

Psych. Research 1 Guide to SPSS 11.0

I. What is SPSS:

SPSS (Statistical Package for the Social Sciences) is a data management and analysis program. It allows us to store and analyze very large amounts of research data. The statistics that SPSS is capable of are far more complex than the stats that we can do in excel which makes it more desirable as an analysis tool. Also, spss allows us to store our data, protocols (syntax), and results (output) in separate files, which makes analysis of large amounts of data much less cumbersome than excel.

II. Goals:

Our goals for this unit include the following aspects:

1. Learn to set up data files and enter/import data (III)
2. Learn to create syntax files (VI)
3. Learn to generate descriptive statistics (V)
4. Learn to generate Frequency statistics (VI)
5. Learn to compute new variables from existing variables (VII)
6. Learn to transform/"normalize" variables (VIII)
7. Learn to use filters (selecting cases) (IX)
7. Learn to compute chi square (single variable, multivariate) (X)
8. Learn to compute correlation coefficients (XI)
9. Learn to compute t-tests : independent sample, repeated measure. (XII)

III. Data Files: Set up, Entering Data, & Importing Data

--For the most part SPSS can be split up into 3 major parts: The Data Editor (where we enter data, name variables, compute new variables, and select cases), The Syntax Editor (where we store & create syntax for our analyses and procedures), and The Output Navigator (where we view the results our statistical tests have generated).

-- The first step in this process is to **set up the data file** in the Data Editor (note: we can create variables and enter data in the syntax editor, but it is beyond the scope of this course).

-if we have already created a data file in another program we can import it by selecting "open" on the "File" pull-down menu. An "open file" box will appear. If the desired file is not an spss file (*.sav), scroll through the "file type" options and choose the appropriate format (e.g. excel (*.xls)). Then open the desired file. (Note: if desired file is not created in spss a second box will appear. If you have variables named in this file then check the "read variable names" box, other wise leave blank. Also, you can designate portions of files to open using the "range" option, see help for details).

Defining Variables - If you are creating a new data file, editing an existing data file or have imported a data file where the variables have not yet been named and/or have not had value labels associated with the variables then you will need to begin here.

-there are 2 parts (or views) to the Data Editor:

1. **The Data View**: This view allows you to view and input data values. The columns represent Variables and the rows represent participants/subjects (often referred to as cases).
2. **The Variable View**: This view allow you to edit variables and add new variables to the data set. Note that the rows represent each variable and corresponds to the columns in the data view. Also, the columns in this view represent different aspects of each variable.

3. To toggle back and forth between the "Data View" and the "Variable View" Click on the small, labeled file tabs located at the bottom left hand

corner of the data editor spread sheet

4. To create variables on a blank data file, select the variable view. Note that no variables are listed.

a. Give the variable a **“Variable Name”**. Variable Names are limited to 8 characters (a throw back from the old Unix days). The Name can not start with a number (though it can have numbers in it). Nor can it have any spaces or symbols (i.e., !@#\$\$%^&*+~ () { } [] ? / > < .) in it except (- and _).

b. Give the variable a **“Variable Label”**, by clicking the appropriate cell in the “Label” column. Variable Labels are more flexible than Variable names. You can have more than 200 characters and you can use spaces and symbols. The variable label allows you to give a more descriptive name to the variable, that will make sense to you when you come back and look at your data after a long period of time has elapsed. Be as precise as possible. Also, when you give a variable a “Variable Label” the variable label will appear on the output of your analyses (see below). However, notice that labels more than 40 or so characters will not be truncated on the output (it will only present the 1st 40 characters, so put the most unique and descriptive information first. Otherwise your outputs may become vary confusing.

c. If the variable is a categorical variable (e.g. gender, class rank, ethnicity, group membership) then you will need to define the **Value labels**. You do not need to do this if your data is ratio, interval, or ordinal data (like 1-7 numerical rating scales). By adding Value Labels you will not have to try to remember what the numbers stand for (e.g., 1 = male, 2 = female). Also, the Value Label will be printed on the output.

c1. To define Value Labels, click on the appropriate cell of the “Values” column. A dialogue box labeled “Value Labels will appear.

c2. In the “**value**” field first type the lowest value (e.g. 0 or 1), then in the “**value label**” field type the name of the category (e.g. female), and finally click the “**add**” button.

c3. next in the “**value**” field type the next highest value (e.g. 1 or 2), then in the “**value label**” field type the name for that category (e.g. male), and finally click the “**add**” button.

c4. continue this process until all the values for all the categories have been named. When this is complete left-click “**OK**”.

-It is always a good idea to make your first variable subject number (subnum) so that if certain subjects have to be excluded from later analyses (e.g. because of missing data or some other criterion) they can be easily sorted out based on subject number.

Entering Data-Once the variables are named and the values labeled you can begin data entry. In the Data View, simply select the cell you want to begin with, type the appropriate value into that cell, and then either press “enter” or one of the direction arrow keys. If entering data on the right-hand numeric-key-pad be sure that the number lock has been turned on and be careful not to turn it off when reaching for the 7 key. Data can be corrected by selecting the desired cell and typing in the new value and pressing either enter or a directional arrow key.

Saving Files—Once you have your variables defined and perhaps some data entered you should periodically save your file. If you have imported your data from another program DO NOT overwrite your original file, make a new file and save it as an spss data file (*.sav). Chose “**save as**” from the “**file**” pull down menu and save the file in the appropriate directory (e.g. a:). Note: you do not have to give the file a name different from the original data file, because it will have a different extension (i.e. .sav not .xls, if the original file was from excel) and will therefore not overwrite your original data file.

-Be sure to save your work often and save it in multiple places (e.g. make backups) so you will not loose anything important.

IV. Syntax Files:

What is Syntax-In the olden days (8-10 years ago) there did not exist a Graphical User Interface (GUI) version of spss and all data entry and analysis was done using an spss syntax language. Now we have a windows based program (a GUI) and we can point & click our way through all analyses. The only problem with this is that when we want to change our analysis procedure we must step through the whole point click process again which is tiresome and potentially error ridden. However, throughout spss there is the option to “**paste syntax**” which will send our point click commands to a syntax sheet, which can be stored as a separate file. This way we can have all the steps in our analysis recorded and stored. Also, we can run all of our analysis from the syntax sheet, write syntax for new analyses, or alter the syntax from previous analyses.

Opening a Syntax Sheet - if no syntax sheet is open when we paste our first procedure then spss automatically opens an untitled sheet for us. Later, we can title and save as a syntax file (.sps). Otherwise it will paste syntax to the bottom of our currently opened syntax sheet.

-To create a new syntax sheet, select “**new**” from the “**file**” pull-down menu and then select syntax from the side menu.

-To open an existing syntax file, select “**open**” from the “**file**” pull-down menu. The “**open file**” dialogue box will appear. You will need to change the “file type” line to .sps. Otherwise the desired file will not appear as an option (only .sav files will). Finally, choose the desired file from the desired directory, and click “open”

Editing Syntax - Syntax can be moved and removed using the copy, cut, & paste commands identical to those found in most word processors and spreadsheet programs.

Running Analyses From Syntax - to run an analysis or multiple analyses from the syntax box, highlight the syntax for the desired analyses and then press the play button on the tool bar. The “**play**” button (>) resembles the play button on a tape recorder or VCR (arrow pointing to the right). Also, analyses can be run from the “**run**” pull down menu where there are four options: **All** (runs all analyses on all syntax sheets that are open), **selection** (runs highlighted area), **current** (runs all analyses on the currently active syntax sheet), and **to end** (runs analyses on the current syntax sheet that fall below the point where the cursor is currently positioned).

V. Output Files:

- Anytime you run an analysis (e.g., Desriptives, Frequencies, Chi-Square, etc.) the results will be presented in a separate window titled Output Viewer. Like syntax, this is a separate file that will need to be save. The extension for these types of files are “.spo” (short for spss output).

Page Setup and Saving

- When you generate output that you plan on saving and printing out, then you should select the “**Page Setup**” option from the pull-down “File” menu. A “Page Setup” dialogue box will appear. Click on the “Options” button. A second Dialogue box will appear titled “Page

Setup: Options”. Here you can give your file a header that will always appear on the printed output. Clicking on the Calendar icon will give the date the output was printed. The clock gives the time it was printed. # gives you the page number (very handy if you accidentally mix the pages). The icon that looks like a sheet of paper will print the file name on the header. Also you can type your own message in the header. This makes keeping track of your output very easy.

– To save the file simply click the floppy disc icon or select save (or save as) from the pull-down file menu.

Opening an Output File -

–To open an existing output file, select “**open**” from the “**file**” pull-down menu. The “**open file**” dialogue box will appear. You will need to change the “file type” line to .spo. Otherwise the desired file will not appear as an option. Finally, choose the desired file from the desired directory and click “open”

Note that all analyses can be run from the output window, but data can not be entered and “Transformations” of the data (See below) can’t be made from this window

V. Descriptive Statistics

What are Descriptive Statistics – This procedure will give you a variety of basic statistical options:

Mean	Sum	Standard Deviation	Variance	Range	Minimum	Maximum
Standard Error of the Mean			Kurtosis	Skewness		

–This type of data is the first analysis of any variable. It allows us to determine if a measure is behaving in the way we want, e.g. does it have enough variance to be useful, is it skewed, is it kurtotic, is the range adequate or is there some kind of ceiling or floor effect.

Generating Descriptive Analyses – in either the Data Editor, Syntax Sheet, or Output Navigator, select “**Descriptive Statistics**” from the “**Analyze**” pull-down menu and then select the “**Descriptives**” option from the side menu. A “**descriptives**” box will appear.

--In the left hand column appear all the variables you have data for. Move the variables you want analyzed to the “**variable(s)**” column. Do this by either double-left-clicking on the desired variable or single-left-clicking on the desired variable and then left-clicking on the boxed arrow between the columns.

--To remove a variable from the “**variable(s)**” list follow the same procedures.

– To choose the descriptive statistics you want, left click the “**options**” button (the default stats are Mean, Standard Deviation, Minimum and Maximum) and check the boxes of the stats you want. In general it is best to ask for all the stats possible just in case you have a need for them later. Then click “**continue**”.

–To paste the syntax for the descriptive stats you want left-click the “**paste**” button in the “**Descriptives**” box.

–To run analyses without pasting them to the syntax sheet, left-click “**OK**”

Interpreting Descriptive Analyses Output - Descriptive output has 2 components: Title, Descriptive Statistics. Each can be displayed or hidden by double-left-clicking on their label in the outline located in the left-hand column. Within the Descriptive Statistics box, the rows represent variables and the columns indicate the various requested statistics. The last row is labeled “Valid N (Listwise)”: this refers to the number of subjects who are not missing any data for any of the variables listed. The casewise N’s are listed above this, these are the number of subjects who are not missing the data for each variable (each case).

VI. Frequency Statistics

What are Frequency Statistics -- The Frequency Analysis allows us to generate ungrouped

frequency tables (tells us how often each score occurred) for our variables of interest while also generating all of the statistics that the Descriptives command allows us to generate (although, we can not get a listwise valid N). The ungrouped frequency tables are useful for providing us with a visual representation of the distributions for our variables, something we can't get from descriptive statistics alone. Further, within the Frequencies command we can generate graphical displays of our data (e.g. histograms, bar charts, pie charts, line graphs, scattergrams, etc).

Generating Frequencies – in either the Data Editor, Syntax Sheet, or Output Navigator, select “**Descriptive Statistics**” from the “**Analyze**” pull-down menu and then select the “**Frequencies**” option from the side menu. A “**Frequencies**” box will appear.

--In the left hand column appear all the variables you have defined. Move the variables you want analyzed to the “**variable(s)**” column by either double-left-clicking on the desired variable or single-left-clicking on the desired variable and then left-clicking on the arrow between the columns.

--To remove a variable from the “**variable(s)**” list follow the same procedures.

– **Statistics** --To choose the descriptive statistics you want (the default stats are None), left click the “**Statistics**” button and a “**Frequencies: Statistics**” box will appear. This box is comprised of four parts: Percentile Values, Central Tendency, Dispersion, and Distribution.

-Percentile Values has three options: Quartiles (this will generate the cut off points for dividing your subjects into 4 equal groups for the selected variables), Cut Point For ? Equal Groups (generates cut off points for any number of equal sized groups you select), Percentile (will give you the raw scores associated with any percentiles you “**Add**” to the percentile list).

- Central Tendency : contains the following options: Mean, Median, Mode, and Sum.

- Dispersion : contains the options: Standard Deviation, Variance, Range, Minimum, Maximum, and Standard Error of the Mean (SEM).

-Distribution : contains options for Skewness and Kurtosis, which tell us how far our distribution deviates from a normal distribution.

Check the boxes of the stats you want. Again, it is best to ask for all the stats possible just in case you have a need for them later, but the percentiles are not that necessary unless you know you will use them for something. Then click “**Continue**”.

–**Charts**--To generate graphical representations (e.g. figures) of your data distributions then click on the “**Charts**” button. There are three main chart types available on in the “**Charts**” box: Bar Chart (for use with discrete/categorical data), Pie Chart (for use with both discrete/categorical & continuous data). Histogram (for use with continuous data). (note more chart options are accessible from the “**Graphs**” pull-down menu on the upper-most tool bar.

-For the Bar chart and the Histogram there is also the an option that allows you to display the estimated normal curve for your data. This option allows you to see how much your true distribution deviates (differs from) the theoretical normal distribution (e.g. how much skewness and kurtosis your data has).

-For all charts you also have the option of representing either the absolute frequency (i.e. the total number of occurrences of each score) or the percentage (i.e. the total number of occurrences of each score divided by the total number of scores).

–**Format**– to specify how you want the frequency distribution to appear left-click the “**Format**” button. There are 5 options in the “**Frequencies: Format**” box:

-Ascending Value: Will order the chart so that the smallest scores are at the top and largest are at the bottom .

-Descending Value: will order the chart so that the largest scores are at the top and the smallest scores are at the bottom.

-Ascending Counts: will order the chart so that the least frequent scores are at the top and the most frequent scores are at the bottom.

- Descending Counts: will order the chart so that the most frequent scores are at the top and the least frequent scores are at the bottom.
- Suppress tables with more than ? Categories: this allows you to exclude tables for variables where all scores have a frequency of 1. In such a case, if you have many subjects your table will be very large and not particularly informative. This option will allow you to still get the stats you want and the valid number of subjects and missing cases for that variable, but without the frequency table.
- Syntax**-To paste the syntax for the descriptive stats you want left-click the “**paste**” button in the “**Frequencies**” box.
- To run analyses without pasting them to the syntax sheet, then left-click “**OK**”

Interpreting Frequency Output: Frequencies output is made up of 2 or 3 major parts, depending on whether you asked for charts.

- 1) Part one is labeled “**Statistics**”. Statistics will give you at least the Valid N and Missing N (Valid + Missing = total N). Depending on what statistics you asked for, various stats will follow the data for N.
- 2) Part two is labeled with the Variable Label. The Rows of this table are labeled with the Value Labels you defined for your variable (if it discrete, if it continuous the rows will have the scores). The Columns are labeled with the following headings: Frequency, Percent, Valid Percent, Cumulative Percent.

Frequency : this is the Absolute Frequency (e.g. the number of participants in that category or with that score.)

Percent: this is the Relative Frequency with respect to the Total N.

Valid Percent: This is the Relative Frequency with respect to the Valid N (number of participants with data for that variable).

Cumulative Percent : This Cumulative Relative Frequency (CRF) tell you the percent of observations (participants) that have accumulated up to the score of interest (e.g. it includes all the scores above it on the table)

For example: the CRF for the second row is the sum of the CRF for the first row plus the RF for the second row. Similarly the CRF for the 3rd row is the sum of the CRF of the 2nd row plus the RF for the 3rd row.

- 3) Part 3 will only be included if you requested Charts. This section will also be Titled with the Variable Label. Charts can be edited by double-left-clicking on the chart itself. A new window will open called “**Charts Editor**”, here you can make alterations to your charts. Several useful options can be accessed by selecting the “**Charts**” pull-down menu. Any changes you make to the chart in the charts editor will be reflected in your output. To return to output, either close the chart editor or use the “**windows**” pull-down menu and select the file you want to return to (e.g. the name of the output file you are working on).

VII. Compute Commands

What are Compute Statements – compute statements allow us to create new variables by transforming existing variables or to alter an existing variable by using some mathematical function. It may be as simple as subtracting 1 from every participant’s score or more complex like computing an interaction score between several variables. The most frequent use of compute statement is to create a scale score by averaging several items from a questionnaire.

Writing Compute Statements – From the “**Transform**” pull-down menu select “**compute**” and the “**compute**” box will appear. In the upper left hand corner is the “**Target Variable**” field where you designate the name of the variable that you want to create (if the name you specify already exists, the program will ask if you want to overwrite, copy over, the existing variable before it executes the compute command).

-- To the right of the “**target variable**” field is the “**Numeric Expression**” field. This is where you

write the formula for the new variable. For example if you want to reverse the direction of scores for a questionnaire item using a 7 pt. likert scale with a variable name of “q1” the numeric expression would be: $8-q1$

The **operators** for these equations are as follows:

+	=	addition	<	=	greater than	~	=	not
-	=	subtraction	>	=	less than	~=	=	not equal to
*	=	multiplication	<=	=	greater than or equal to	&	=	and
/	=	division	>=	=	less than or equal to		=	or
**	=	exponentiation	=	=	equal to	()	=	Grouping operator e.g. order of operations

– **Functions**-- The functions list provides some useful shortcuts for certain operations. (Note: brief descriptions of the functions can be accessed by right-clicking on the function in question. A small dialogue box will appear explaining what the function does)

Mean(numexpr, numexpr,)	Will average the variables (numexpr) that you include in the parentheses.
Sum(numexpr, numexpr,)	Will sum the variables that you include in the parentheses.
LG10(numexpr)	Returns the base 10 log of a variable
LN(numexpr)	Returns the natural log (base e) of a variable
SQRT(numexpr)	Returns the square root of a variable
ABS(numexpr)	Returns the absolute value of a variable

– **Conditional Compute Statements** --the “**IF**” button allows you to make the compute statement conditional. That is, you can have SPSS perform the compute statement only for cases (subjects) where certain conditions are met. For example, if you only wanted the data for male participants to be transformed you could include the IF condition ($gender = 1$).

– **Syntax** – to paste the syntax for the descriptive stats you want, left-click the “**paste**” button in the “**compute**” box.

–To run a compute statement without pasting it to the syntax sheet, left-click “**OK**”

VIII Transformations : “Normalizing” Variables

Why Normalize Variables – when a variable is found to significantly deviate from the normal distribution (i.e. it has a skewness or kurtosis with an absolute value greater than 1.00) then we need to transform the data so that it has a more normal distribution. This can be done using the compute statement (see compute statements above) appropriate to the type of problem you have. For the most part you can follow these rules:

--If positively skewed (skewed right) = use a logarithmic function of some sort (e.g. LG10 ()).

--If negatively skewed (skewed left) = use an exponential function of some sort (e.g. square it or cube it).

IX Filters : Selecting Cases / Creating Subsets:

What are Filters: Filters are kind of what they sound like, they allow you to exclude cases (participants) that you do not want included in an analysis. For example, if you wanted to know the mean IQ of only your male participants you could apply a filter that excludes the female participants and then generate the descriptive or frequency statistics for the IQ variable. Also, filters can be used to exclude certain participants based on their subject number. For example, if subjects 10 and 25 did not complete a questionnaire you could apply a filter that excludes all participants with subject

numbers equal to 10 & 25 (there would be one of each, unless you accidentally repeated a subjected number in your data entry).

Selecting Cases : the “select cases” option makes use of filters to exclude unselected cases. That is, unless cases satisfy some condition that you specify then the case will be “filtered” out. To select cases, choose the “select cases” option from the “Data” pull-down menu. A “select cases” box will appear with several options. The default option is “All Cases”, which mean all participants are included and no cases are excluded. To select a subset of cases, check the “If condition is satisfied” circle. The “IF..” box will now be active and you should left-click on it. A “Select Cases : If..” box will open. The upper-right-hand field is where you write the conditional statement that you want satisfied.

Examples:

- Males only (female = 1 & male = 2) = Gender = 2
- All subject except 10 & 25 = Subnum \neq 10 & Subnum \neq 25
- Subjects 100 and up = Subnum \geq 100 (Also, subnum > 99)
- if scale score greater than 4 = Mean(q1, q2, q3) > 4

--Once the conditional statement is specified, then left-click continue and you will return to previous “select cases” box.

--Before continuing on, be sure that the “Filtered” option in the “Unselected Cases Are” field (located at the bottom of the dialogue box) is select an not the “Deleted” option (the “deleted” option will delete the unselected cases from your data file).

--**Syntax**--to paste the select cases command syntax to the syntax sheet, left-click the “paste” button. Note: the filter will not be applied until you run the syntax or return to the “select cases” option and click “OK”.

- to apply the filter without pasting to the syntax sheet, then just left-click “OK”.

Note: If you are not sure whether filters have been activated, you can check by viewing the data in the data editor. If a filter has been applied then you will see the appropriate row markers (left hand side of spreadsheet) will have diagonal slashes through them. If no row markers have slashes through them, then you have not activated any filters. Also, be sure to deactivate filters by returning to select cases and selecting “All Cases”. There nothing more frustrating than running a bunch of analyses and discovering that you have only included half of the subjects because you left a filter on.

X Chi Square:

What is a Chi Square: A chi square is one type of nonparametric (distribution free) test. A chi square allows you test whether the frequency of occurrence for different categories of a discrete (categorical) variable (or variables) are significantly different from the frequencies expected by chance alone. That is, it tells us if there is some systematic difference in the number of people we have in each category or whether the differences can be explained by random chance. There are two types of chi squares we will focus on here: Univariate (one discrete variable) and Bivariate (two discrete variables).

Univariate Chi Square :

--From “Analyze” pull-down menu select “Nonparametric Tests” and then select “Chi-Square” from the side menu. A “chi square” box will appear. From this command box you can request several separate chi-square tests at once.

--Enter all variables that you want to be analyzed in the “Test Variable List” by either double-left-clicking on the desired variable or by selecting the variable and then left-clicking on the boxed arrow.

--The Expected Range option allows you to remove certain groups from the test. For example, if you have 6 groups but only want to test 4 of them, then you can set your range

from 2 to 5 (excluding 1 & 6).

–The Expected Values option allows you to customize your expected frequencies if you have a theoretical reason for doing so (e.g. gender distributions not really 50/50 (female/male) but rather 56/44 so if you are testing gender you may want to customize your expected frequencies to reflect this natural pattern.

–The “Options” button opens a box where you can choose between some statistics (descriptives and quartiles) and treatment of cases with missing data. “**Excluding cases test by test**” means that each chi-square test will include all participants that have data for that particular variable being tested. “**Excluding cases listwise**” means that each chi-square test will only include participants who have data for all the different variables listed in the “**Test Variable List**”.

–**Syntax**– to paste syntax to the syntax sheet, left-click the “**Paste**” button in the “**Chi-Square Test**” box.

–to run the chi-square analysis without pasting to the syntax sheet, left-click “**OK**”.

Interpreting a Univariate Chi Square Output.

– **Descriptive Statistics** – these will only be presented if you selected descriptives from the “**options**” box. They present the N(number of cases analyzed), mean and standard deviation (which is rather meaningless for categorical data), and the minimum and maximum values.

– **Contingency Table**– This table will be labeled with the variable label that you specified when you created the variable. The first column will present either the values (numbers) of the categories or the value labels (names of the categories) if you identified them when you created the variable. Column 2 shows the Observed Frequencies for each category. Column 3 shows the expected frequencies for each category. Column 4 shows the Residual (Observed - Expected).

–**Test Statistics**– This box gives you the value of Chi-Square statistic, the degrees of freedom (number of groups -1), and the Significance level achieved (anything less than .05 means that the observed frequency is significantly different from the expected frequency).

Bivariate Chi Square:

– From the “**Analyze**” pull-down menu select “**Descriptive Statistics**” and then select “**Cross Tabs**” from the side menu. From the “**Cross Tabs**” box you can perform a variety of multivariate non-parametric tests. To designate a Chi-Square test, left-click on the “**Statistics**” button. From the “statistics” box, select the chi-square option (upper left-hand corner) and then click continue. Enter the variable with the fewest number of categories in the **Columns(s)** box by left-clicking the desired variable and then left-clicking the boxed arrow pointing to the **Columns(s)** box. Enter the variable with the most categories in the **Row(s)** box, following the same procedure described above.

–**Cells**– the “**Cells**” button gives you options regarding the information that will be displayed in the contingency table. You can ask for both the “**Observed**” cell frequencies and the “**Expected**” cell frequencies. Also, you can also ask for the “**Row**”, “**Column**”, and “**Total**” percentage to be included in each cell. The “**Residuals**” (Observed Frq. - Expected Frq.) can be included as unstandardized, standardized, and adjusted standardized scores. Residuals are standardized by dividing them by a mean error estimate and they have a mean of 0 and standard deviation of 1. Adjusted Standardized Residuals are reported in standard deviation units from the mean.

Note: It is generally a good idea to ask for all of the options in the cells box.

Although it does clutter the chart some, this information can give you better idea of what is happening with your data.

- Format**- this gives you the option of determining whether your groups will be but in ascending or descending order. Either is fine.
- Syntax**- to paste syntax to the syntax sheet, left-click the “**Paste**” button in the “**Cross Tabs**” box.
- to run the chi-square analysis without pasting to the syntax sheet, left-click “**OK**”.

Interpreting a Bivariate Chi Square Output.

- Case Processing Summary**- gives you a breakdown of the number of valid cases (individuals with all data that were included in the analysis) number of missing cases (individuals missing one or more data points and were excluded).
- Contingency Table**-- This table will be labeled with the variable labels that you specified when you crated the variables. The columns and rows will be labeled with either the values (numbers) of the categories or the value labels (names of the categories) if you identified them when you created the variable. The Observed Frequencies, Expected Frequencies, Residuals (Observed - Expected), and requested percentages are displayed in each cell.
- Chi Square Tests**- The results of the Chi-Square can be obtained from the first row of this table (labeled: Pearson Chi-Square). The test statistic is located in column 2. The degrees of freedom are shown in column 3. The significance level is shown in column 4. A chi-square statistic is significant if the significance value is less than .05.

XI Correlations

What are Correlations: A correlation allows you to test the direction, magnitude, and significance of the association between two continuous variables. This procedure produces the r statistic (also called Pearson’s Product Moment Correlation Coefficient), which has an absolute value ranging between 0 and 1. A positive correlation indicates a positive relationship between variables (e.g. as one increases the other increases). A negative correlation indicates a negative relationship (e.g. as one variable increases the other variable decreases) What a correlation tells us is how much variance in a variable is shared by the variance in another variable (this is called covariance). The r statistic itself is a ratio of the Covariance divided by the Total Variance. By squaring the r statistic we get an estimate of the amount of variance in one variable that is accounted for by another variable (Coefficient of Determination).

Generating Correlations: From the “**Analyze**” pull-down menu, select “**Correlate**,” then select “**Bivariate**” from the side menu. In the “**Bivariate Correlations**” dialogue box, enter the variables to be tested in the “**Variables**” field by either double-left-clicking on the desired variable or left-click on the desired variable and then left-click the boxed arrow. You must enter at least two variables in the “Variables” field in order to preform a test. After listing your test variables select the desired correlation coefficient: in this case we want the “**Pearson**” option. You must also select the desired test of significance. In general, we will always want the “**Two Tailed**” option. Also, check the box to “**Flag Significant Correlations.**” This will make it easier to identify significant associations on the output tables.

-**Options**- left-clicking on the “**Options**” button will open the “**Options**” dialogue box with options for stats and treatment of missing data. In the stats options you can ask to include means and standard deviations and the cross products and covariances of the variables (these are not necessary). In the missing values options you can ask to exclude cases pairwise (this will exclude participants only in analyses where they do not have values for each variable being tested) or listwise (this will exclude participants for a all tests if they are missing a data point for any of the variables you have put in the Variables list).

-**Syntax**- to paste syntax to the syntax sheet, left-click the “**Paste**” button in the “**Bivariate**

Correlations” box.

–**“With” Syntax**– one option that is available from the syntax sheet that is not available from the dialogue box is the ability to include the With command. By inserting with between variables in the syntax as follows:

```
CORRELATIONS
/VARIABLES=height with pantsize weight
/PRINT=TWOTAIL NOSIG
/STATISTICS DESCRIPTIVES XPROD
/MISSING=PAIRWISE .
```

You can tell the computer to only generate 2 correlations (height with pantsize & height with weight) instead of the standard 3X3 matrix with 9 correlations (height x height, height x pantsize, height x weight, pantsize x height, pantsize x pantsize, pantsize x weight, weight x height, weight x pantsize, weight x weight). This can make interpretation of your output much easier.

–to run the chi-square analysis without pasting to the syntax sheet, left-click “OK”.

Interpreting Correlation Output

Output for Correlations produces a table called a Correlation Matrix. The format of this matrix will depend on the stats you requested and on the syntax you used to produce it (i.e. whether you used the “with” command or not).

–**Standard Matrix (no with)**-- the rows of this matrix will be made up of 3 major sections (Pearson Correlations, Significance, and N) unless you asked for stats option 2, in which case the matrix will include a section for sums of squares and cross-products and a section for covariance. The columns of the matrix will be titled with the labels (not the names) of the variables you selected.

–Pearson correlations– this section will have X^2 number of correlations (where X = number of variables requested). In the left to right diagonal you will see a series of 1.00, these are the correlations for each variable with itself, which is a perfect positive correlation. You will also notice that the correlations above (e.g. to the right of) the diagonal are identical to the correlations below the diagonal, this is because they are the same tests only with order of the variables in the equation reversed. Correlations with asterisks (*, **) by them are significant at least at the $p < .05$ level (e.g. 95% confidence level).

–Significance – more specific significance levels are reported here. The maximum significance value reported is $p < .000$, this means it is at least significant at the $p < .0001$ level. Note: $df = n-2$

–**“With” Syntax Matrix**– This matrix differs from the standard matrix in that it will not include the correlations of the variables with themselves or the correlations between the variables that appear on the same side of the with statement in the syntax. For example with the syntax “/Variables= AGE IQ With SHOESIZE BEDWETNG “ you will get the following correlations: Age x shoesize, age x bedwetng, iq x shoesize, iq x bedwetng. But you will not get: age x age, age x iq, iq x iq, shoesize x shoesize, shoesize x bedwetng, or bedwetng x bedwetng.

XI t-Tests

What is a t-Test – t-Tests allow us to determine whether two groups have significantly different averages for some continuous dependent variable. The independent variable in this case is a discrete variable made up of group membership (e.g. test group vs. control group). For example, assume that you wanted to see if drinking beer makes you burp more than not drinking beer, so you randomly assign your participants to one of two groups (Beer grp. Vs. No Beer grp.), and then you count the number of belches that occur during a one hour test period.

–**Three Different Types of t-Tests**–

Dependent Sample t-Test (the One Sample T Test)– For this type of test we are comparing a sample mean with the population mean. These types of tests are rare, because it requires that we know a population parameter. One example of the use for this test would be to compare the number of children a sample of Midwestern families have to the number of children all US families have (Population Parameter obtained from Census Bureau).

Independent Sample t-Test – For this type of test we have two sample groups for which we have averages on some continuous variable. An example of this can be seen in the Beer and Burping example given above.

Repeated Measures t-test (Paired Sample T Test) – For this type of test we again only have one sample group, but we have data drawn from two different time points. For example, if I wanted to know if taking statistics improved your overall math ability, I could test your math ability on the first day of class and then again on the last day of class and compare the class before and after averages.

–Generating a t-Test–

–**One Sample T Test**– From the “**Analyze**” pull-down menu select “**Compare Means**” and then select “**One Sample T Test**” from the side menu. This test requires you to identify one continuous variable in the “**Test Variable(s)**” Field by double-left-clicking on the desired variable(s). Also you will need to enter the Population Mean (a value that you determine, there is no variable designated for this) in the “**Test Value**” field by clicking on the field and typing in the value.

–Options– the “**options**” button will present an “**options**” dialogue box where you can change your confidence limits (something you don’t really need to worry about) and choose how you want to deal with missing data (e.g. pairwise vs. listwise, see above test options for details).

–Syntax– to paste syntax to the syntax sheet, left-click the “**Paste**” button in the “**One Sample T Test**” box.

–to run the chi-square analysis without pasting to the syntax sheet, left-click “**OK**”.

–**Independent Sample T Test**– From the “**Analyze**” pull-down menu select “**Compare Means**” and then select “**Independent Sample T Test**” from the side menu. This test requires you to identify your Continuous Dependent Variables in the “**Test Variable(s)**” field and your Discrete/Categorical Independent Variable (you can only test one grouping variable at time) in the “**Grouping Variable**” field using the appropriate boxed arrows.

–**Defining Groups**--After Identifying the grouping variable, you will need define your groups (that is you have to tell the computer what the numbers are that you used to identify your two groups of interest). Define your groups by left-clicking the “**Define Groups**” button. In the “**Define Groups**” dialogue box you will have two options:

- 1) Use specified values (which you type into the grp 1 and grp 2 fields).
- 2) Cut Point (this option is for Independent variables that are continuous, e.g. if you wanted 2 groups based on IQ, one Below Average IQ and one Above Average IQ, you could choose a cut point of 100).

–Options– the “**options**” button will present an “**options**” dialogue box where you can change your confidence limits (something you don’t really need to worry about) and choose how you want to deal with missing data (e.g. pairwise vs. listwise, see above test options for details).

–Syntax– to paste syntax to the syntax sheet, left-click the “**Paste**” button in the “**One Sample T Test**” box.

–to run the chi-square analysis without pasting to the syntax sheet, left-click “**OK**”.

–**Paired-Samples T Test**– From the “**Analyze**” pull-down menu select compare means and then select “**Paired-Samples T Test**” from the side menu. This tests requires you to identify the time 1 and time 2 variables. The first variable you select should be the pretest and the

second variable should be the post-test. When you have both variables selected, then left-click the boxed arrow to move them to the “**Paired Variables**” field.

–**Options**– the “**options**” button will present an “**options**” dialogue box where you can change your confidence limits (something you don’t really need to worry about) and choose how you want to deal with missing data (e.g. pairwise vs. listwise, see above test options for details).

–**Syntax**– to paste syntax to the syntax sheet, left-click the “**Paste**” button in the “**One Sample T Test**” box.

–to run the chi-square analysis without pasting to the syntax sheet, left-click “**OK**”.

Interpreting T Test Output

–**One Sample T Test**– There are two parts to this test’s output: 1) One-Sample Statistics (including: N (s), Mean for Variable(s), Standard Deviation(s), Standard Error of the Mean (standard deviation divided by the square root of N)). 2) One-Sample Test (including: t value, degrees of freedom (n-1), significance level, mean difference (sample mean - population mean), and confidence intervals.

–Sample means Population means are considered to be significantly different if the significance level is less than or equal to .05 (e.g. $p < .05$).

–**Independent Sample T Test**– There are three parts to this test’s output: 1) Group Statistics (including: Group N’s, Group Means, Group Standard Deviations, and Standard Error of the group Means). 2) Independent Samples Test (Levene’s Test for Equality of Variances) (including: an F statistic and the Significance Level). 3) Independent Samples Test (t-tests for Equality of Means) (including: t, df (n-2), significance, mean differences, standard error of difference, and confidence levels.

– In the “**t-test for equality of means**” table you will notice that there are two rows of statistics for each test performed: 1) Equal variances assumed. 2) Equal Variances not Assumed. The stats you use depends on whether Levine’s Test for Equality of Variances is significant or not. If Levine’s test is not significant (e.g. sig. Greater than .05), then use the statistics in the first row. If Levine’s test is significant (e.g. sig. Less than or equal to .05) then use the statistics in the second row.

--Two independent group means are considered to be significantly different if the significance level is less than or equal to .05 (e.g. $p < .05$).

–**Paired Sample T Test**– There are three parts to this tests output: 1) Paired Sample Statistics which gives stats for the pre and post-test variables (including: Means, N’s, Standard Deviations, and Standard Error of the Mean). 2) Paired Samples Correlations which gives the correlations between the pre and post-test variables (including: N, Correlation Coefficient, Significance). 3) Paired Sample Test (including: Differences between means, standard deviation of mean diffs., Standard error of mean differences, confidence intervals, t, df (n-1), and significance level).

–Time 1 and Time 2 means are considered to be significantly different if the significance level is less than or equal to .05 (e.g. $p < .05$).