

Basic Statistics in SPSS

Office for Nursing Research
University of Washington School of Nursing
Martha J. Lentz, RN, Ph.D.
Maggie Baker, RN, Ph.C.

1999

Martha J. Lentz, RN, PhD
Noppawan Piaseu, RN, MS, MSN

2000

You've got data *everywhere!!!*

What's next?



Let SPSS help you with your data analysis.

Remember: *SPSS is your friend!*

Introduction to SPSS

Welcome to the world of SPSS. This software program can become your friend, if you're nice to it!
Today, we'll learn to:

- Open a file
- Recognize the layout of variables and cases in a file
- Distinguish between string variables and numeric variables
- Identify missing data
- Label variables and define values
- Recode 'string' variables into numeric variables
- Check reliability of a scale
- Produce statistics

This will be fun! Let's get started.....



Starting the SPSS Program

Start the SPSS program by clicking on START in the lower left corner of the computer screen.

- ⇒ Click on 'Programs'
- ⇒ Click on 'SPSS'
- ⇒ Click on 'SPSS 10.0' (not SPSS 10.0 Production Facility)
- ⇒ Click on 'File', then 'Open'.
- ⇒ Look in: Data on 'G' drive ('Welby'(G:))
 - ⇒ 'Cdata'
 - ⇒ 'NMeth521C'
 - ⇒ 'WomensHealth'

Variables/Columns

Now that you have successfully opened the file, you should see columns going from left to right across the top. This is where **VARIABLES** go – in the columns! Always set the variables up in the columns.

- Scroll to the right and look at how many variables there are (for now) in this data set.

In SPSS, variable names must have 8 or less characters, and must always begin with a letter. Notice that all of these variable names have eight or less characters *and* they begin with a letter.

Cases/Rows

Now, look at the numbers along the left border. You see rows going horizontally. This is where **CASES** (subject #s, specimen #s) go – in the rows! Always set the cases up in the rows.

- Scroll down, and you'll see that the numbers go from 1 to 659.

There are 659 cases (or participants, in this example) in this data set.



A Quick Look at the Data

Go back to the beginning of the variable list. Notice that the first 17 variables *describe* important information about the people in the study.

V1	?????
year	year data collected
yrborn	year subject born
marital	marital status
currentp	currently employed?
income	annual family income, all sources, before taxes
polparty	political party affiliation
polspec	political philosophy
howrelig	how religious?
educ	highest grade completed
ethnic	ethnicity
diffgsl	difficulty getting to sleep?

stasleep	difficulty staying asleep?
sweetday	type of sweets last week?
height	height (inches)
weight	weight (pounds)
percepwt	perception of weight
drinkdas	# days last week drank at least 1 serving of any alcoholic beverage

- Now, slowly scroll again to the right. You can do this by using the \Rightarrow arrow on the keyboard, by clicking on the arrow at the bottom right corner of the screen, or by moving the box at the bottom to the right.
- The rest of the variable names indicate answers for the Attitudes Toward Women (ATW) Scale, the Rosenberg Self-Esteem Scale, and the CES-D Scale. Some scale totals are found at the end of the variable columns.
- Return to the beginning of the variable list (**v1**).
- Now scroll down through the cases. You can do this by using the 'page down' button on the keyboard, by clicking on the down arrow at the bottom right, or by moving the box down along the right edge.

Notice that some of the cells only have a dot (period) in them. This is because data are missing for those questions. See the variable 'percepwt' for an example

Also, notice that some of these data are numerical (numbers) and some are letters. (See 'marital' and 'stayslp' for examples). Data recorded as letters are known as 'string' variable types. It's OK to record or enter data as string-type, but in order to do much of anything with it, you usually have to convert it to numeric, so....you may want to consider entering it as numeric data from the start. (We'll recode these string variables to numeric variables in a little while.)

Naming Variables

Under "name" column, let's label the first variable, which contains data about each participant's identification number. (The rest are already done for you.)

- Click twice in the first column where it says 'v1' (or click on 'variable view' button on the left side at the end of the spreadsheet)
- Under the 'Name' box, type in 'ptidno' (for patient identification number)

Now look in the first column. The first variable now has this name.

Labeling Variables

OK. So now you know about *naming* variables. When you *label* variables, you are assigning word labels to the numeric answers provided by the respondents.

Let's start labeling!

- Go to the variable 'ethnic' under the 'Name' column.
- Move the cursor to the fifth column from left and type in 'Ethnicity' under this 'Label' column
- Double-click in the box under the 'Values' column
- A box appears called 'Value Labels' (in the blue bar on top)
 - For VALUE: enter '1'
 - For VALUE LABEL: enter 'American Indian or Alaskan native'
- Click on 'Add' (The cursor should go back to the box labeled 'VALUE')

- For VALUE: enter '2'
- For VALUE LABEL: enter 'Asian or Pacific Islander'
- Click on 'Add' (We'll repeat the process)
 - For VALUE: enter '3'
 - For VALUE LABEL: enter 'Caucasian/White'
- Click on 'Add'
 - For VALUE: enter '4'
 - For VALUE LABEL: enter 'Hispanic'
- Click on 'Add'
 - For VALUE: enter '5'
 - For VALUE LABEL: enter 'Other'
- Click on 'Add'
- Click on 'OK'

Congratulations!!! You have just labeled this variable *and* assigned labels to the possible values. To see this, just click once in the little box on the right under the 'Values' column.

Recoding

As mentioned earlier, some of the answers have been entered as string-type answers (alpha characters instead of numbers). We won't be able to do much analysis with these string-type data until they are in numeric form. So...let's recode them into numeric data.

Out of the descriptive variables, you can see that two of them are string-type variables:

- marital
- stayslp

To RECODE the answers into numeric data:

Go to **TRANSFORM**

⇒ **Recode**

⇒ **into different variable**

A box appears with the title 'Recode into Different Variables'

FIRST, we'll identify the old variable ('marital'), and then name a new one ('marstat'):

- Click on 'marital' (in the variable list – note the 'A' in the box next to the variable name, indicating that you have a string variable as opposed to a numeric variable.)
- Click on the arrow (in the little box) to move 'marital' into the middle box
- Place the cursor in the 'Output Variable:Name' box
 - Type in 'marstat'
 - Click on 'Change'
- In the 'label' box, type in 'marital status'
- Click on 'Old and New Values'
 - In 'Old Value', type in 'never married'
- Hit TAB twice
 - In 'New Value', enter '1'
 - Click on 'Add'
- In 'Old Value', enter 'married or living with partner'
- Hit TAB twice
 - In 'New Value', enter '2'
 - Click on 'Add'

- In 'Old Value', enter 'divorced/separated'
- Hit TAB twice
 - In 'New Value', enter '3'
 - Click on 'Add'
- In 'Old Value', enter 'widowed'
- Hit TAB twice
 - In 'New Value', enter '4'
 - Click on 'Add'
- In 'Old Value', enter '9'
- Hit TAB twice
 - In 'New Value', click on 'system missing'
 - Click on 'Add'
- Click on 'Continue'
- Click on 'OK'

For more practice, let's recode a variable with a yes/no answer. We'll change the yes/no answer for 'Difficulty staying asleep?' into '1' for 'yes', '0' for 'no'.

Go to **TRANSFORM**

⇒ **Recode**

⇒ **into different variable**

A box appears with the title 'Recode into Different Variables'

- Click once on 'stayslp' to select it
- Click once on the arrow (in the box) to move 'stayslp' into the middle box. (It should now say 'stayslp' → ? in the middle box.)
- Place the cursor in the 'Output Variable:Name' box (this is where the name of the new variable will go)
- Type in 'stasleep'
- Click on 'change'
- In the label box, type in 'Difficulty staying asleep?'

Now, in the middle box, it should say 'stayslp' → stasleep'.

'stayslp' is positioned under 'string variable'

'stasleep' is positioned under 'output variable'

NEXT, we'll assign numeric values to the letters.

- Click on 'Old and New Values' (Your cursor should be blinking on the 'Old Value' side of the box)
- In 'Old Value', type in 'no'
- Hit TAB twice. Your cursor should be blinking on the 'New Value' side of the box.
- Enter '0'
- Click on 'Add'. In the 'Old → New' box, it should say 'no → 0'

Notice your cursor is blinking in the 'Old Value' box again.

Repeat the procedure, this time recoding 'yes' into '1'.

- Click on 'Continue'
- Click on 'OK'

- In 'Old Value', enter 'missing'
- Hit TAB twice
 - In 'New Value', click on 'system missing'
 - Click on 'Add'

Once you have finished this, you have assigned a number corresponding to each yes/no answer.



Good job!

Now, let's look at the new variables you've created!

- On your keyboard, press the 'End' key (next to the 'Delete' key).

You will go to the last variables in your file, and there are your new variables!

Descriptive Statistics

Ready for much more *interesting* moves?

The first approach to data is always to describe them – such as numbers (frequencies, mean, median, mode, skewness, kurtosis), charts or graphs.

Frequencies/Measure of Central Tendency

Let's check on how frequently answers occur in the data set, and look at measures of central tendency. We'll check on respondents' 'height' and 'weight'.

Click on **Analyze**

⇒**Descriptive statistics**

⇒**Frequencies**

From the variable list,

- click on 'height' and then click on the arrow in the box to move it over
- click on 'weight' and move it over by clicking on the arrow.

Under the '**Variables**' box, you should see both of these variables that you just selected.

- At the bottom of the box, click on '**Statistics**'.

- Under 'Central Tendency', click on 'mean, median and mode'.
- Under 'Dispersion', select 'Std. deviation'
- Under 'Distribution', select 'Skewness' and 'Kurtosis'
- Click on 'Continue'
- Click on 'OK'

A window will now fill your screen from the SPSS Viewer.

The output at the top contains the descriptive statistics you just asked for.

- Scroll down and you will see the other statistics you just requested.

What do you notice about these data? For example, what have you learned about the height and weight of the respondents in this survey?

(When you are done, minimize this window by clicking on the dash in the upper right hand corner of the screen.)

Another way to generate descriptive statistics...

There is another way to generate descriptive statistics, without the frequency tables. This option gives you a nice, neat little table of stats.

Go to **Analyze**

⇒ **Descriptive statistics**

⇒ **Descriptives**

Place the same two variables ('height', weight') in the 'Variable' box.

- Click on 'Options'
- Click on 'mean, std.deviation., minimum, maximum, kurtosis, and skewness'
- Click on 'Continue'
- Click on 'OK'

The SPSS Viewer window now appears with descriptive statistics, without the frequencies.



Visualize...

There are many ways to visually display your data. You can do charts, boxplots, pareto charts, scatterplots, histograms, and a lot of other really cool things. Let's do a *histogram* as a way to look at political party affiliation.

Go to **GRAPHS**

⇒**Histogram**

- scroll down and select '**polparty**' from the variable list
- click on the arrow in the box to move '**polparty**' over to the right
- click on '**Titles**'
- for 'Title, Line 1', type in '**Political party affiliation**'
- click on '**Continue**'
- click on '**OK**'

The **SPSS Viewer** window **magically** appears again...but now with a histogram of '**polparty**'. (If you ever need to know how to edit a chart like this, or export it to your *Power Point* presentation, or a *Word* document (*Fancy, Fancy!*), see an SPSS resource person in the Office for Nursing Research (T643). You'll receive free advice *and* a handout of your very own!

Scatterplots are an interesting way of plotting two variables against each other, to see if there is a correlation. We will look at the correlation between scales later (such as the correlation between depression and self-esteem), but for now, let's plot height against weight.

Go to **GRAPHS**

⇒**Scatter**

- Click on Simple
- Click on Define
- Select '**weight**' from the variable list; move it into the box for the Y axis
- Select '**height**' from the variable list; move it into the box for the X axis
- Click on '**OK**'

Does there seem to be a correlation between height and weight?

Reliability



Before we start drawing conclusions now from these data, let's look at the reliability of the scales that were used.

Go to Analyze

⇒Scale

⇒Reliability Analysis

First, we will check the reliability of the General Well-Being subscales. In order to do that, we need to select the variables that go with that subscale. So...

- Select the variables for Subscale 1: "General Well-Being: Anxiety" (listed below) and move them over using the (famous) little arrow in the box.
Variables: cesdp21 thru cesdp24

Once you have them all moved over...

- Click on 'Statistics'
 - select 'scale' and 'scale if item deleted'
 - select 'Tukey's test of additivity'
- Click on 'Continue'
- Click on 'OK'

The output will appear in an *SPSS Viewer* window.

Repeat this process for the other two General Well-Being subscales. The variables for each subscale are included below.

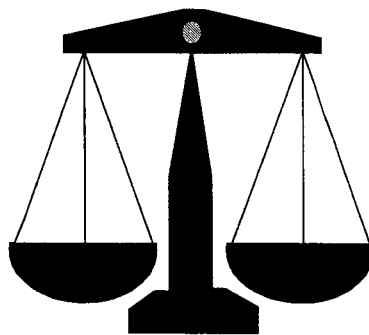
Subscale 2: General Well-Being: Positivity

Variables: cesdp 25 thru cesdp 28

Subscale 3: General Well-Being: Vitality

Variables: cesdp 29 thru cesdp 32

If you want to look at the entire scale:

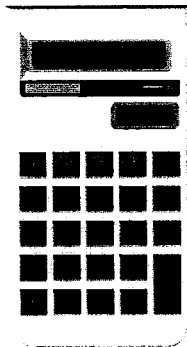


- Select all of the variables from **cesdp21 thru cesdp32** by clicking on **cesdp21** and dragging down to **cesdp32**.
- Click on the arrow in the box to move all of them over.
- Then follow the rest of the process, and you can check the overall reliability of the scale!!!

(You didn't think doing a Cronbach's alpha was this easy, did you?)

For more practice, check the overall reliability of the CES-D Scale (depression) by selecting variables **cesdp1** thru **cesdp20** and repeating the process above.

Each time you complete the process and click on 'OK', the SPSS Viewer window will appear with your answers.



The 'Compute' Command

The next step is to compute *sums (total scores)* of the scales. The ATW Scale and the Rosenberg Self-Esteem Scale totals are already done for you.

Go to **TRANSFORM**

⇒ **Compute**

- Under 'Target Variable', we'll call the 1st sum 'totcesd' for the total CES-D scale score, so type in 'totcesd'
- Place the cursor in the next box labeled 'Numeric Expression'.
- Select 'cesdp1'; move it over using the arrow
- Click on the '+' sign in the area that looks like a calculator
- Select the next variable ('cesdp2'), move it over, click on the '+' sign.
- Keep doing this until you have added (literally) all the variables that comprise subscale 1.
 - In the 'Numeric Expression' box, it should look like this: **cesdp1 + cesdp2 + cesdp3 + cesdp4 + cesdp5 + cesdp6 + cesdp7 + cesdp8 + cesdp9 + cesdp10 + cesdp11 + cesdp12 + cesdp13 + cesdp14 + cesdp15 + cesdp16 + cesdp17 + cesdp18 + cesdp19 + cesdp20**
- When it *does* look like that, click on 'OK'
- You can't see it now, but there is a new variable at the end of your variable columns, labeled 'totcesd'. (OK, you can look if you really want to!)

(Remember to click on 'Reset' if necessary to clear out labels and expressions after each operation.)

Now, repeat the process for the General Well-Being subscales. Follow the procedure above.

'Target Variable' labels and the 'Numeric Expression' contents for each are listed below.

Target Variable	Numeric Expression
gwbax	cesdp21 + cesdp22 + cesdp23 + cesdp24
gwbpos	cesdp25 + cesdp26 + cesdp27 + cesdp28
gwbvital	cesdp29 + cesdp30 + cesdp31 + cesdp32

When you are done, press the 'End' key. You will see your computed variables at the end of your variable list.

Take a deep breath. We're almost done!

T-Test



In the section above, you successfully completed subscale scores for all of the respondents. Now, let's do a t-test to see if there is any difference in the depression or general well-being scores between two groups (or types) of participants, such as those who *do not* have difficulty getting to sleep and those who *do* have difficulty.

Go to Analyze

⇒ Compare means

⇒ independent samples T-Test

- Scroll down the list of variables until you get to the variables that indicate the total scale scores you just computed.
- Select the variables which are the total scores for the ATW, Rosenberg Self-Esteem, CES-D and the General Well-Being scales/subscales.
- Select 'diffgs1' and move it over into the box for 'Define Groups'; then enter the two groups as '0' (those who do not have difficulty getting to sleep) and '1' (those who *do* have difficulty getting to sleep.)
- Click on 'Continue'
- Click on 'OK'

Data appear in the *SPSS Viewer*. The first box of data shows you 'Group Statistics'. Scroll down to see 'Independent Samples Tests'.

The Finish Line!



The last exercise is to do a **Chi-Square test**. We are testing to see if there is a significant difference between the observed and expected values for categorical data.

Let's check to see if there is a difference between those who have difficulty getting to sleep and those who have difficulty staying asleep.

Go to **Analyze**

⇒ **Descriptive statistics**
Crosstabs

- Scroll down the variable list and select **'diffgs1'**. Move this variable into the box for **'Row(s)'**
- Select **'stasleep'** and move it into the box for **'Column(s)'**
- Click on **'Statistics'**
 - Check **'Chi-Square'**
- Click on **'Continue'**
- Click on **'OK'**

In the *SPSS Viewer*, you'll see the **'Case Processing Summary'**, **'Stasleep Crosstabulation'**, and **'Chi-Square Tests'**.

You're done! Congratulations! Good job!

