value to the alloy. The other variable is used to quantify the number of pits on the plate after being underwater for 60 days (see Figure 9).

	alloy	pits	var
1	1.00	15.00	
2	1.00	17.00	
3	1.00	11.00	
4	1.00	18.00	
5	1.00	24.00	
6	2.00	62.00	
7	2.00	61.00	
8	2.00	58.00	
9	2.00	68.00	
10	2.00	54.00	

Figure 9 – Alloy Data File

2. In **Data View**, click the **Analyze** menu, point to **Compare Means**, and then click **One-Way ANOVA**. The **One-Way ANOVA** dialog box opens.
3. Select the **pits** variable in the box on the left and move it to the **Dependent List** box.
4. Select the **Alloy [alloy]** variable in the box on the left and move it to the **Factor** box (see Figure 10).

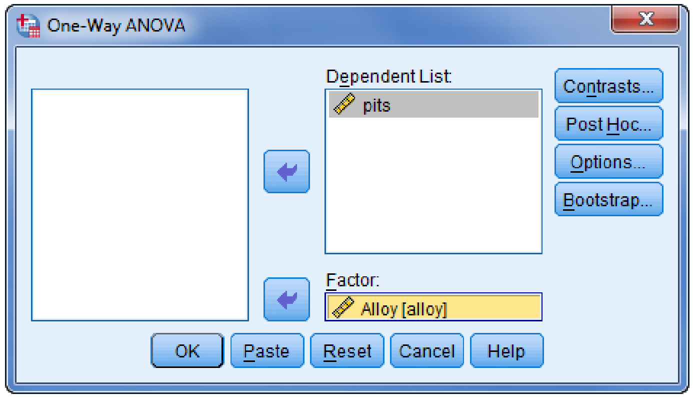

Figure 10 – One-Way ANOVA Dialog Box

5. Click the **Options** button. The **One-Way ANOVA: Options** dialog box opens.
6. Select the **Descriptive**, **Homogeneity of variance test**, and **Means plot** check boxes (see Figure 11).
7. Click the **Continue** button.

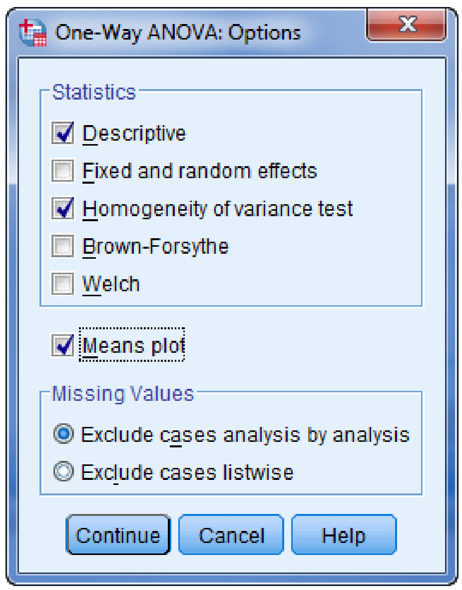

Figure 11 – One-Way ANOVA: Options Dialog Box

8. Click the **OK** button. The **Output Viewer** window opens.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	5	17.0000	4.74342	2.12132	11.1103	22.8897	11.00	24.00
2.00	5	60.6000	5.17687	2.31517	54.1721	67.0279	54.00	68.00
3.00	5	34.6000	4.92950	2.20454	28.4792	40.7208	28.00	40.00
4.00	5	19.4000	3.20936	1.43527	15.4151	23.3849	15.00	24.00
Total	20	32.9000	18.29840	4.09165	24.3361	41.4639	11.00	68.00

Figure 12 – ANOVA Descriptive Output

Levene Statistic	df1	df2	Sig.
.535	3	16	.665

Figure 13 – Output for Test of Homogeneity of Variances

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6026.200	3	2008.733	95.768	.000
Within Groups	335.600	16	20.975		
Total	6361.800	19			

Figure 14 – ANOVA Output

Reporting the analysis results:
H_0: Reject in favor of **H_1**.
H_1: The four alloys do not exhibit the same kind of behavior. They are statistically different from one another.

Explanation: Figure 12 lists the means, standard deviations, and individual sample sizes of each alloy. Figure 13 provides the degrees of freedom and the significance level of the population; *df1* is one less than the number of sample alloys (4-1=3), and *df2* is the difference between the total sample size and the number of sample alloys (20-4=16). Figure 14 lists the sum of the squares of the differences between means of different alloy populations and their mean square errors. In Figure 14, the *Between Groups* variation *6026.200* is due to interaction in samples between groups. If sample means are close to each other, this value is small. The *Within Groups* variation *335.600* is due to differences within individual samples. The *Mean Square* values are calculated by dividing each *Sum of Squares* value by its respective degree of freedom (*df*). The table also lists the F statistic *95.768*, which is calculated by dividing the *Between Groups Mean Square* by the *Within Groups Mean Square.* The significance level of *0.000* is less than the threshold value of 0.05 indicating that the null hypothesis can be rejected. In conclusion, the alloys are not all the same.

Post Hoc Tests

In ANOVA, if the null hypothesis is rejected, then it is concluded that there are differences between the means ($\mu_1, \mu_2, \ldots, \mu_a$). It is useful to know specifically where these differences exist. Post hoc testing identifies these differences. Multiple comparison procedures look at all possible pairs of means and determine if each individual pairing is the same or statistically different. In an ANOVA with α treatments, there will be α*(α-1)/2 possible unique pairings, which could mean a large number of comparisons.

Research Question # 5
Is the mean difference between alloy sets statistically significant?

The rejection of the previous null hypothesis leads to the conclusion that all the alloys do not exhibit the same behavior. The next part of the analysis determines if the mean difference between individual alloy sets is statistically significant.

H_0: $\mu_0 = \mu_1 \ldots = \mu_a$
H_1: $\mu_0 \neq \mu_1 \ldots \neq \mu_a$

To run post hoc tests:

1. In **Data View**, click the **Analyze** menu, point to **Compare Means**, and then click **One-Way ANOVA**. The **One-Way ANOVA** dialog box opens (see Figure 15).

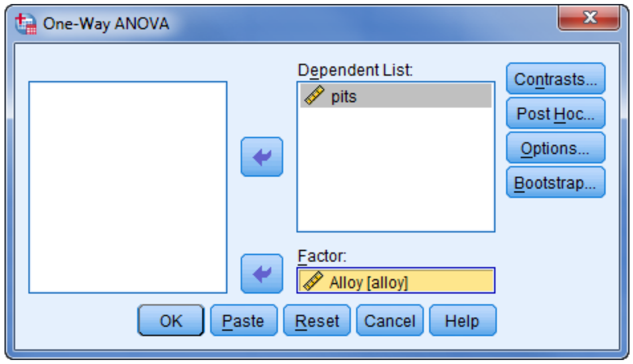

Figure 15 – One-Way ANOVA Dialog Box

2. Click the **Post Hoc** button. The **One-Way ANOVA: Post Hoc Multiple Comparisons** dialog box opens.
3. Select the **LSD** check box (see Figure 16).
4. Click the **Continue** button, and then click the **OK** button. The **Output Viewer** window opens.

   NOTE: LSD stands for *List Significant Difference*, which compares the means one by one.

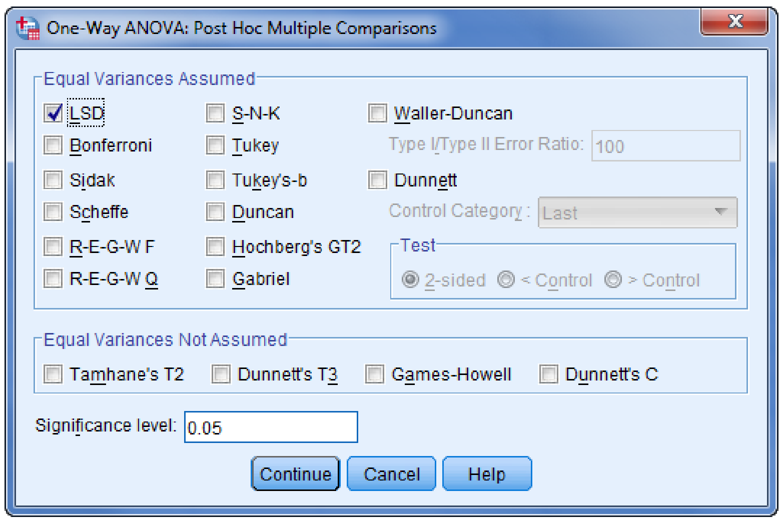

Figure 16 – One-Way ANOVA: Post Hoc Multiple Comparisons Dialog Box

(I) Alloy	(J) Alloy	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	-43.60000*	2.89655	.000	-49.7404	-37.4596
	3.00	-17.60000*	2.89655	.000	-23.7404	-11.4596
	4.00	-2.40000	2.89655	.420	-8.5404	3.7404
2.00	1.00	43.60000*	2.89655	.000	37.4596	49.7404
	3.00	26.00000*	2.89655	.000	19.8596	32.1404
	4.00	41.20000*	2.89655	.000	35.0596	47.3404
3.00	1.00	17.60000*	2.89655	.000	11.4596	23.7404
	2.00	-26.00000*	2.89655	.000	-32.1404	-19.8596
	4.00	15.20000*	2.89655	.000	9.0596	21.3404
4.00	1.00	2.40000	2.89655	.420	-3.7404	8.5404
	2.00	-41.20000*	2.89655	.000	-47.3404	-35.0596
	3.00	-15.20000*	2.89655	.000	-21.3404	-9.0596

*. The mean difference is significant at the 0.05 level.

Figure 17 – Multiple Comparisons Output

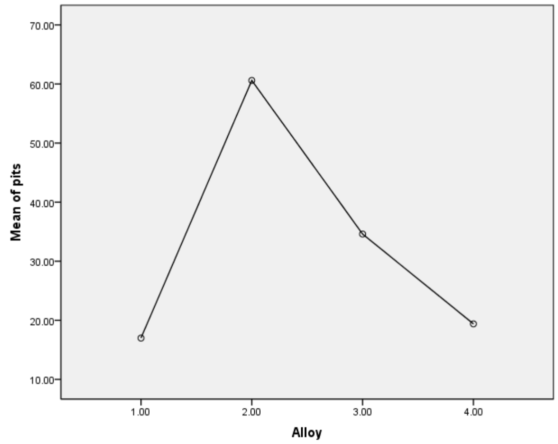

Figure 18 – Means Plot

<u>Reporting the analysis results</u>:
H_0: Reject in favor of **H_1**.
H_1: At least one of the means is different.

Explanation: Figure 17 shows the results of comparing pairs of means between different alloy sets. Each row indicates the difference between the two corresponding treatments. Alloys *1* and *4* have a mean difference of *2.4* (a relatively small value). Also, the significance level of *0.420* indicates that the null hypothesis cannot be rejected for the comparison of alloys *1* and *4*. There is no statistically significant difference between them.

Alloy pairs *1* and *2*, *1* and *3*, *2* and *3*, *2* and *4*, and *3* and *4* have large mean differences with significance values of *0.000*. In these cases, the null hypothesis can be rejected, leading to the conclusion that they are statistically different. Also, the means plot (see Figure 18) shows that alloys *1* and *4* have average mean values of pits very close to each other. Because alloys *1* and *4* have the lowest mean number of corrosive pits, they are the best choices for the array. Depending on the relative costs of the two alloys, the one that is more cost effective can be chosen.

Two-Way Analysis of Variance

Two-way analysis of variance (Two-Way ANOVA) is an extension of the one-way analysis of variance. With Two-Way ANOVA, two or more independent variables can be tested instead of just one. Using multiple variables has two advantages: increased efficiency and an increase in the result's statistical power.

<u>Research Question # 6</u>
Will test anxiety and different teachers affect student test scores?

To answer the question, two classes were given a cumulative standardized test. Before the test, students were asked about the level of anxiety they were feeling: none, some and lots. The experiment compared two different classes taught by teachers A and B using the same test. After evaluating the standardized test, each group's mean score was examined.

H_0: Test anxiety and teacher do **<u>not</u>** affect student test scores.
H_1: Test anxiety and teacher do affect student test scores.

To run Two-Way ANOVA:

1. Open the **Two-Way-ANOVA.sav** file (see Figure 19).

	ANXIETY	TEACHER	SCORE
1	none	A	23
2	none	A	32
3	none	A	25
4	some	A	29
5	some	A	30
6	some	A	34
7	lots	A	31
8	lots	A	36
9	lots	A	33
10	none	B	32
11	none	B	26
12	none	B	26
13	some	B	34
14	some	B	41
15	some	B	35
16	lots	B	23
17	lots	B	26
18	lots	B	32

Figure 19 – Two-Way ANOVA Data File

2. In **Data View**, click the **Analyze** menu, point to **General Linear Model**, and then click **Univariate** (see Figure 20). The **Univariate** dialog box opens.
3. Select the **SCORE** variable in the box on the left and move it to the **Dependent Variable** box.
4. Select the **ANXIETY** and **TEACHER** variables in the box on the left and move them to the **Fixed Factor(s)** box (see Figure 21).

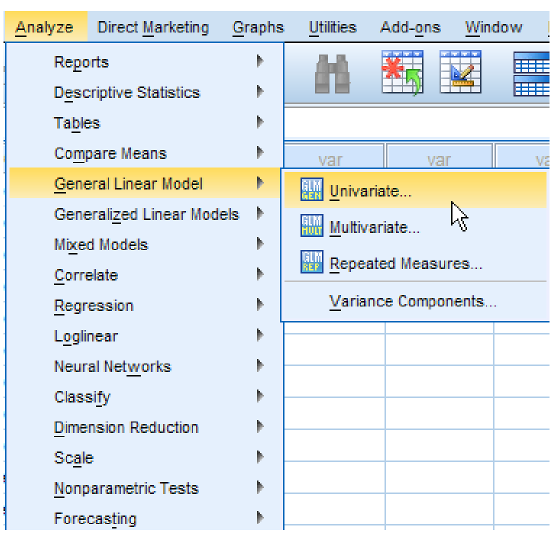

Figure 20 – Analyze Menu When Selecting Univariate

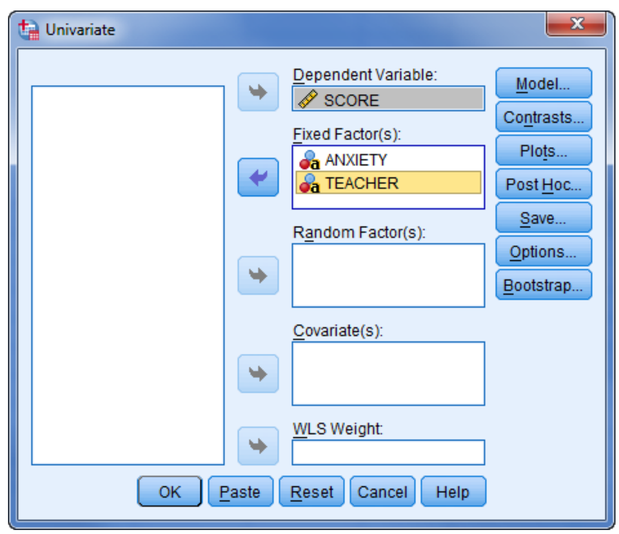

Figure 21 – Univariate Dialog Box

5. Click the **Options** button. The **Univariate: Options** dialog box opens.
6. Select the **Descriptive statistics** check box, and then click the **Continue** button (see Figure 22).
7. Click the **OK** button. The **Output Viewer** window opens (see Figure 23 and Figure 24).

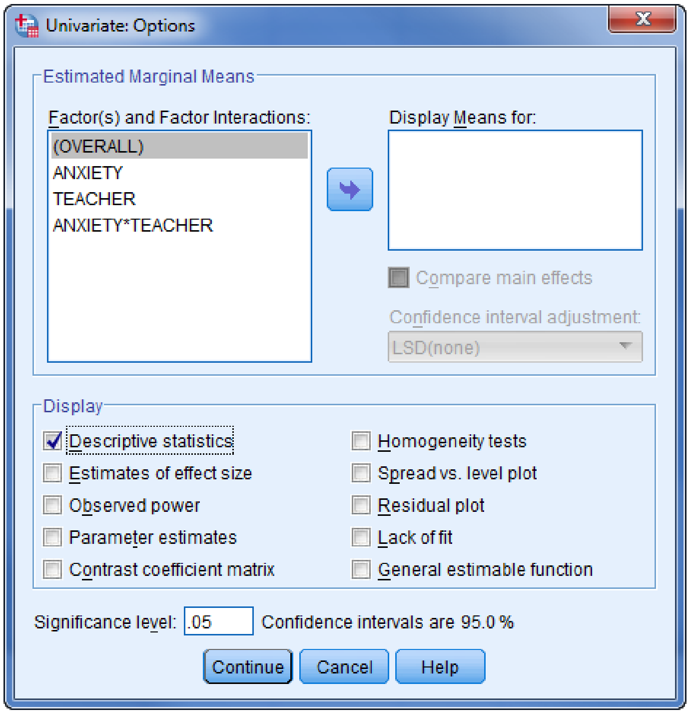

Figure 22 – Univariate: Options Dialog Box

Descriptive Statistics

Dependent Variable: SCORE

ANXIETY	TEACHER	Mean	Std. Deviation	N
lots	A	33.33	2.517	3
	B	27.00	4.583	3
	Total	30.17	4.792	6
none	A	26.67	4.726	3
	B	28.00	3.464	3
	Total	27.33	3.777	6
some	A	31.00	2.646	3
	B	36.67	3.786	3
	Total	33.83	4.262	6
Total	A	30.33	4.183	9
	B	30.56	5.747	9
	Total	30.44	4.878	18

Figure 23 – ANOVA Descriptive Output Table

Tests of Between-Subjects Effects

Dependent Variable: SCORE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	238.444[a]	5	47.689	3.447	.037
Intercept	16683.556	1	16683.556	1206.040	.000
ANXIETY	127.444	2	63.722	4.606	.033
TEACHER	.222	1	.222	.016	.901
ANXIETY * TEACHER	110.778	2	55.389	4.004	.047
Error	166.000	12	13.833		
Total	17088.000	18			
Corrected Total	404.444	17			

a. R Squared = .590 (Adjusted R Squared = .419)

Figure 24 – Output Table for Tests of Between-Subjects Effects

<u>Reporting the analysis results</u>:
H_0: Reject in favor of **H_1** for **Anxiety** and the interaction between **Anxiety** and **Teacher** (**Anxiety*Teacher**).
H_1: Testing anxiety and teacher affect student test scores.

Explanation: Figure 23 lists the means and standard deviations from three anxieties in two classes. Students who have *some testing anxiety* and are in *Teacher B's* class have the highest mean score (mean=36.67). Because the significance value of the *Teacher* variable (0.901) is more than the threshold value (0.05) as indicated in Figure 24, it can be concluded that the *Teacher* factor alone does not affect test scores. The significance values of the *Testing Anxiety* variable (0.033) and the interaction between the two factors *Anxiety*Teacher* (0.047) are less than the threshold value (0.05), leading to the conclusion that both *Testing Anxiety* alone and the combination of *Testing Anxiety* and *Teacher* (*Anxiety*Teacher*) do affect student test scores.

Importing and Exporting Data

SPSS Statistics can be used to analyze data in a Microsoft Excel spreadsheet. SPSS Statistics provides the ability to import an Excel spreadsheet directly into the Data Editor window and automatically create variables based on the spreadsheet's column headings. Data can also be exported from SPSS Statistics into Microsoft Excel and PowerPoint.

To import an Excel spreadsheet into SPSS Statistics:

1. Click the **Open** button on the **Data Editor** toolbar. The **Open Data** dialog box opens.
2. Click the **Files of type** arrow and select **Excel (*.xls, *.xlsx, *.xlsm)** from the list.
3. Select the **demo.xls** file, and then click the **Open** button (see Figure 25). The **Opening Excel Data Source** dialog box opens (see Figure 26).

   <u>NOTE</u>: If the Excel file contains multiple worksheets, select the desired worksheet by clicking the **Worksheet** arrow (see Figure 26). To import a specific range of cells, specify the range in the **Range** box.

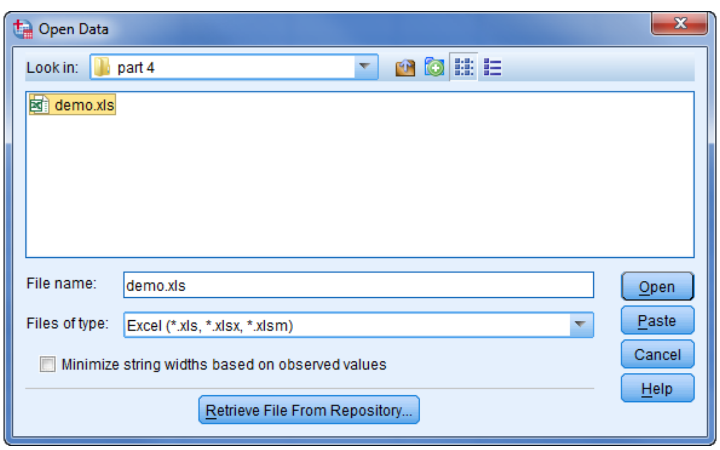

Figure 25 – Open Data Dialog Box

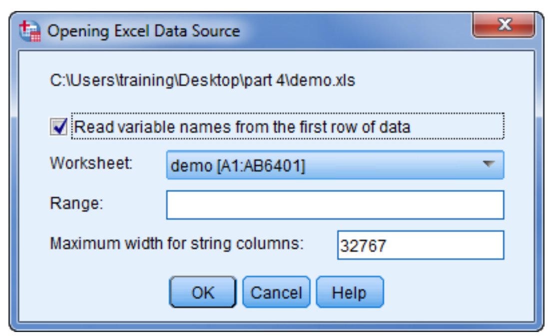

Figure 26 – Opening Excel Data Source Dialog Box

4. Click the **OK** button. SPSS Statistics processes and reads the Excel file and converts all first row column headings into variables using the best approximation for the variable attributes (see Figure 27 and Figure 28).

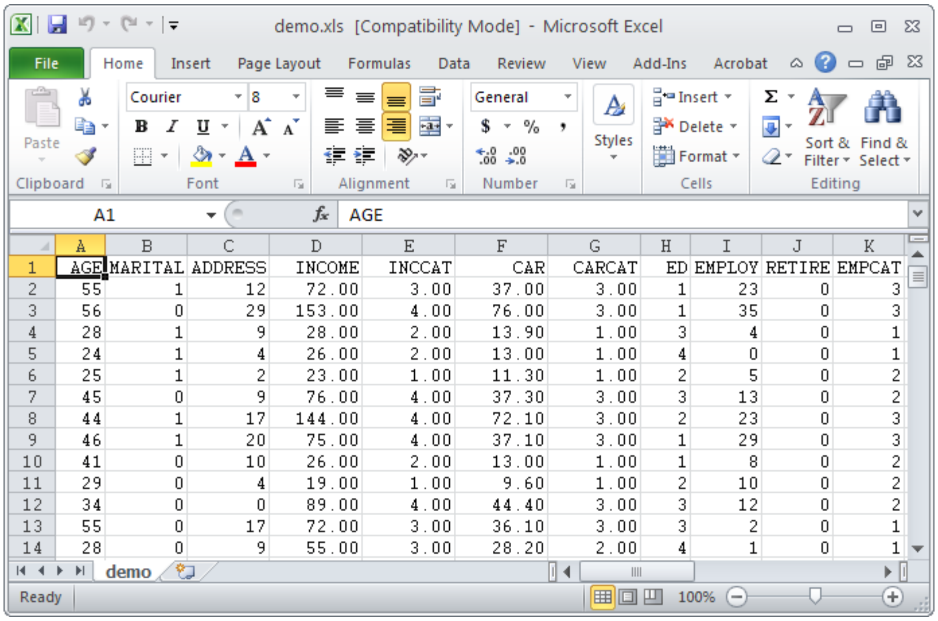

Figure 27 – Excel File

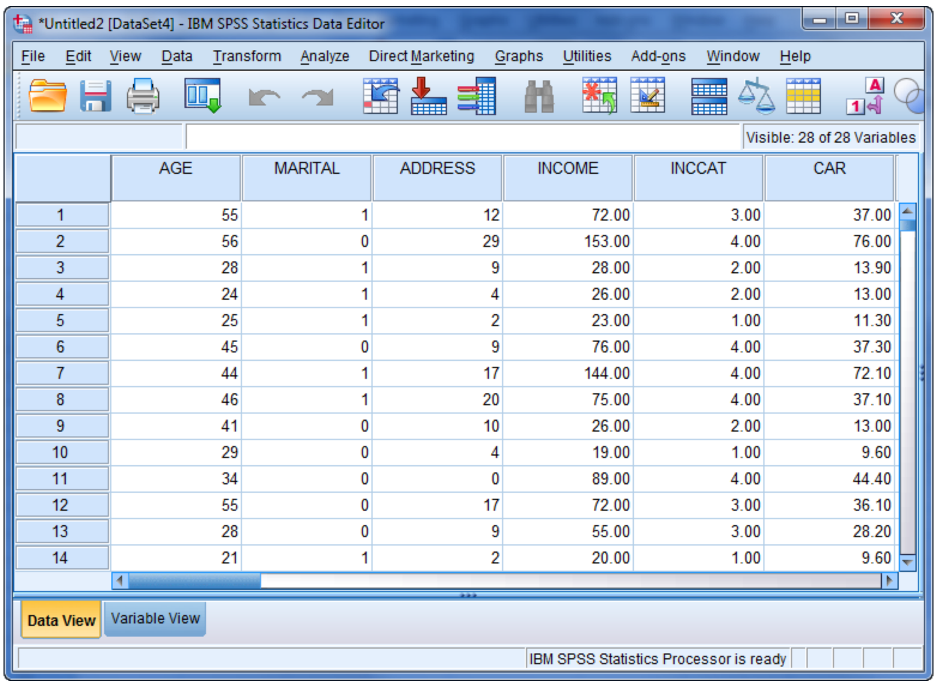

Figure 28 – Excel File Imported into SPSS Statistics

The reverse situation may also arise, where data in an SPSS Statistics file must be analyzed using Excel. This can be accomplished by exporting the contents of the Data Editor window into an Excel spreadsheet.

To export SPSS Statistics data into an Excel spreadsheet:

1. In the **Data Editor** window, click the **File** menu, and then click **Save As**. The **Save Data As** dialog box opens.
2. Click the **Save as type** arrow and select **Excel 97 through 2003 (*.xls)** or **Excel 2007 through 2010 (*.xlsx)** from the list (see Figure 29).

   NOTE: Selecting the **Write variable names to spreadsheet** check box will cause SPSS Statistics to write the variable names as column headings in the spreadsheet.

NOTE: If only certain variables from the **Data Editor** window are desired in the spreadsheet, users can click the **Variables** button and select or deselect variables in the **Save Data As: Variables** dialog box (see Figure 30).

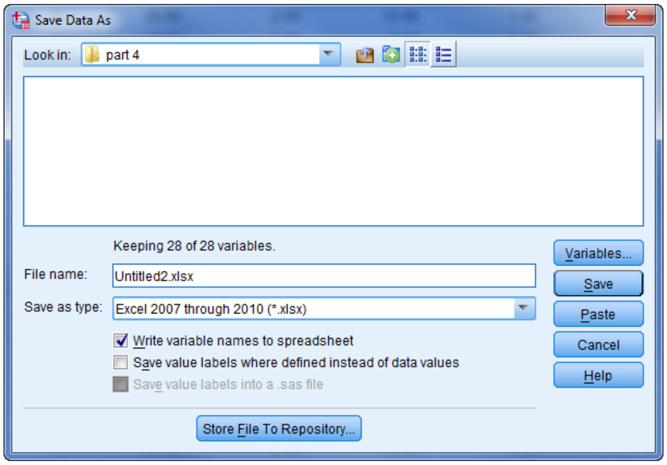

Figure 29 – Save Data As Dialog Box

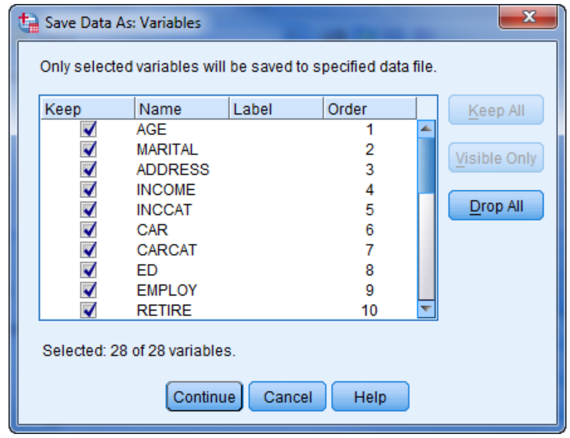

Figure 30 – Save Data As: Variables Dialog Box

3. Click the **Look in** arrow and select a location to save the file.
4. Type a name for the Excel file in the **File name** box.
5. Click the **Save** button. The **Output Viewer** window opens with a report summarizing the details and results of the export operation (see Figure 31).

```
Data written to C:\Documents and Settings\Desktop\Part4\Export.xlsx.
28 variables and 6400 cases written to range: SPSS.
Variable: AGE                Type: Number   Width:  11   Dec: 0
Variable: MARITAL            Type: Number   Width:  11   Dec: 0
Variable: ADDRESS            Type: Number   Width:  11   Dec: 0
Variable: INCOME             Type: Number   Width:  11   Dec: 2
Variable: INCCAT             Type: Number   Width:  11   Dec: 2
Variable: CAR                Type: Number   Width:  11   Dec: 2
Variable: CARCAT             Type: Number   Width:  11   Dec: 2
Variable: ED                 Type: Number   Width:  11   Dec: 0
```

Figure 31 – SPSS Statistics Export Output Report

To export an SPSS Statistics Output chart into a PowerPoint slide:

1. In the **Output Viewer** window, click to select the table. A box appears around the table and a red arrow to the left of it.
2. Click the **File** menu, and then click **Export**. The **Export Output** dialog box opens.
3. Click the **Type** arrow and select **PowerPoint (*.ppt)** from the list (see Figure 32).
4. Click the **Browse** button. The **Save File** dialog box opens.
5. Click the **Look in** arrow and select a location to save the file.
6. Type a name for the PowerPoint file in the **File name** box (see Figure 33).
7. Click the **Save** button.
8. Click the **OK** button.

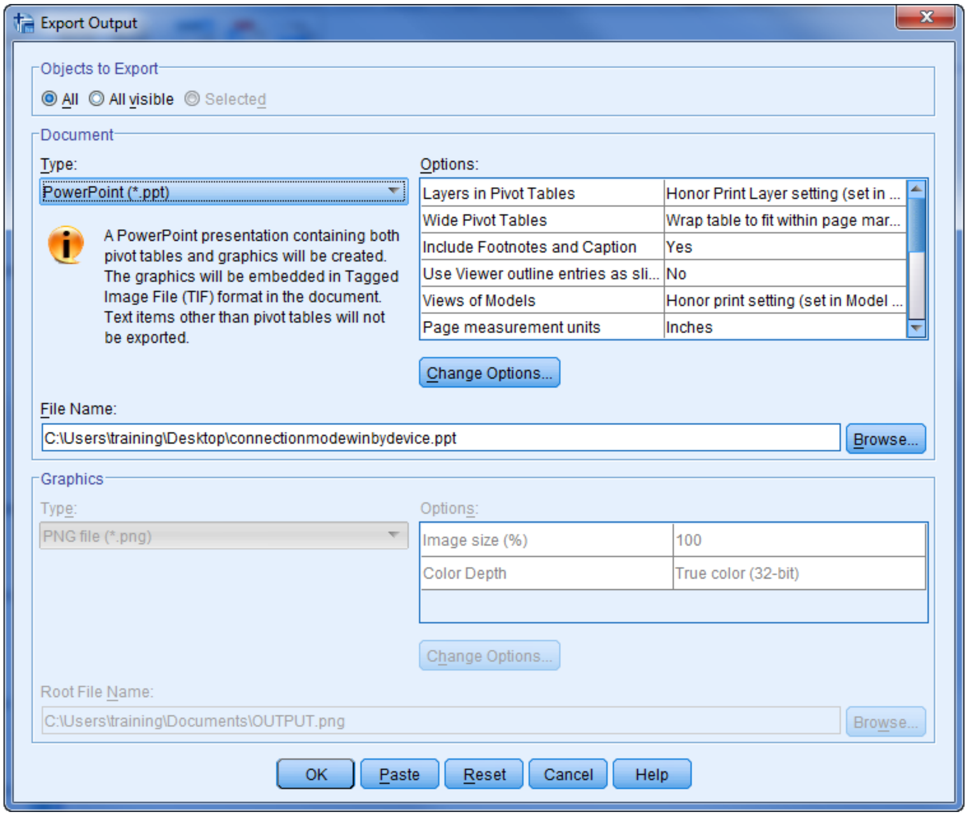

Figure 32 – Export Output Dialog Box

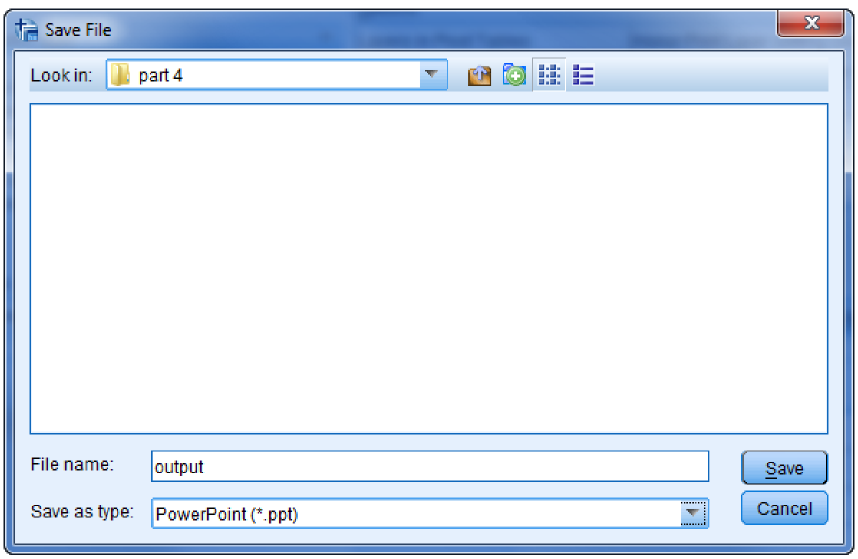

Figure 33 – Save File Dialog Box

Using Scripting for Redundant Statistical Analyses

Every statistical analysis used by SPSS Statistics is executed through a special programming language. The code used for each analysis can be captured, stored as a script file, and edited if necessary. A series of scripts in a script file can be run either individually or all at once. Scripting automates a series of statistical analyses that are performed on a file that always has the same variables, but contains data that changes. Scripts are captured and edited in the *IBM SPSS Statistics Syntax Editor* window.

The following example illustrates the benefits of capturing, storing and running scripts. The sample data is taken from a classroom setting for a weeklong course. At the end of each week, each student's data is compiled. The variables in the set include the subject name, gender, pretest scores, posttest scores, grade point average, computer ownership, and method of administering examinations. Each week, a report is generated that answers a series of questions about the class from the previous week. The questions answered and the statistical analyses used are the same every week, as described in Table 1.

Table 1 – Scripted Questions and Statistical Techniques

Question	Statistical Technique(s) to Answer Question
Does the data set include equal numbers of each gender and each test method?	Split the file Crosstabs
Is there a difference between the male and female pretest scores?	Select all cases Independent-Samples T Test
Is there a difference between the male and female posttest scores?	Independent-Samples T Test
Is there a difference between the overall pretest and posttest scores?	Paired-Samples T Test
Do gender, computer ownership, and test method affect test scores?	Three-Way ANOVA
Do gender, computer ownership, and test method affect test scores differently depending on gender?	Split the file Two-Way ANOVA
Is there a linear relationship between the pretest and posttest scores for each gender?	Scatter plot graph with file split
Can pretest scores predict posttest scores for each gender?	Simple regression with file split
Is there an overall linear relationship between pretest and posttest scores?	Select all cases Scatter plot graph
Can pretest scores predict posttest scores?	Simple regression

To construct a script file that will automatically run the analyses:

1. Open the **ClassData.sav** file.
2. Click the **Edit** menu, and then click **Options**. The **Options** dialog box opens.
3. Click the **Viewer** tab, select the **Display commands in the log** check box, click the **Apply** button, and then click the **OK** button (see Figure 34).

   NOTE: The script file is built by performing each statistical analysis in the desired order. All analyses must be performed manually one time while the file is being built. In this example, the file will first be split before creating a crosstabs table.

4. Click the **Data** menu, and then click **Split File**. The **Split File** dialog box opens.
5. Select the **Compare groups** option, and then move the **gender** variable to the **Groups Based on** box.
6. Click the **Paste** button to add the command to the script file. The **Split File** dialog box closes and the **IBM SPSS Statistics Syntax Editor** window opens with the pasted command displayed (see Figure 35).
7. In the **IBM SPSS Statistics Data Editor** window, click the **Analyze** menu, point to **Descriptive Statistics**, and then click **Crosstabs**. The **Crosstabs** dialog box opens.

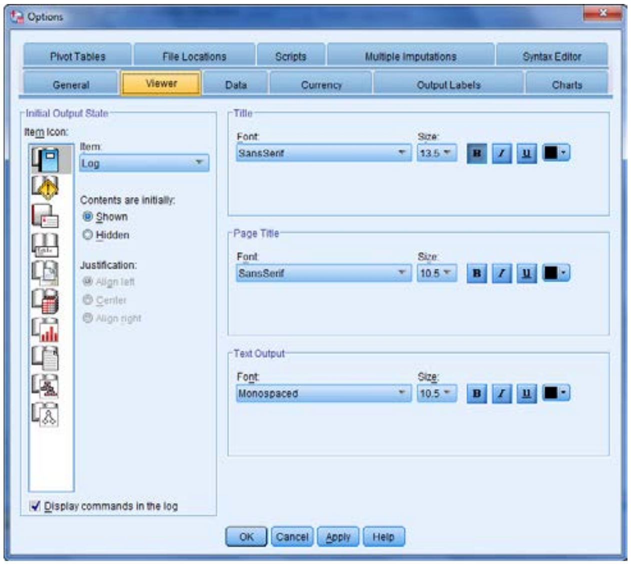

Figure 34 – Options Dialog Box

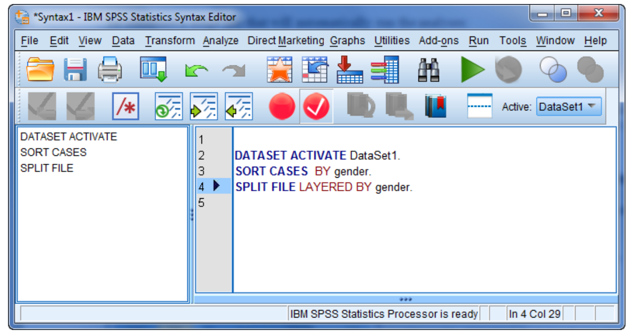

Figure 35 – IBM SPSS Statistics Syntax Editor Window

8. Move the **gender** variable to the **Row(s)** box and the **method** variable to the **Column(s)** box.
9. Click the **Paste** button. The **Crosstabs** dialog box closes and the command is pasted in the **IBM SPSS Statistics Syntax Editor** window (see Figure 36). The first question in Table 1 has been entered into the script file.

   NOTE: Scripts for each of the remaining analytical techniques can be entered into the script file by choosing the desired parameters in each dialog box, and then clicking the **Paste** button.

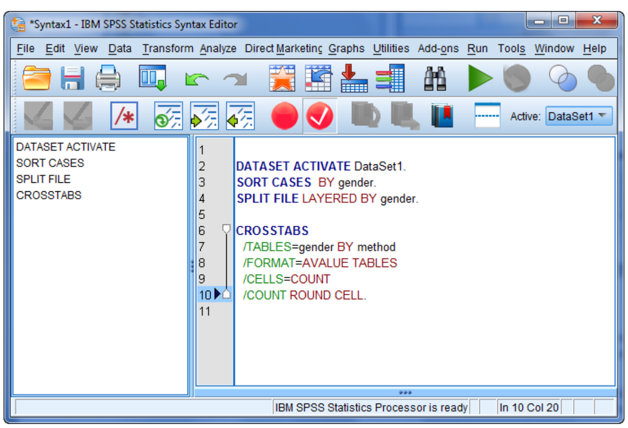

Figure 36 – IBM SPSS Statistics Syntax Editor Window

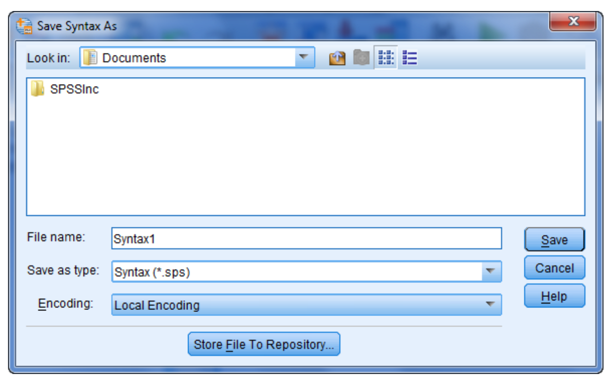

Figure 37 – Save Syntax As Dialog Box

10. To save the script file, click the **File** menu in the **IBM SPSS Statistic Syntax Editor** window, and then click **Save As**. The **Save Syntax As** dialog box opens (see Figure 37).
11. Select a location to save the file, enter a file name, and then click the **Save** button.

SPSS Statistics script files have the *.sps* file extension. The program provides several options for running script files. The *Run* menu of the *IBM SPSS Statistic Syntax Editor* window contains commands for *All*, *Selection*, and *To End* (see Figure 39).

To run an existing script file:

1. In the **Data Editor** window, click the **File** menu, point to **Open**, and then click **Syntax** (see Figure 38). The **Open Syntax** dialog box opens.
2. Locate and select the **WeeklyAnalysis.sps** syntax file, and then click the **Open** button. The **IBM SPSS Statistics Syntax Editor** window opens with the script displayed.
3. In the **IBM SPSS Statistics Syntax Editor** window, click the **Run** menu, and then click **All** (see Figure 39). Every command in the script file is executed and the results are displayed in the **Output Viewer** window.

   NOTE: If the **Display commands in the log** check box on the **Viewer** tab of the **Options** dialog box remains selected, individual script commands will appear with the output in the **Output Viewer** window.

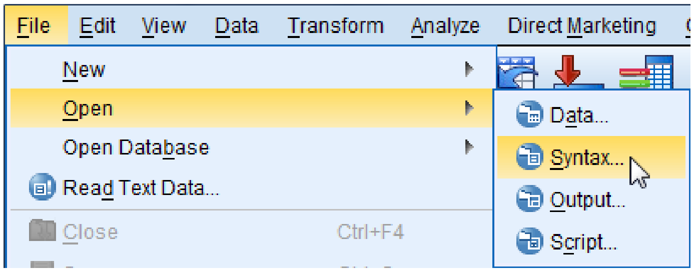

Figure 38 – File Menu When Selecting Syntax

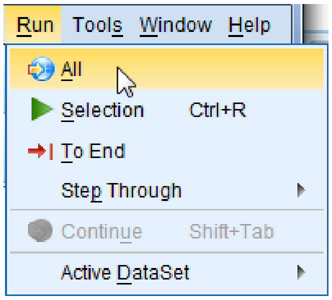

Figure 39 – Run (Syntax) Menu