What to do now that you have your completed surveys??

1. Provide each survey a unique identification number.

* Put a number on the top of each one
* Reason is because this makes it easy to find a survey at any time during the process. Say your find that you have the # 091 entered for gender when there are only three reasonable codes (0, 1, 9). By having the id number, you can go directly to that instrument and correct the error.
* Perhaps this seems silly with your project, but if you have a project with 15,000 returned surveys, it is a big deal.

2. Create a code book.

* One method is to take a blank copy of your survey. Write “Codebook” on the top. Then for each variable, you put codes for the possible responses next to the actual items. Advantage is that the codebook LOOKS like the instruments so it’s easy to refer back and forth. Disadvantage is that it can get messy.
* A second method is to create a separate document that looks something like this:

**SPSS Survey Code Book (an Example)**

Variable Label and Description Variable Name Variable Width

I.D. Number

01— ID 4

Satisfaction with Curriculum Q1 3

1 Strongly disagree

2, 3, 4, 5, 6

7 Strongly agree

99 not ascertained

Satisfaction with Instruction Q2 3

1 Strongly disagree

2, 3, 4, 5, 6

7 Strongly agree

99 not ascertained

Satisfaction with Internship and Q3 3  
Work Experience Opportunities

1 Strongly disagree

2, 3, 4, 5, 6

7 Strongly agree

99 not ascertained

Satisfaction with Financial Assistance Q4 3  
and Scholarship Opportunities and

Availability

1 Strongly disagree

2, 3, 4, 5, 6

7 Strongly agree

99 not ascertained

Satisfaction with Equipment and Facilities Q5 3

1 Strongly disagree

2, 3, 4, 5, 6

7 Strongly agree

99 not ascertained

Satisfaction with Advisement Q6 4

1 Strongly disagree

2, 3, 4, 5, 6

7 Strongly agree

99 not ascertained

Academic Status Q7 4  
1 Freshman

2 Sophomore

3 Junior

4 Senior

99 not ascertained

Self-reported GPA (A.BC) Q8 5

1.00 – just key-in the self-reported GPA (up to 2 decimal points)

99 not ascertained

Respondent's gender Q9 4

1 male

0 female

99 not ascertained

Advantages, neat looking, easy to read, Disadvantages is that it doesn’t look like the survey and in some ways requires more work.  
===============================================================

* You also identify variable names (8 digits or less is preferred)
* Add codes for missing data. 9, 99, 999. Using 9 is the convention.
  + We code this so that we can exclude this from some statistics. Plus, it is useful to know what % of an answer has missing data.
* Be sure to provide codes for Don’t Know, N/A and every possible answer.
* For variables with ordered or greater levels of measurement, make the coding make sense. That is higher numbers mean more of something and lower numbers mean less of something.
* For dummy variables, the convention is 0=absence of trait; 1=presence of trait
* Some of you will have some not applicable codes required. This is a part of skip patterns. For instance you will have a question that some folks are not supposed to answer. These also need a code. One convention I’ve seen is that the use of 8, 88, 888.
* Start with widest coding possible. You can always collapse several categories into “other” during spss session. You cannot expand other to something else easier later on though.

3. Examples

* Everyone’s first variable will be ID. The coding is simply the actual number. So survey #14 will have a code of 14.
* Gender; male; female
  + Variable= Gender or sex or Q2 or whatever. Make it make sense to you though. 0=female; 1=male; 9=missing data/refusal (when no answer was given).
  + What is your current age in years? \_\_\_\_. Use the actual number provided by the respondent as this code. See how interval/ratio level measures often take on the actual value the respondent provides?
  + Race? White, black, amer Indian, 9. Also, if you let them fill it in, they you create the actual categories and codes. Then as you look at what folks write, you decide where they go. Any difficult decisions or decision rules you use needs to be discussed in your paper. In the end, I should be able to code your surveys using your codebook in exactly the same manner as you or anyone else.
  + Opinion/attitude Very strong agree sa a d sd and vsd. Each of these categories gets a code. Make it make sense in terms of values/order. Also, if you use a matrix question, then each aspect/line/item of it is its own variable with its own code.
  + 3) I love bunny rabbits.   
    Strongly Disagree  
    Disagree  
    Agree  
    Strongly Agree
  + For select all that apply questions, each of the options is a variable and is coded as a dummy variable (0/1).
* 15) Matrix question example

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Highest | High | Medium High | Medium-Low | Low | Lowest |
| Blah blah |  |  |  |  |  |  |
| Ark ark |  |  |  |  |  |  |
| Oink oink |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Once you have your codebook done, you can start setting up SPSS to create variables and enter data.

Open SPSS software

* There are many versions out there and they all will do what you need them to do for this class. Do not panic if your screen doesn’t look like mine.
  + Search for the same function
  + Take the tutorial
  + Look in books
  + Google!
  + You cannot hurt yourself hunting around on the software.
* Please note there are three types of files for SPSS.
  + Command file where you can type code. You won’t need this.
  + Data file where the data are stored. This file has a .dat file type.
  + Output file where output is shown. This file has a .spv file type.

Open SPSS. It may take a second….

* In spss, there are multiple ways to do most anything. I will be showing you one way.
* When you open it, a dialogue box usually opens that allows you to click on the file you want if you recently used it. Either open your file or cancel past it.
* Please note at this screen some things.
  + There is a ‘data view’ tab and a ‘variable view’ tab. The default is to open to the data view tab.
  + See how it says “var” a bunch of times across the top? This is where your variable names will ultimately reside. Thus, each column represents the values for that variable you will ultimately input.
  + See the numbers down the left side? These are each case. That is, each number, each line represents each survey/respondent.

Start with the variable view tab as your first step is to create your variables. This is done here. You enter data after your variables are created and that is done on the data view tab.

* + Start with the first variable: ID
    - Fill in the variable name (e.g., ID)
    - Type refers to type. In our work, they are all numeric which is the default. You can enter words here but we aren’t doing that.
    - Skip Width. By default you have enough width. You could make this much longer or shorter if you want.
    - The “Variable label” box lets you write a longer description of the variable name. That is, you could say “q3, opinion about blah blah blah”. Fill this in to make it understandable to you. \*please note that a variable label and a value label are different things!
    - Next you need to fill in the value labels on the “Values” column. This is a label for each numeric code. Next you associate your codes with what they mean.
    - BE SURE TO HIT ADD for each one.
    - BE SURE TO PUT IN YOUR N/A codes and your missing data codes here too. Every code should have a label.
    - Next click on missing values. Here you can input up to three discrete missing data codes (9,8). Or you can put in a range. Or both. You must tell SPSS which numeric codes are to be treated as missing data. Giving it a label is not enough. Hit OK when done.
    - Define your Measure – in terms of level of measurement. (optional).
    - Your variable now appears on the data screen right?
    - Go through and add all your variables and define them. It’s a wee bit tedious, but it will go faster than it likely seems it will right now.
    - Note that you can now copy/paste value labels! Lucky you!
  + Next you need to input all the data from your surveys. Go to “Data View” to do this.
  + Once done, you’ll have a sheet full of numbers showing on the data view tab.
    - Please note that at any time you can go in and add more variables, you can check on the definition of any variable, and you can add more cases/respondents.
  + NEXT, check your data. Best way to do this is to get some descriptive statistics for every single variable.
    - Analyze
    - Descriptives
    - Frequencies (offers lots of stats such as mean, median, mode, standard deviation, variance, etc).
    - Select variable of interest (or highlight all you want) and click on the little arrow key to move them over to the “variable” box.
    - Click on “statistics” to choose stats you want for these variables.
      * New box opens giving you the choice of what statistics you’d like to see for each of these.
      * Go for it! Mean, median, mode, std. dev, min, max variance. We will discuss how some of these are used only for particular levels of measurement next week.
      * Hit ‘continue’
      * Now hit “ok”
      * A new screen with results shows up.
        + The text at the top is what we used to before the windows version came out - type to write code for spss. You can ignore this, but note it tells you what you just did to get the output.
        + Note the information it provides you.
        + N, missing, mean, etc. for each variable.
        + Look at these findings. Are there weird things going on? Are there two surveys with the # 45? If so, address that in your data file. This can easily be rectified because you have ID numbers on your surveys. Do you have a # instead of a label for an outcome? If so, this needs to be dealt with.
      * You’ve been asked to provide a Descriptives table in your paper that describes your sample. This is where you get this information. By descriptives I mean measures of central tendency, dispersion, percentages etc.
        + You can copy and paste these boxes out of spss and into excel. In excel it is easier to make the tables pretty.
        + And a key thing I’ll be looking for is that you provide the appropriate stats for each variable. Note that spss will give you a mean on nominal level stuff though it is nonsense. It is up to you to NOT provide nonsense to me. It demonstrates your understanding of this which will be covered in an upcoming class and in the reading.
        + MAKE TABLES ATTRACTIVE. COPY AND PASTE IS NOT ENOUGH.
      * Note that you have two windows in SPSS open now. An output window and the data window. You can save both. BE SURE TO SAVE YOUR DATA OR IT WILL VAPORIZE AND YOU WILL BE DISTRAUGHT.
      * It is up to you if you want to save your output. If you don’t then you have to rerun it again if you need it again. Note that you will be turning in your data file and our output file as appendices in your paper.
* You may want to (later) continue with any further analyses on this screen too. Don’t be afraid to play around.
* For instance, want to see your IV and DV in a table? Go to statistics, summarize, crosstabs. Put your DV in for the row; and your IV and others in the column. Next, go to and click on ‘column percentages’.
* For instance, you want to see if your variables are correlated?
  + Analyze
  + Correlate
  + Bivariate
  + Chose variables, arrow over to variables box, click okay.

In some cases, you will want to **create a new variable** based on values of existing variables. For instance, if you asked 5 questions to measure some underlying concept, you may look at responses for each of the five questions, and you may want one score that applies to all five scores. An easy way to do this is to add the score on all five variables to get that ‘summary value’.

To do this:

* When in the data view, hit ‘transform’
* hit ‘compute variable’.
* In the ‘target variable’ type in the new variable name. E.g., “sumscore”.
* Click on ‘type and label’ and fill in this information. A label for that variable and the level of measurement.
* Hit continue.
* Next move var1 over and type + then move var2 over and type + etc. (or whatever you want to do, e.g., ‘salary/12’; kids\*2).
* Hit ‘ok’. If you look at the end of your data, you’ll see your new variable.
* If adding, make sure you have no reverse coding issues.
* Be sure to save your data or your new variable will be lost forever!

REVERSE CODING. A research often writes multiple questions to measures an underlying abstract concept. For example, one could ask a single question “how satisfied with your job are you?”. Or a researcher may ask five questions designed to measure that concept: relationship with their boss and co-workers, the amount of vacation time they receive, the amount of any bonuses they receive, how well they like their health plans, etc. Asking several questions allows one to better measure a concept generally.

One can get statistics on each of these separate questions. But one may also want to get the summary score (described above). Adding the scores across questions is a common way to do this. Although this process works well, **you need to be careful of any reverse-coded questions that might appear in your survey**. Reverse-coding occurs when a question is phrased in a way that is opposite to most of the other questions in a survey. For example, imagine assessing job satisfaction on a 6-point rating scale with the endpoints, *1 = Strongly Disagree* to 6 *= Strongly Agree*. You might have items like these:

|  |
| --- |
| The bonuses I receive are fair. We have good company health plan. I have a good relationship with my co-workers. I do not have a good relationship with my boss. |

This last item is reverse-coded because a response of *6 = Strongly Agree* indicates **low** satisfaction, while with the other items, a response of 6 *= Strongly Agree* indicates **high** satisfaction.

As we noted in lecture and in the readings, reverse-coded items are generally used to encourage/ensure respondents to actually read survey questions and not just provide the same response to each question.

In the example above, however, one cannot simply tally the scores on all the items to compute an overall job satisfaction score because of the reverse coding. First, you need to **recode** the reverse-coded item.

To recode a reverse-coded item in SPSS,

* click on Transform
* click on Recode into Different Variable.
* Name the new variable you will create. This is under the ‘output variable’ and ‘name’.
* Click on ‘old and new values’. “Old Value” and specify all of the original values for the variable, and go to “New Value” and specify all of the new values for the variable you are creating.

After you've created the new variable, you can add up the scores on all of the items to produce your overall score (as described above).