The following books are available in my office. Both are excellent for learning or re-learning SPSS:


If an option is not mentioned, the default is ok (but it might not be mentioned because I overlooked it).

Also, most of this was written using earlier versions of SPSS. I updated some things I knew had changed, but there are likely to be some obsolete sections.

Please let me know about omissions or errors.

The page numbers in the table of contents may not be correct (because of additions).

"OPTIONS" TO FORMAT SPSS OUTPUT ................................................................................................. 3
PRINTED OUTPUT SETTINGS ........................................................................................................... 3
  Add A Footer With The Date And Pages ..................................................................................... 3
  Compact Printed Output ................................................................................................................ 4
COPY AND PASTE SPSS OUTPUT INTO WORD OR POWER POINT ........................................ 4
  Printed Output .............................................................................................................................. 4
SCATTER PLOT WITH REGRESSION LINE .......................................................................................... 5
  A. Scatter Plot – Single Line Per Chart .................................................................................... 5
    The Scatter Plot ....................................................................................................................... 5
    Regression Line and Correlation Coefficient .......................................................................... 5
    Replicate Scatter Plots for Groups Of Cases ......................................................................... 5
    Printed Output ......................................................................................................................... 5
    Edit the graph ....................................................................................................................... 5
  B. Scatter Plots With Two Trend Lines ..................................................................................... 6
    Add trend lines ........................................................................................................................ 6
    Edit the graph ......................................................................................................................... 6
EDITING SPSS GRAPHS ...................................................................................................................... 6
  General ..................................................................................................................................... 6
  Background (called Fill by SPSS) ............................................................................................. 6
  Titles For Graphs and Labels For the X and Y Axis ................................................................. 7
  Change the line pattern or Line Markers .................................................................................. 7
  Legend with Labels For the Trend Lines .................................................................................. 7
  Change line graphs to bar charts .............................................................................................. 7
  Tick Marks And Grid Lines ....................................................................................................... 7
  Save the Output File .................................................................................................................. 7
  And Copy Chart To Power Point Or Word .............................................................................. 7
CORRELATION .................................................................................................................................... 8
NUMBER OF DECIMAL PLACES TO USE WHEN REPORTING RESULTS .............................................. 25
DATA MANAGEMENT AND OTHER INFORMATION ........................................................................... 24
COMPARE MEANS: ............................................................................................................................. 12
LOGISTIC REGRESSION (LOGIT) ........................................................................................................ 10
SCALES ............................................................................................................................................... 15
ANALYSIS OF VARIANCE (ANOVA) AND COVARIANCE (ANCOVA) ............................................. 12
MULTIPLE REGRESSION .................................................................................................................... 9
TRANSFORMING AND COMPUTING VARIABLES .............................................................................. 20
Z, ZP, STANINE, AND STEN SCORES ............................................................................................... 22
Z-cores ............................................................................................................................................... 22
ZP Score ............................................................................................................................................. 22
Stanine and Sten Scores ...................................................................................................................... 23
Variable Names and Label Suffixes For Stanine And Sten Scores ...................................................... 23
Stanine Score Syntax ............................................................................................................................ 24
Sten Score Syntax ................................................................................................................................. 24
DATA MANAGEMENT AND OTHER INFORMATION ........................................................................ 24
Create A List Of Your Variables Using Descriptives ........................................................................... 24
Variable Names ................................................................................................................................ 25
NUMBER OF DECIMAL PLACES TO USE WHEN REPORTING RESULTS .............................................. 25
Percentages ......................................................................................................................................... 26
Other Rates ........................................................................................................................................ 26
F, t, and Chi-Square ............................................................................................................................. 26
Correlation coefficients ....................................................................................................................... 26
Odds Ratios ......................................................................................................................................... 26
P Values From Tests Of Significance ..................................................................................................... 26
Regression coefficients ......................................................................................................................... 26
"OPTIONS" TO FORMAT SPSS OUTPUT

1. Open SPSS.

2. Click on Edit in the SPSS data editor. In the pull down menu, click on “options.”

3. Click on the “General” tab.
   - In the left side, check Display Names and Alphabetical.
   - Under the heading “Session Journal,” check the append option.
   - Check Open Syntax At Startup.
   - In the right side, check No Scientific Notation and below that Regular
   - Leave the rest of the options as the default.

5. Click the Viewer tab. On the right, you will see the heading “text output page size.” Set the width at custom 100. Set the length at custom 70. Then change the “text output font” to Courier New and the font size to 8.

6. Click the Draft Viewer tab. Under the heading “page breaks between,” check procedures. Also, make sure Display Commands in logs is checked.

7. Click the Output labels tab. Change the four boxes down the column to read:
   - Names and labels
   - Values and labels
   - Names and labels
   - Values and labels

8. Click the Charts tab
   - Make the font Ariel
   - Click Lines (bottom of the page) and click Grouped Data and Continue
   - Click Markers and click Grouped Data and Continue
   - Click Apply and then OK to leave the Options dialog window

PRINTED OUTPUT SETTINGS

Add A Footer With The Date And Pages

This is essential in all output done for Straus (and I strongly advise it for everyone)

Assuming you are in the output window and want to print that output

Click on FILE and then on Page Setup
   - In the pull down menu click on LANDSCAPE
   - Click on Options
     - Delete the Page & [page] from the Header box
     - Center the Page & [page] in the Footer box
     - Click the Date icon. It will paste the data into the footer. Add a comma and a space to separate the data from the Page
     - You can also add the file name if you want, but for that to be meaningful you have to have named the output file first.

SPSS HELP15 A.DOC, 9/29/2008, Page 3
Compact Printed Output

SPSS output can go on for pages and pages. To produce a much more compact output, but with type that is small and, for some people, difficult to read

- Click on File and then on Print
- Click on Properties
- Click on Landscape and Long Edge
- Change Pages Per Sheet from 1 to 2

It is a good idea to save your output in addition to printing it. However, for this to be useful, you need to name the output file. The easiest way is to use the date as the file name. Better would be to use the date and also a name that identifies the content, such as Cross Tabs By Sex.

COPY AND PASTE SPSS OUTPUT INTO WORD OR POWER POINT

Double click the table or chart you want to copy
Right click and select Export

If a variable has many categories, say 10 or more, Cross tabs is not usually a good mode of analysis. Sometimes categories can be combined to make a more manageable number of categories. When there are many categories for the dependent variable it is usually better to run Means. Consult with me about any such variables.

- Open SPSS and then open your data file
- On upper tool bar, click on Analyze
- In drop down menu, click on Cross Tabs
- Decide on which will be the row variable. If one variable has more categories, make that the row variable because the output will fit on the page better.
- Put you pointer on the variable you want to use as the row variable, and click it into the box for Row Variable.
- Chose the column variable and click it into the dialog box.
- Click on Statistics, and choose Chi-Square
- Click on Cells. If the row variable is the one you regard as the “independent” or “cause” variable, ask for row percentages. If the column variable is the independent variable, ask for column percentages. If you are not sure, ask for both row and column percentages. (“total" percentages are rarely useful,)
- Click OK

Printed Output

- After you have done the cross tabs you want, and deleted the cross tabs that are wrong or you do not want, click on Output at the top of the left column of the output window. This should highlight all the items in that column
- Follow the instructions above on MORE COMPACT PRINTED OUTPUT

It is possible to specify more than one row and more than one column variable. This can be very efficient, but it can also result in many cross tabs you do not want. If you list three row variables and 5 column variables, SPSS will produce 3 * 5 = 15 cross tabs.
SCATTER PLOT WITH REGRESSION LINE

There are two sets of instructions, A and B. They overlap but are different in that the B set permits you to have two trend lines in the one graph, for example the same graph showing one line for males and another for females.

For both A and B see EDITING SPSS GRAPHS for information on how to change the background, the line markers, etc.

A. Scatter Plot – Single Line Per Chart

The Scatter Plot
➢ Open SPSS and then open your data file
➢ On upper tool bar, click on Graphs
➢ In drop down menu, click on Scatter
➢ Click on “Simple Scatter:
➢ Decide on which variable will be the Y axis variable. It should be the variable you regard as the “dependent” or “outcome”
➢ Put your pointer on the Y variable and click it into the box for Y variable.
➢ Decide on the X variable. It should be the variable you regard as the “independent” or “cause” variable.
➢ Chose the column variable and click it into the dialog box.
➢ If you are analyzing the International Dating Violence Study data on nations, put your pointer on WUNName and click it into the box for Label Cases By
Click OK

Regression Line and Correlation Coefficient
➢ Put pointer in the scatter plot and left double click. The graph should now be framed by blue lines.
➢ Click the icon in the toolbar for Add Fit Line at Total
➢ Click on Linear
➢ Click on the chart, and then on Close in the dialog box
Compute the correlation coefficient by taking the square root of the R squared linear coefficient

Replicate Scatter Plots for Groups Of Cases
It is often important to examine a relationship for different groups, such as males and females, or Blacks, Hispanics, and Whites. To do this, first run the scatter plot of the total sample and then use Split File to specify the groups. When you run scatter plot again after invoking Split File, it will repeat the scatter plot for as many groups as are specified by the Split File variable. For example if you use Gender of the respondent as the Split File variable, SPSS will create one scatter plot for the male cases and one for the female cases.

Printed Output
➢ After you have created the scatter plots you want, delete those that are wrong or you do not want, click on Output at the top of the left column of the output window. This should highlight all the items in that column

Save the file so that you can later copy and paste a scatter plot into your paper.

Edit the graph
This is an essential step. See the section EDITING SPSS GRAPHS
B. Scatter Plots With Two Trend Lines

These instructions permit you to have one scatter plot in which there are trend lines for sub-groups, such as males and females; whereas the previous instructions result in a separate scatter plot for each subgroup. One scatter plot with two trend lines is usually better.

1) Go to Graphs → Scatter/Dot.
2) Select Simple Scatter and click Define.
3) A dialog box titled Simple Scatter will appear.
   a) Move the appropriate independent and dependent variables into the boxes marked “X axis” and “Y axis.”
   b) If you want to graph by a group (e.g. gender), move the variable you want to group by into the box labeled “Set markers by.”
   c) If you want to label the data points, move the variable you want to use for the labels (e.g. WUNNAME) into the box labeled “Label cases by.”
   d) Click OK.
4) A scatter plot will appear in your output window.

Add trend lines
1) In output, double-click on your scatter plot, this should open the SPSS chart editor.
2) Click on Elements → Fit line at Subgroups.
3) A dialog box labeled “Properties” will appear, click close.

Edit the graph
This is an essential step. See the section EDITING SPSS GRAPHS

EDITING SPSS GRAPHS

These instructions were written to edit output from the Scatter Plot procedure. If you are editing graphs from some other procedure, such as ANOVA in General Linear Model, some details may be different. If so, I would appreciate your letting me know so that I can update this.

General
Get into the SPSS chart editor by double-clicking on the chart. This opens the SPSS chart editor.

Background (called Fill by SPSS)
SPSS uses a light shading as the default fill. For most purposes it is best to have no background color or shading. To get rid of the shading, click on the background.
1) Click anywhere in the chart area (which will be gray) to highlight it. You will know it is highlighted because a second line will appear around the chart area.
2) Go to Edit → Properties.
3) A dialog box labeled “Properties” will appear, the tab labeled “Fill & Border” should be showing, if not, select the Fill & Border tab.
   a) To change the color of the background, under “Color” click on the box next to the label “Fill” (on the left side of the “color” box), and then click on the new color from the color pallet on the right side of the “color” box. Clicking on the white box with a red slash across (in the top right of the color chart) will set the background color to transparent.
   c) When you are done, click Apply and then click Close.
Titles For Graphs and Labels For the X and Y Axis

It is convenient to use the titles and labels that SPSS automatically prints based on the names of the variables used in the analysis. However, once you paste the chart into your paper or to power point, the labels and some other parts of the chart cannot be edited in Word or Power Point. This is a problem because it is almost always necessary to make changes later on. To be able to make these changes, save the SPSS output file. This permits you to edit it again in SPSS and then export that new version to paste into your paper.

Another approach, which provides more flexibility than the SPSS chart editor, is to delete all the titles and labels before you export the chart. Then add them using the options in the Draw tool bar of Word or Power Point. Even doing that, it is best to also save the SPSS output because there will be changes that cannot be made in Word or Power Point.

Change the line pattern or Line Markers. You can change the default set of line styles and line markers by choosing a different sequence in the Charts tab of the Options window.

Legend with Labels For the Trend Lines

The default places these in a box outside the main chart area. A better arrangement is to have the legend inside the main chart area, in the upper left or right corner. You can click on the legend and then drag it to inside the chart area.

Change line graphs to bar charts

The General Linear Model ANOVA program outputs line charts. However, when the independent variable consists of categories of a nominal variable such Protestant, Catholic, Jewish, this needs to be changed to a bar chart, as follows:

- Double click on the graph
- Double click on the graph lines and highlight the lines
- Right click and scroll to the bottom of the box that comes up and choose “change to clustered bar.” If more appropriate “change to stacked bar” could also be chosen.
- Then go to the properties box to edit the size, spacing and color of the clustered bars.

Tick Marks And Grid Lines

These help the reader of your charts know the value of at any point on the trend line or bar.
1. Click on the Y axis. It will darken.
2. In the Dialog box, for Major tick marks, choose Through, For Minor tick marks choose Inside and select the number of minor tick marks you want to have between each major tick mark. Usually 10 works well. But if that crowds things too much, ask for 5
3. For grid lines, the usual patter is to have only horizontal grid lines. The default is a grid line for each major tick mark. Click on the Y axis, right click, and from the menu that appears, click on grid lines. If it produces vertical as well as horizontal grid lines, click on Undo, and try it again.

Save the Output File

It is important to save the edited output file in order to be able to make changes later. This is because charts that have been pasted into Word or Power Point cannot be revised to make changes that might later be necessary.

And Copy Chart To Power Point Or Word

To copy into power point, put the pointer into the chart area and right click. 1) Choose Copy from the menu. 2) In power point, open a new completely blank slide. 3) Put pointer in the middle of the black slide and paste the chart. You may need to first use Paste Special. 4)
Adjust the size of the chart by clicking on it and dragging one of the corners. 5) Drag the chart to center it on the slide.

For Word, charts go at the very end of the paper. Start a new page for each chart.

**CORRELATION**

There are many different measures of the degree of association between two variables. The most frequent is the Pearson correlation, also called product-moment correlation, and zero-order correlation.

**Pearson Correlation**

- In the top menu bar, click on Analyze
- In the pull-down menu, click Correlate
- Click Bivariate
- In the dialog box, check Pearson
- If you have a hypothesis about whether the correlation will be positive or negative, click on one-tailed
- Highlight the independent variable or variables and click them into the Variables box. Then highlight the dependent variable or variables and click them into the box. This method of organizing things will usually make an output that is easier to read and understand.
- Click OK

**To Produce A More Compact Output**

SPSS produces a bulky output which is difficult to export into Word and edit into a format that a journal will accept. To get a compact matrix that you do not have to retype, use the correlation part of the procedure scaling.

- In the Analyze menu, click on Scale and then on Reliability
- Highlight the variables to be in the matrix and click them into the Items box. In most cases you will be better off to click them in one-by-one so as to first click in the independent variable(s) and then the dependent variable(s). This makes for an easier to interpret output.
- Click Item Labels
- Click Statistics and choose Correlations and Continue and OK
- Double click on the matrix. It will then have a box around it. Edit the matrix to be one-line-per-variable by dragging the first vertical line from the left margin enough to the left so that each variable label takes up only one line.
- Right click the mouse and choose Export, and specify Word as the file type. It will save the matrix as a Word table in the folder that you specify.

**Partial Correlation**

Partial correlation allows you to control for (holding constant) variables you specify. It is very useful for an exploratory analysis because you can get an output that includes several dependent variables in one analysis, as compared to regression which requires you to do another analysis for each dependent variable.

The options in the following list are the ones I suggest, but they may not fit your needs.

Before starting SPSS, list (in the following order) the variables you regard as the independent variable or variables, the variables you regard as the dependent variables, and then the control
variables. Your output will be much easier to understand if you use this sequence for the variables.

- Open SPSS and Open your data file
- Click Analyze and, in the drop-down menu, click Correlate, and then choose Partial
- Highlight the independent variables and click them over into the Variables box. Then do the same for the dependent variables
- Highlight and click the control variables into the Controlling For box
- Click box for One-Tailed tests (if that is appropriate for your study)
- If you are correlating many variables, it is usually better to have the compact output
- Click on Options
  - Choose Zero-Order Correlations. This will give you both the ordinary correlations (the zero-order) and the partial correlations
  - Choose Pairwise deletion, especially if there are many variables in the matrix
- Click on Continue and then OK

It is usually important to use the compact output (see above) because that gets more of the variables on the same page where they can be more easily read, and because if there are many variables it can be a large output.

**MULTIPLE REGRESSION**

Multiple regression will tell you how much the dependent variable increases with each increase of one point of the dependent variable. The program will also add a variable to your data file that is the predicted value of the dependent variable. This can be can be used to graph the results. The graphing instructions are in the last part of the Logistic Regression section.

**A. Steps To Run OLS Regression**

The most usual form of linear regression is OLS (Ordinary Least Squares).

- Open SPSS and open your data file
- Click Analyze and, in the drop-down menu, click Regression
- Choose Linear
- Highlight the dependent variable and click it into the dependent variable box
- Highlight the main independent variable(s) and click them into the Independent(s) box
- Do the same for the independent variables that you classify as control variables
- If you want to run the regression for just one part of the cases, for example, only males, click that variable into the Selection Variable box, and then click on Rule to identify which category of the variable is for males
- Use Case Labels only if you ask for plots, and only if when the number of cases is small (perhaps less than 100) because with many cases the labels overlap and are not readable. If the study uses nations as the cases, the case label would be the name of the nation (abbreviated)
- Click on Statistics and choose Estimates, Model fit, and Descriptives (of course you can chose other statistics if you know how to interpret them)
- Click on Plots and choose Produce All Partial Plots. However, this works best for samples that are not large.
- Click on Save, under Predicted Values, click either Unstandardized or Standardized, or both. This will add a new variable to your data file with the name PRE_1. It is the probability of the dependent variable “predicted” on the basis of the independent variables. You can use this to plot your results. When there are many results, plot the
one or two that you want to emphasize in your paper. See section below on how to create the graph.

- Click on Options and choose one of the three ways of dealing with missing values. Each is useful under certain circumstances. The most usual is “listwise” but if there are many independent variables, it can result a large reduction in the number of cases used for the analysis
- Click OK to run the regression

B. Plot The Predicted Values Of The Dependent Variable

- Rename the variable(s) added to your data file (PRE_1, PRE_2 etc. to indicate what variable is. For example, if you ran a regression to determine the extent to which a set of variables “predicts” the score on the Anger Management Scale, the variable giving the predicted values will be named PRE_1. You should rename PRE_1 to be AMT1M_P and label it as “Predicted probability of Anger Management Mean Total”
- See the instructions for Graphs in part C of the Logistic Regression instructions.

LOGISTIC REGRESSION (LOGIT)

Logistic regression will tell you how much the odds of the dependent variable occurring increases with each increase of one point of the independent variable.

- Open SPSS and open your data file
- Click Analyze and, in the drop-down menu, click Regression
- Choose either Binary Logistic or Multinomial Logistic

A. Binary Logistic
Use when the dependent variable is a dichotomy, such as 0 versus 1, or 1 versus 2

- Dependent Variable: Highlight and Click into the box

- Independent Variables
Continuous independent variables: Click them into the Covariates box
Categorical independent variables:
  - Click them into the Covariates box
  - Click on Categorical
  - Click them over into the categorical variables box

NOTE: In SPSS, the category you want to use as the “reference” category with which the other categories are compared, should have the highest number. For example, if you wanted to compare African Americans and Hispanic Americans to Euro Americans, the Euro Americans should coded as category 3.

- Interaction effects: You must first create an interaction variable using COMPUTE and then include it as one of the independent variables.

- Output: The column headed Exp(B) is the odds ratio. That is usually the preferred measure of “effect size” to discuss in a paper.

- To graph the results, see the instructions in part C of this section.

B. Multinomial Logistic Regression
Use when the dependent variable has more than two categories, for example: None, Minor Only, Severe; or None, Property Crime Only, Violent Crime only, and Both. NOTE: In SPSS,
the category you want to use as the “reference” category with which the other categories are compared, should have the highest number. In both these examples, the None category should be coded as 4.

Independent Variables
Continuous independent variables: Click them into the Covariates box
Categorical independent variables: Click them into the Factors box

- Model: If you check Full Factorial, you get all possible interactions. If you have more than two independent variables, this is usually more than you want. Instead click on Custom/Stepwise. Then click on Build Terms, and click over the combination of independent variables for which you want an interaction to be calculated. Usually this will be the interactions of your main independent variable with each of the other variables, i.e., a set of two-way interactions. Three-way and four-way interactions are more difficult to interpret and are rarely used.

- Dependent Variable – The referent category must be the last (highest numbered) category. If that is not the way it is in your data set, use RECODE INTO to create a version of the variable in which it is the last category.

- Independent Variable
Continuous independent variables: Click them into the Covariates box
Categorical independent variables: Click them into the Factors box

- Interactions:
Click on Model: If you check Full Factorial, you get all possible interactions. If you have more than two independent variables, it will usually produce more interactions than you want. Instead click on Custom/Stepwise. Then click on Build Terms, and click over the combination of independent variables for which you want an interaction to be calculated.

- Saved Variables: Click on Predicted. This will create a new variable with the name PRE_1. It is the probability of the dependent variable “predicted” on the basis of the independent variables. You can use this to plot your results. When there are many results, plot the one or two that you want to emphasize in your paper. See section below on how to create the graph.

- Change the name of the PRE_1 variable to indicate what variable is it is the predicted value for, and to add a Variable Label that spells this out. For example, the predicted dependent variable is the score on the Anger Management Scale, rename PRE_1 as AMT1M_P and label it as “Predicted probability of Anger Management Mean Total”

- How to Read the Output: The column headed Exp(B) is the odds ratio. That is usually the preferred measure of “effect size” to discuss in a paper. It gives the amount by which a change of one point on the independent variable changes the dependent variable. If the odds ratio is 1.2 it means that each increase of one point increase the dependent variable 1.2 times, i.e. by 20%. If the odds ratio is .85, it means that each increase of one point decreases the dependent variable by 10% (because you are multiplying it by a fraction).

C. Graph The Results
What is usually wanted is a plot of the net effect of one of the independent variables, or a plot of the interaction of one independent variable with another (“moderator” effect).
This can be done using the program Expost. You can get a copy of the program and the instructions from me.

If you want to plot the combined effect of all the independent variables together (which is rarely desired), it can be done with SPSS by asking the logistic regression program to save the predicted probability values as a new variable. Then:

- On the tool bar, click the Graphs icon and in the pull down menu choose Line Graph.
- Click on the box with for Simple. if you are plotting the results for the total sample. If you are plotting results for different groups, such as Males and Female, click on Multiple
- Choose Summaries for Groups of Cases and click Define
- Highlight the predicted probability variable that was outputted by SPSS and click it into the box Line Represents. This will produce a trend line of the mean predicted probabilities. If you want the Median (which is better for non-normal distributions) click Change Statistic and then Median
- Highlight the independent variable for which you want to plot the effect on the dependent variable and click it into the Category Axis box. (If the independent variable is continuous, unless the sample is extremely large, the results are likely to be more clear if you first transform it into a limited number of categories such as quintiles, or deciles. See the section on TRANSFORMING AND COMPUTING VARIABLES)
- Click Ok

**COMPARE MEANS:**

**ANALYSIS OF VARIANCE (ANOVA) AND COVARIANCE (ANCOVA)**

There are many reasons for using ANOVA. The most usual is because it lets you compare differences between a variable with nominal categories, such as Black, White, Asian, and Other. In addition, I often use ANOVA because it enables me to examine means that have been adjusted for all the other independent variables, i.e. as a way to provide a view of the results, and because all relevant interaction can be tested and viewed without having to create interaction term variables and without having to plot regression lines to know what the interaction effect is.

**Means**

This should be used instead of cross tabs for dependent variables that are “continuous” ("measurement variables"), or for ordinal variables with many categories.

The program allows for more than one independent variable, but it does not compute the interaction of those variables.

- Open SPSS and then open your data file
- On upper menu bar, click on Analyze
- In drop down menu, click on Compare Means and then on Means
- Highlight the dependent variable (or variables) and click them into the Dependent List box
- Highlight the independent variable (or variables) and click them into the Independent List box
- Click on Options and:
  - Add Minimum and Maximum to the three default statistics
  - Click on both statistics (ANOVA table and Test for Linearity)
- Click Continue and then OK
NOTE: If you specify more than one independent and more than one dependent variable, it can be very efficient, but it can also result in many analyses you do not want. If you list three independent and 5 dependent variables, SPSS will produce $4 \times 5 = 20$ analyses.

**General Linear Model, Univariate ANOVA and ANCOVA**

If you want only bivariate F tests (i.e., no control variables, no interaction tests), use Means. This has the advantage of being able to test many dependent variables in one run. If you want to examine the interaction of the independent variables with each other, or you want to control for other variables, use General Linear Model.

**General Linear Model, “Univariate”** is the most frequently used method. It refers to analyses with one dependent variable. Typically there are two or more independent variables. It is probably the most frequently used type of ANOVA. If you specify covariates, it is an analysis of covariance (ANCOVA).

- Click on Analyze, then on General Linear Mode, and then on Univariate
  - Dependent variable: Highlight and click it into the Dependent Variable box.
  - Independent variable(s): click them into the Fixed Factors box
  - Covariates: click into the covariates box. These are the control variables.

- Click on Model
  - Click Custom
  - Highlight each of the variables in the left box and click each into the right box
  - Specify interactions. Highlight the pairs of variables that you want to investigate and click each pair over to the Model box. (Most researchers want only the interaction of the main independent variable with each of the other independent variables, for example: Corporal Punishment with Gender, Corporal Punishment with SES, Corporal Punishment with Race/Ethnic group. Only rarely are the interactions of the other fixed factors variables with each other asked for.)

- Contrasts: Skip this unless you want a different set of contrasts than provided by Compare Main Effects in the Options screen. If you do use Contrasts, each one chosen must also be put into effect by clicking Change. (For the IDVS study, when Nation (WUNName) is an independent variable, select the Deviation contrast. For the category to omit, check last.)

- Plots: A plot is usually very helpful, and is essential if you asked for an interaction. For an independent variable with ordinal categories, the default line plot chart is great. For nominal categories, you can use the chart editor to convert it to a bar chart. See the section on
  - Each time you specify a plot, you must also click ADD or it will not take effect.
  - To plot a main effect, click it into the Horizontal Axis and then click Add.
  - To plot an interaction, click the main independent variable into the Horizontal Box and the variable with which it interacts into Separate Lines, and click Add.
  - If the chart plots an interaction and you want to print it out in black and white, you will need to edit the trend lines to differentiate one from another and edit other things. See the section on EDITING SPSS GRAPHS.

- Options
  - Click all the variables in the left box into the “Display Means For” box on the right
  - Click Compare Main Effects. If contrast has already been asked for (see above), you may not also want this and you should uncheck it if it is checked. It produces pairwise
comparisons (each category of the independent variable with every other category). If there are many categories, or example 25 nations, it can produce a very large output).

- Click Descriptive Statistics if you want the unadjusted means for main effects and for each cell. This can also be a large output.
- Check Estimates of Effect Size. This gives you the Eta coefficient.

Output

- Use landscape output format (see section on Printed Output) to help avoid having the left half of a table on one page, and the right half on another page. This makes for a difficult to read table and also a long and expensive to print printout.

- If the left column headings occupy two or more lines, it may sometimes be helpful to make them fit on one line: To do this:
  o Locate the pointer in the table, right click to get a pull down menu with Edit Pivot Table at the bottom and click that and choose Edit.
  o Drag the vertical line separating the left column of the table to the right, just enough to make the heading fit on one line. But do no do this if it forces the right hand part of the table to be printed on the next page. Also, there is no need to do this if, regardless of the length of the line, that part of the table would require two lines.

**General Linear Models “Multivariate”** is used when you want to estimate the effect of the independent variables on two or more dependent variables. The program provides results for the effect of the independent variables on each of the dependent variables after adjusting for the overlap between the dependent variables. If there is a lot of missing data, use of this program can result in a much reduced sample.

**Analysis Of Covariance For Repeated Measures**

**Using Menus:**

“Repeated measures” refers to the dependent variable which is repeated
“Between subjects factors” refers to the independent variables

Analyze \ General Linear Model\ Repeated Measures

First dialog box :

- Within Subjects Factor Name Type in a name that covers the repeated measures
- Number of levels: Enter the number of repetitions, for example, for mothers and fathers, type in 2. For three treatments, type in 3
- Click ADD
- Click DEFINE

2nd Dialog box

- Highlight the repeated measures and click them into upper box
- Highlight the independent variables and click into the Between Subjects Factors box
- Highlight the control variables and click them into the Covariates box

Click Model

- Select Custom
- Within Subjects: click the variable into right box
- Between subjects
  - Select Main Effects: Highlight all the variables and click into right box
  - Select Interactions: Highlight the combination(s) you want and click into right box
Contrasts: Ignore except if the Compare Main Effects in Options is not what you want
Plots: Select as desired. Note you must click ADD after each one
Click Continue

Options Dialog Box
Factors: highlight all variables & click into Display Means box
Check Compare Main Effects
Check Descriptive Statistics
Check Estimates of Effect Size
Click Continue and then OK to run program

SYNTAX

The following example has two dependent variables (the repeated measures and two independent variables (the between subjects variables). The charts show the main effect of one of the independent variables and a two-way interaction.

GLM CXcpFFS CXcpMFS BY CTaUMr AXsdr2
/WSFACTOR = Parent 2 Polynomial
/METHOD = SSTYPE(3)
/PLOT = PROFILE( CTaUMr CTaUMr*Parent CTaUMr*AXsdr2 )
/EMMEANS = TABLES(OVERALL)
/EMMEANS = TABLES(CTaUMr) COMPARE ADJ(LSD)
/EMMEANS = TABLES(Parent) COMPARE ADJ(LSD)
/EMMEANS = TABLES(CTaUMr*Parent)
/EMMEANS = TABLES(AXsdr2) COMPARE ADJ(LSD)
/EMMEANS = TABLES(AXsdr2*CTaUMr)
/EMMEANS = TABLES(AXsdr2*Parent)
/EMMEANS = TABLES(AXsdr2*CTaUMr*Parent)
/PRINT = DESCRIPTIVE ETASQ
/Criteria = ALPHA(.05)
/WSDESIGN = Parent
/DESIGN = CTaUMr AXsdr2 AXsdr2*CTaUMr.

SCALES

The most frequent methods of combining several variables (also called “items” or “indicators”) to create a composite scale or index are (1) The sum of the scores for the items making up the scale. (2) Compute a factor analysis and have SPSS output a “factor score” for each factor that it finds. A limitation of these two methods is that they use “listwise deletion.” This means that if even one of the items is missing, the score for the entire scale is missing for that case. (3) A third method avoids losing cases with just one or two missing items is to compute the mean of the items for which there is data. This can be used if all the variables have the same range of scores, for example, if they are questions that all use response categories such as 1 through 4 or if you first transform all the items to Z scores.

Whenever a scale is created, in addition to the variable name, add a “variable label” of up to 40 characters. If you do not, a month or two later there is risk that you will not remember, for example what is measured by AMT1M or the difference between that and AMT2M, AMT3.

Direction of Scale Scores

It is very desirable to create scales in which a higher score means more of whatever is being measured. If the items are scored so that higher scores mean less of the variable, you
should reverse the items before combining them to create the scale. This is even more important if some of the items are scored in one direction and others in the opposite direction. If you combine items scored in different directions, the scale will be mathematically correct, but the alpha coefficient of reliability will be wrong.

**Sum of Item Scores**

This example is to compute a socioeconomic status (SES) scale by combining summing three indicators of SES: EDUC Years of education, the education, INCOM Annual income, and OCCUP Occupational prestige score. Because these three items have very different ranges of scores, the SPSS program DESCRIPTIVES was used to transform them to Z scores. The transformed variables have the same variable name, but with a prefix of Z.

Transform
   Compute
      Target Variable: SES1
      Type in ZEDUC + ZINCOM + ZOCCUP, and click Continue

This adds the SES scale score to the records for all cases. Then add the variable label. You can do this by going to the Data Editor window and use Variable View, or go to the Syntax window and use the following syntax.

VAR LABELS SES1 “Socioeconomic Status – EDU+INC+Occ”.
FREQUENCIES SES 1.

**Mean Of The Items For Which There Is Data**

If the scale is to be created by combining items that differ in the response categories or minimum and maximum values, Z score the items before combining to form the scale.

SPSS allows you to specify the number of variables that must have non-missing values. To do so, type a period and the minimum number after the function name MEAN, as in:

COMPUTE Anger scale = MEAN.3 (var1, var2, var3 var4).

The above syntax creates a scale score which is the mean of the items that are not missing, provided there are at least three of the four items with data. If there are only two items with data for a case, that case will be coded as having a missing value on this scale.

If the scale consists of the mean of items with answer categories of 1 through 4, the scale score will have a minimum possible score of 1 and a maximum possible score of 4, with most cases having scores such as 2.3, 3.5, etc.

These scores can be transformed into Z scores, ZP scores, T scores, or Stanine scores. See below.

**Factor Scores**

This example illustrates creating a socioeconomic status (SES) scale by combining variables for education, income, and occupational prestige score.

Analyze
   Data Reduction
      Factor
Choose EDUC, INCOM, and OCCUP

Descriptives:
Under Statistics choose Univariate Descriptives and Initial Solution
Under Correlation Matrix choose Coefficients and Significance Levels and click Continue
Extraction: Skip this because the default options or ok
Rotation
Verimax
Rotated solution, and Continue
Scores:
Save as Variables
Display Factor Score Coefficients, and Continue
Options
Missing values: Replace with Mean, and Continue
Click OK to run the analysis

The section of the output labeled TOTAL VARIANCE EXPLAINED gives the percent of variance in the three items that is explained by the factor.

In addition to the statistics that are printed out, the program will create and save a new variable named Fac1_1. This gives each case a score on the factor in Z score format. Rename the variable. In this example the name would be SESF1 and give it the label SES FACTOR SCORE.

If you were doing this with, for example, eight questions designed to measure past criminal behavior, the analysis might reveal two factors, one for violent crime and one for property crime. In that case, it would create two new factor score variables.

Compute FREQUENCIES. Whenever a new variable is created, it is essential to carefully inspect the frequency distribution to check for such things as: illegal values, outliers, skewness, and whether the variable seems to make sense given your knowledge of the data.

Syntax To Create An "Indicator" Version Of A Scale

Copy and paste the following syntax into the Syntax window of SPSS. Don't worry about the lines of COMMENT. Their presence will not affect the operation of SPSS.

This example is for the Criminal History (CH) scale. Replace all variables in this syntax that begin with CH with the equivalent variables for the scale you are doing.

COMMENT The count command creates a new variable with indicates the number of items that are missing for the scale. CAUTION if there is a "reversed" version of a variable (indicated by the suffix R), use that rather than the original version of the item.

COUNT CHT1MV= CH02 CH03 CH04 CH06 CH08 CH11 CH14 CH15 (SYSMIS).
VARIABLE LABEL CHT1MV 'CRIM HIST TOT- CH 2,3,4,6,8,11,14,15- # MISS.'
COMMENT The RMV command replaces missing values, in this example, with the mean of all cases in the sample.

RMV
/ch02_1=SMEAN(ch02)
/ch03_1=SMEAN(ch03)
/ch04_1=SMEAN(ch04)
COMMENT The DO IF command creates the scale only if there are 2 or fewer items with missing values that have been replaced. For longer scales, the DO IF can be 3 or fewer missing items.

COMMENT This example uses items for which the original coding is 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree. The COUNT command in this example creates a scale in which each item with a score of 3 or 4 adds a point to the scale. The scale therefore indicates how many of the eight items the respondent agreed with.

DO IF (CHT1MV LE 2).
COUNT CHT1MD= CH02_1 CH03_1 CH04_1 CH06_1 CH08_1 CH11_1 CH14_1 CH15_1 (3 THRU HI).
END IF.

VARIABLE LABEL CHT1MD 'CRIM HIST TOT- CH 2,3,4,6,8,11,14,15- #AG+SA.'

FREQUENCIES
  VARIABLES=CHT1MD
  /STATISTICS=STDDEV MEAN MEDIAN SKEWNESS SESKEW KURTOSIS SEKURT
  /HISTOGRAM NORMAL
  /ORDER= ANALYSIS.

EXECUTE.

Check that the frequency distribution is within the legal range and looks plausible.

Scale Indicating The Percent of Dichotomous Items With A Positive Code

The section will be revised as soon as I get the chance. If you want to use this method now, let me know and I will revise it immediately.

If the items are dichotomies, using the mean of variables for which data is present does not work. Imagine a scale of four items scored 1 = Yes, 0 = No. The theoretical range is 0 to 4. A respondent who omits one item, but answers Yes to the other three would have a score of 3 divided by 4 = 0.75. But he should have a score of 4 to indicate his answer was always Yes. The following might be better for dichotomous items:

Assume Scale 1 (Cognitive Neglect) has five items

This syntax will create a scale score with a legal value range of 0 to 100. 100 indicates that, for the questions the child answered, all indicated neglect. 80 means that, of the questions the child answered, 80% indicated neglect.

COUNT MIS1=ITEM1 ITEM2 ITEM3 ITEM4 ITEM5 (MISSING).
VAR LABELS MIS1 "NUMBER MISSING FOR SCALE 1."
  This results in a score ranging from 0 (no items with missing data) to 5 (all five have missing)

RECODE ITEM1 ITEM2 ITEM3 ITEM4 ITEM5 (MISSING = 9999).
  This replaces the missing data code with 9999. There will be no cases with missing data, only cases with legal values and with 9999

COMPUTE SCALE1 = 8888.
  This initializes the scale with a value that can be easily recognized if the following steps fail to replace it with a legal value for the scale (0 to 100)
IF (MIS1 GE 3) SCALE1= 9999.
This is to make the scale score a missing value if the child did not answer at least
3 of the five items. You can set some other criterion, such as at least 2
answered or at least 4 answered.
IF (MIS1 LE 2) SCALE1 = (((ITEM1+ITEM2+ITEM3+ITEM4+ITEM5)-(MIS1*9999) / (5
– MIS1)) * 100.
If the child answered at least 3 of the five items, the scale is computed on the
basis of dividing by the number of items that were answered.

MISSING VALUES SCALE1 (9999).
VAR LABELS SCALE1 ‘COGNITIVE NEGLECT – % OF ITEMS ANSWERED’.

FREQUENCIES SCALE1.

I am not sure if the parentheses are correct, and there may be other errors. This method does
work, even though my syntax may have errors. Just try it out and if there is a problem, I will
figure it out and correct the syntax.

The same procedure can be used to compute a scale on the basis of the mean of the answered
questions (which is different than replacing missing data by the mean of the sample).

NEGLECT SCALE SYNTAX TO
AVOID LOSING CASES WITH MISSING DATA ON SOME ITEMS

The following example is for Subscale 1. Assume Scale 1 has five items and is the Cognitive
Neglect subscale.

This syntax will create a scale score with a legal value range of 0 to 100. 100 indicates that, for
the questions that the child answered, all indicated neglect. 80 means that, of the questions
the child answered, 80% indicated neglect.

COUNT MIS1=ITEM1 ITEM2 ITEM3 ITEM4 ITEM5 (MISSING).
VAR LABELS MIS1 "NUMBER MISSING FOR SCALE 1."
This results in a score ranging from 0 (no items with missing data) to 5 (all five have
missing)

RECODE ITEM1 ITEM2 ITEM3 ITEM4 ITEM5 (MISSING = 9999).
This replaces the missing data code with 9999. There will be no cases with missing
data, only cases with legal values and with 9999

COMPUTE SCALE1 = 8888.
This initializes the scale with a value that can be easily recognized if the following steps
fail to replace it with a legal value for the scale (0 to 100)

IF (MIS1 GE 3) SCALE1= 9999.
This is to make the scale score a missing value if the child did not answer at least 3 of
the five items. You can set some other criterion, such as at least 2 answered or at least 4
answered.
IF (MIS1 LE 2) SCALE1 = (((ITEM1+ITEM2+ITEM3+ITEM4+ITEM5)-(MIS1*9999) / (5 - MIS1)) * 100.
If the child answered at least 3 of the five items, the scale is computed on the basis of dividing by the number of items that were answered.

MISSING VALUES SCALE1 (9999).
VAR LABELS SCALE1 'COGNITIVE NEGLECT - % OF ITEMS ANSWERED”.

**Alpha Coefficient of Reliability**

- Go to Statistics
- Click on Scale
- Click on Reliability Analysis
- Highlight the variables in the scale, and click them into the items box. Note that all items must be scored in the same direction. The scoring direction should usually be so that the higher the score, the more of whatever the scale is intended to measure. If it is not in that direction, use RECODE INTO or COMPUTE to reverse the scoring.

- Model = alpha
- Click list item labels
- Then click on statistics
- In the descriptive box, click on item, scale and scale if item deleted
- In the summaries box, click on means and correlations
- In the inter-item box, click on correlations

**TRANSFORMING AND COMPUTING VARIABLES**

**Avoid Writing Over an Existing Variable**

Every recoded or transformed variable should have a new name. This will usually be the same name as the source variable but with a suffix letter or number appended (see the section below on variable names). This preserves the source variable in its original form. You will then have both the original the recoded version. That way, if you made a mistake or you want to try a second method of recoding, you still have the original to go back to.

If RECODE is used, use RECODE INTO NEW VARIABLE rather than RECODE. Name the recoded variable with a suffix letter or number

If COMPUTE of IF is used to create a new version of an existing variable, start with:

```plaintext
COMPUTE V64R = V64, or IF(………….) V64R = ..... 
```

**Add a Variable Label**

Whenever a new variable or a new version of a variable is created, it is essential to also give it a 40 character variable label. Otherwise, a week or a month later, there is a good chance that you will not be able to figure out what it is.

See the section on what to use for variable names and labels. For transformed or recoded variables, use the same variable name and variable label as the original, but with the nature of the transformation appended, for example:

VAR LABELS V64Q “Assaults, previous year - Quintiles

**Recoding To Create New Categories or a Dichotomy**
Examples of recoding include changing a variable with six race/ethnic groups into one with four categories, such as Black, Hispanic, White, Other; and dichotomizing a four category variable into one with two categories. This is done with RECODE INTO NEW VARIABLE. See above about naming the new version of the variable.

**Adjusting Outliers**

Outliers, even when they are correct data for a case, can exert a disproportionate influence on the results. This is most likely when the N is small. A small n can occur in a large N study when the sample is partitioned into sub categories either in the data file or when interactions are tested. A small N characterizes most cross-national studies.

Tabachnick and Fidell (1989) suggest using values greater than 3.29 SD as the criterion for identifying outliers, and recoding those cases to values that corresponded with 3.29 standard deviations below or above the mean. This method uses a clear criterion and is relative easy to implement. However, it can still leave outliers in the sense a case recoded to a Z score of 3.29 could be more than a SD higher then the next highest case, and therefore still be an outlier even though not as extreme an outlier. To deal with this, I prefer to examine the frequency distribution to find cases with a value that is discontinuous, i.e. much greater than the highest preceding value. I then recode such cases to be just higher (or lower) than the highest cases which are not separated by a large gap from the preceding cases. The disadvantage of this method is that the criterion is less clear and it calls for judgments that may or may not be correct.

**Grouping Variables Into Percentiles Such As Quartiles, Quintiles, Or Deciles**

**Reasons For Grouping A Variable Into Fewer Categories.** For example, if there is a continuous variable with a score that ranges from 6 to 43, you might want to instead have categories such as the lower third, middle third, and highest scoring third. (“Third” refers to the percent of the sample, not third of possible scores). There are several reasons for grouping into fewer categories: 1. To meet a theoretical specification, for example, a threshold effect. 2. To compare with previous research. For example, research on income inequality often compares the income of the top fifth with the bottom fifth of the population. 3. To have a small enough number of categories to use in cross-tabs or as the independent variable in ANOVA. 4. To convert scores with values that are meaningless to readers into units that have more meaning. For example, a socioeconomic status scale with scores that varies from 12 to 61 can be converted into z-scores. That is fine for researchers, but is even more meaningless for everyone else. Use of the SPSS procedure Ntiles, enables you to categorize a variable into terciles, quartiles, quintiles, deciles, and therefore converts these scores to units that have a known meaning, i.e. bottom fifth, top fifth, etc. This justification for quintiles is not statistical or theoretical, but primarily as a device to aid communication of the results to readers.

- In the Data Editor screen, click on the pull-down menu for Transform
- Click on Rank
- Highlight the variable or variables you want to categorize and click them into the Variables box
- Click Rank Type and delete the check in the Rank box
- Click on Ntiles and then type in the number of categories into which you want to group the variable, such 5 for quintiles, 10 for deciles
- Click Continue and then OK
Rename and re-label the grouped variable. If you grouped a variable named SES1, the above will create a variable with same variable name as the source variable, but with the prefix N, e.g. NSES1, and with a variable label “Ntiles of SES1.”

- In the Variable View part of the Data Editor, go to the end, where you will find the NSES1 variable. Rename it to have the N as a suffix, and to indicate the number of categories, e.g. SES1N5.
- Replace the variable label with the same label as the source variable, but with “ – Quintiles” appended” or if space is short, just “5ths”

Run Frequencies to check that it is ok

Z, ZP, STANINE, AND STEN SCORES

Z-cores

DESCRIPTIVES will transform as many variables as you identify into z-scores. The z-scored variable will have the same name as the raw score, but with a prefix of Z. Therefore, to keep the z-scored version next to the original version in the list of variables, each z-scored variable will need to be renamed to put the Z as a suffix, rather than a prefix. You must also add a variable label that is the same as the original, but with a Z at the end.

ZP Score

A ZP score is a z score with an additional Transformation to create a variable with a mean of 50, a minimum of zero, and a maximum of 100. It avoids the negative numbers of a Z score and uses a scale (zero to 100) that is familiar to everyone.

The first example uses the menus to transform just one variable (X1 SES INDEX) into a normalized ZP score. If syntax rather than the menus is used, several variables can be transformed at one time. The second example uses SPSS syntax and illustrates doing several variables at a time.

Using The Menus

TRANSFORM
RANK X1
Normal

DESCRIPTIVES X1
SAVE Z SCORE OUTPUT  The output variable will probably be named ZRX1

COMPUTE X1ZP = (50+(20*ZRX1)).

RECODE X1ZP (LO THRU 0 = 0) (100 THRU HIGHEST = 100).

VAR LABELS X1ZP ‘SES INDEX – ZP’.

FREQUENCIES X1ZP.  Histogram.  Normal

Using Syntax Window (copy and paste into syntax window)

RANK
VARIABLES=ecil04p ecun05 edun01r hlci04 hlhd12p hlhd31 hlhd32 hlif06 plwb10r plwb14r plhd01p plhd02p schd01 (A) /NORMAL /RANK /PRINT=YES /TIES=MEAN /FRACTION=BLOM.

Rename vars (optional at this point
RENAME VARIABLES
(ncen04p necun05 nedun01r nhlc04 nhld12p nhld31 nhld32 nlif06 npwb10r npwb14r nplhd01p nplhd02p nschd01=ecil04pn ecun05n edun01rn hlci04n hlhd12pn hlhd31n hlhd32n hlif06n plwb10rn plwb14rn plhd01pn plhd02pn schd01n).

Calculate Z score
DESCRIPTIVES
  VARIABLES=ecil04pn ecun05n edun01rn hlci04n hlhd12pn hlhd31n hlhd32n hlif06n plwb10rn plwb14rn plhd01pn plhd02pn schd01n /SAVE /STATISTICS=MEAN STDDEV MIN MAX .

RENAME VARIABLES
(zecen04p zecun05n zedun01rn zhlc04n zhld12pn zhld31n zhld32n zhif06n zplwb10rn zplwb14rn zplhd01p zplhd02p zschd01n=ecil04pz ecun05pz edun01rz hlci04nz hlhd12pn hlhd31nz hlhd32nz hlif06nz plwb10rz plwb14rn plhd01pzplhd02pz schd01nz).

Transform into ZP
COMPUTE
ecil04zp = (50+(20*ecil04pz)).
RECODE
ecil04zp (LO THRU 0 = 0) (100 THRU HIGHEST = 100).
VAR LABELS
ecil04zp 'Strikes lockouts - ZP'.

[Repeat above for additional variables]

Stanine and Sten Scores

Stanine stands for "standard score of nine categories (Cronbach, 1970). The scores range from 1 to 9. Each identifies cases in a band of half a standard deviation from the mean. The middle category (5) identifies cases that are from one quarter of a SD below the mean to one quarter of a SD above the mean. The mean is 5.

Sten scores (Canfield, 1951) also group cases into bands of half a SD, but the Sten score uses ten categories. There is no middle category. A score of 6 starts at the mean and includes cases up to half a SD above the mean. The highest and lowest categories begin two SD from the mean and therefore may have too few cases unless the sample is large.

Normalized Stanine and Sten. Either of them can be normalized by first normalizing the original score as indicated above for ZP scores. Then Z score the normalized score, and then use the following syntax to transform into Stanine or Sten score units.

Variable Names and Label Suffixes For Stanine And Sten Scores

| Stanine | Names: S9 or just S | Labels: STA9 or just S9 if necessary |
| Sten    | Names: ST or just S | Labels: Sten or just St if necessary |
Stanine Score Syntax

RECODE [variable name] (LOWEST THRU -1.76 =1) (-1.759999 THRU -1.26 = 2) (-1.259999 THRU -.76 = 3) (-.759999 THRU -.26 = 4) (-.259999 THRU .259999 = 5) (.26 THRU .759999 = 6) (.76 THRU 1.259999 = 7) (1.26 THRU 1.759999 = 8) (1.76 THRU HIGHEST = 9).

VAR LABELS [variable name with S9 appended or just S] "........- Sta9".

Sten Score Syntax

RECODE [variable name] (Lowest thru -2.0 = 1) (-1.99999 thru -1.5 = 2) (-1.49999 thru -1.0 = 3) (-0.99999 thru -0.5 = 4) (-0.49999 thru -0.0 =5) (0.0 thru 0.49999 = 6) (0.5 thru 0.99999 = 7) (1.0 thru 1.49999 = 8) (1.5000 thru 1.99999 = 9) (2.0 thru highest = 10).

VAR LABELS [variable name] “........- Sten”.

DATA MANAGEMENT AND OTHER INFORMATION

Create A List Of Your Variables Using Descriptives

Open SPSS and open your data file
Click Toolbar for ANALYZE
Click DESCRIPTIVE STATISTICS
Click DESCRIPTIVES\ Highlight all the variables in the left box by clicking on the first one, hold down shift, and press the down arrow until you get to the end of the list. Then click arrow to move them to the right hand box
Click OPTIONS
Click ALPHABETICAL and then CONTINUE
Click OK
The variable names will be listed in many rows. If you do not see this, scroll up
Click on that and a frame will appear
Press Delete to get rid of that part of the output
Double click the DESCRIPTIVES output
Drag the vertical line just before N to the M in MAXIMUM The result will be a list of all your variables, with one variable to a line
Save the file in the folder for the project or paper you are working on:
In the Toolbar, click on FILE
Click on EXPORT
In the EXPORT FORMAT box, choose HTML and click ok
Open the file in Word.
Insert a line space wherever the topic of the variables changes
Save the file as a Word document
Print the file.
When new variables are created, if they are simple transformations or recodes, you can just pencil them in. But if there are new scales or indexes, or a lot of transformed variables, use Descriptives again to create an updated listing.

**Variable Names**

**Avoid ACRONYMS.** Acronyms are a good idea in principle, but if there are a large number of them, it is often hard to figure out what they mean. So I prefer to name the variables Q1, Q2, Q3, etc if there are not very many. If it is a large data set, then I divide it into parts A, B, C, etc to reflect the topical grouping of questions in a typical questionnaire. Then I name them A1, A2, A3, etc, and B1, B2, etc.

**Names for Computed Scales and Indexes.** Name all scales, scores, indexes with the letter X followed by a letter or letters to indicate the topic, for example F for family and C for crime. Each scale is given a number

- XF1, XF2, XF3, etc. for the three scales measuring family characteristics
- XC1 for a scale measuring crime

**Suffixes.** Z for Z score. When SPSS creates Z scores it puts the Z first, you need to rename the variable with the Z after the original variable name.

- ZP for ZP scores (but Z can be used if there number of characters will exceed 8)
- R followed by a number for recoding into categories, such as R2 for a dummy variable recode, R3 for recoding into 3 categories
- V or M for a variable that has missing data replaced
- J for a variable that has had outliers adjusted
- L for normalized
- S for Stanine or Sten scores

**Names for Variables in Panel and Trend Studies.** Either add the year to which the data refer or assign letters to each of the years covered and append the letter, e.g., Q23A for 1992, Q24B for 1993 etc. Letters are usually better because the actual year can create a variable name that is longer than eight characters, especially when suffixes for recoding are needed.

**NUMBER OF DECIMAL PLACES TO USE WHEN REPORTING RESULTS**

There are two good principles to follow, but they conflict with each other. (1) The fewer decimal places the more readily understandable the table or the text. The most readily understandable is if there are no decimals at all, i.e., integers only. (2) The more decimal places, the more precisely the results are presented. In practice, our data usually has so much measurement error and sampling error associated with it that many decimal places is a false precision. For this reason the ease of understanding criterion usually takes priority and I recommend usually using only one or two decimal places.
My guess about the reason more than one or two decimals makes a table or text less readily understandable is that the first digit or two are usually the most important. If there are numbers after that, it draws attention away from the important first two digits.

**Percentages**

Percentages are rates per 100. Percentages are the most usual way of avoiding decimal fractions, i.e. giving a rate of 24% rather than the proportion .24. Use one decimal at most. In writing for the general public, or when the N is small (under 100), use whole percentages only.

**Other Rates**

In demography, public health, and criminology, the conflict between the two principles is resolved by multiplying by a number large enough to eliminate all decimals and create an integer. Thus, rather than giving the homicide rate in a certain year as .0002 for women and .0013 for men, these numbers are multiplied by a number larger enough to make them both integers. For murder 100,000 is used, and the murder rates become 2 per 100K for women and 13 per 100K for men. It is much easier to grasp that the rate for men is more than six times higher when 2 and 13 are used compared to .0002 and .0013.

**F, t, and Chi-Square**

The convention is two or three decimals. I prefer to use only two decimals because, as explained above, it is more readily understandable by readers. Most statisticians would probably recommend three places. .. If the value is less than 1, use a leading zero.

**Correlation coefficients**

Use two decimal places except if there is a specific reason why, for the study at hand, three is needed. This will rarely be the case.

**Odds Ratios**

Odd ratios may be used for subsequent calculations, in which case three or four decimal places may be needed. But for presentation in the text of an article or in a table that readers will refer to, two is best.

**P Values From Tests Of Significance**

The convention is to NOT use leading zeros because they are all less then 1 and putting a 1 before .03 (i.e. 1.03) adds only extra text, not additional information. Two and three decimal places are both acceptable.

**Regression coefficients**

It is customary to use four decimal places. This is probably because, for standardized coefficients, there is no integer before the decimal. Therefore the significant digits begin after the decimal. In addition, regression coefficients are used to compute point estimates and regression lines, and for this reason the 2nd principle (“accuracy”) has more importance than for a percentage.