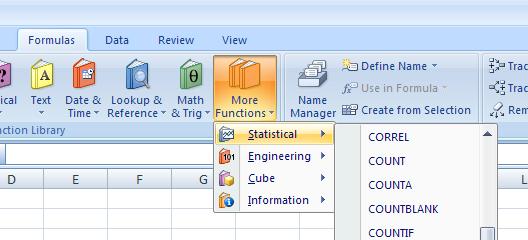
**8 Handy Things About MS Excel**

Most graduate students are familiar with the basic features of MS Excel – adding/averaging groups of numbers, using formulas, etc. In psychological measurement research, Excel is useful for many other things, such as performing basic calculations and making output from other programs (e.g., SPSS, Mplus) easier to read and interpret. Through years of trial, error, and Google, I’ve discovered a number of tricks and techniques that make Excel easier and more useful. The following are things I use all the time when dealing with measurement data:

1. **(Slightly) More Advanced Statistics**

You are probably familiar with performing basic calculations (+, -, \*, /) in Excel, but did you know that Excel will also calculate other statistics for you? On the “Formulas” tab, click “More Functions,” then “Statistical” for a long list of statistical calculations Excel will perform for you.



“CORREL,” “COVAR,” and “VAR” are a few that are invaluable for this course. If you can’t tell what a formula is from the name, point the mouse at the name and it will give you more of a description. There’s no need to reconstruct these formulas by hand – it’s just tedious and introduces lots more possibilities for little errors.

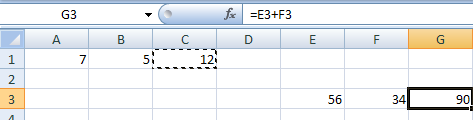
1. **Copying Formulas**

If you do need to create a formula from scratch (e.g., calculating a factor loading), you don’t need to retype it every time you want to use it (again, creating extra potential for errors). Instead, you can change the data the formula refers to.

To do this, it is important to understand how Excel “thinks.” When you refer to another cell in a formula (e.g., “ = A1+B1”), Excel keeps track of where those cells are located **in relation to the cell you’re typing in**. In other words, if you’ve typed this formula in cell C2, Excel remembers that you want to add the two cells to the left of the cell the formula is in.



If you copy and paste this formula in the normal way, then Excel will add the two cells to the right of the cell you just pasted the formula in:



See? When I copy and paste the formula, it automatically changes to E3+F3.

Sometimes you don’t want Excel to do this – sometimes you want to create several formulas that refer to the same set of data. You can “lock” the cell you are referring to by using the $ sign.

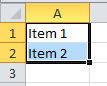
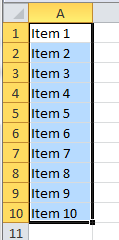
* Typing “=$A$1+$B$2” means that wherever you copy and paste the formula, it will ALWAYS refer to cells A1 and B1.
* If you type “=$A1+$B2,” the formula will always refer to a cell in column A and a cell in column B, but the row will change automatically according to the row your formula is in. This may or may not be in the same row as the formula – if you originally typed the formula two rows lower than the data, the new formula will still refer to data two rows above.
* If you type “=A$1+B$2,” the same as the above will happen, but with reference to rows instead of columns.

1. **Sequences**

This is a small trick, but a remarkably handy one. Sometimes you want to enter the same value several times, or fill in a sequence of data (for example, numbering items or participants). Instead of typing in every value, you can have Excel do this for you. You have to get the sequence started – enter at least the first two numbers (or the same number twice, if that’s what you want) in the appropriate row or column. If you have a more sophisticated sequence (e.g., 5, 10, 15), you may need to enter a little more of it to get the right results.

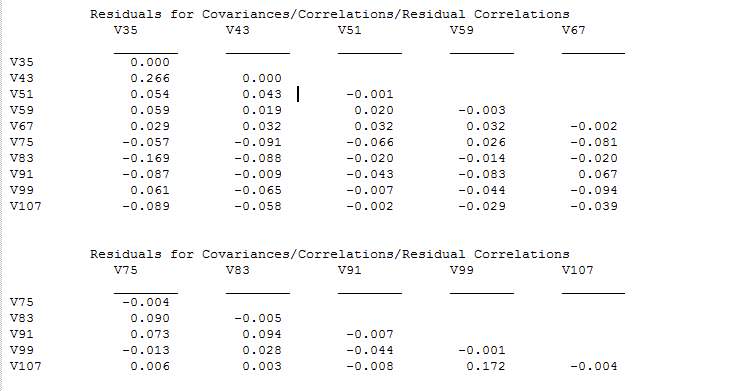
Then highlight both cells and point your mouse at the lower right-hand corner of the bottom-most or right-most cell. When it turns into a black plus sign (not the open white plus sign, and not the four-headed arrow), click and drag as far down or to the right as you want the sequence to extend.

This can also work with numbers that are part of text (e.g., “Item 1”), as long as you get the sequence started. Excel also knows the days of the week and the months of the year.



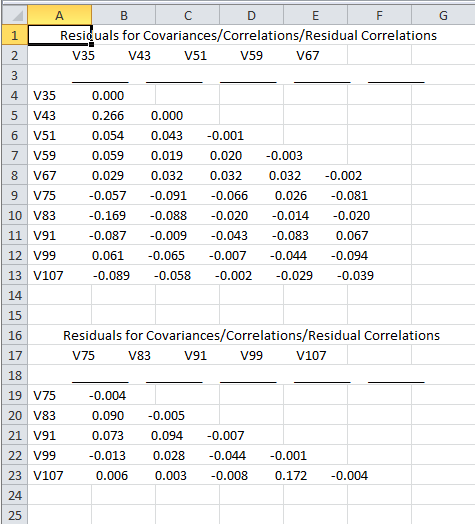
1. **Text to Columns**

One reason I use Excel in measurement analyses is because it makes data easier to look at. The output from Mplus and R is, frankly, not very pretty. It can be difficult to read a correlation matrix or other data from these programs, because the fixed width structure of the output will break up a large matrix. For example, here I want to look at the residual matrix for a CFA with ten items. Mplus will only show 5 columns of data at a time, so the last 5 columns of the matrix are printed below the first 5 columns:

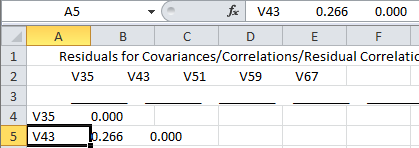


This is a little annoying for 10 items. It is a royal pain with any more than that. Also, if I wanted to perform any kind of additional calculations on these values, I would have to copy each value by hand. That introduces a lot of potential for error.

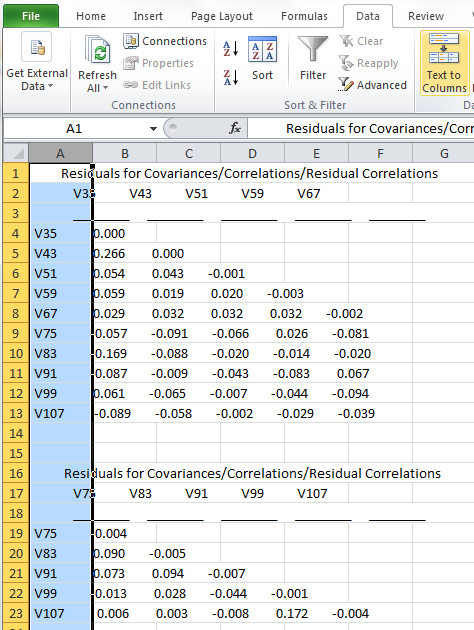
So I often use Excel to view results like this. If you just copy and paste directly into Excel, you get this:



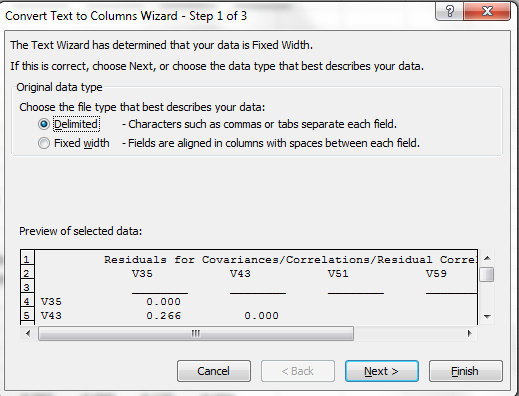
That doesn’t look much better, does it? In fact, if you try to click on an individual value, you’ll see that each entire line of data has been pasted into a single cell:



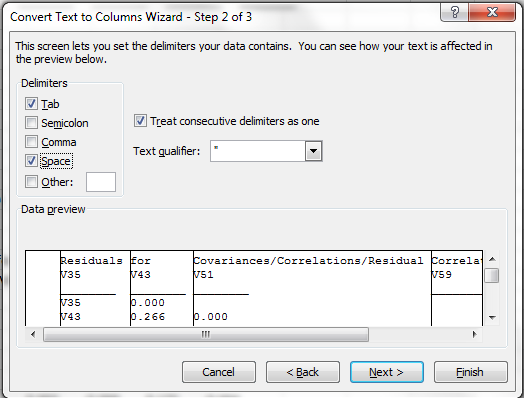
So we need a way to make Excel understand that this is a table. First, highlight the column that the data has been pasted into. Then, go to the “Data” tab and click on “Text to Columns.”



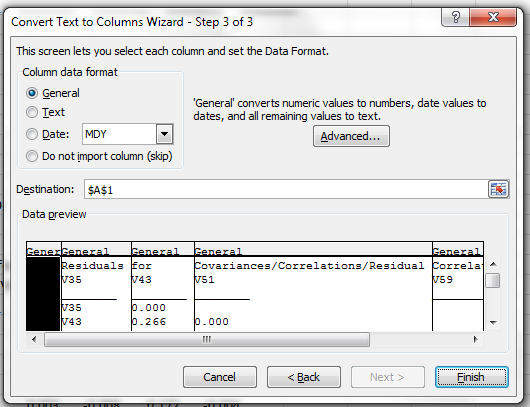
This brings up a dialog box that looks like this:



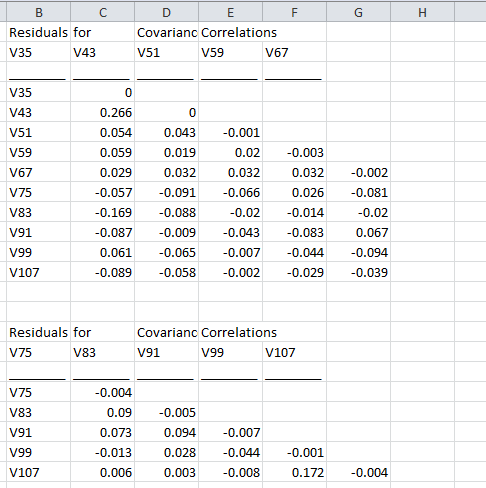
The default setting is “fixed width,” but I usually find it easier to choose “Delimited.”



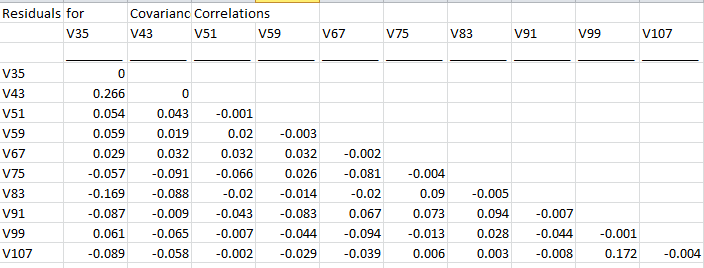
Now you have to tell Excel *how* your data is delimited. “Space” is usually a good choice for Mplus files. Note that the window at the bottom shows you what your data will look like once the text-to-columns process is done – the vertical black lines show you where the columns will be divided. Check this to make sure it’s what you want. The next screen looks like this:



If you want to, you can set the format for each individual column here. Frankly, I think it’s easier to format them once they’re in the spreadsheet, so I usually skip this step and just hit “Finish.” Now my data are in columns:



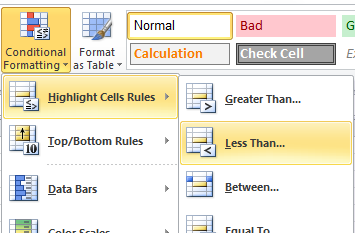
I can then cut and paste the bottom section of the matrix next to the top section. Note that the variable labels at the top of the matrix are now all shifted one column to the left of where they ought to be, but this is easy enough to fix by either “Insert Cells” or just cutting and pasting. Voila:



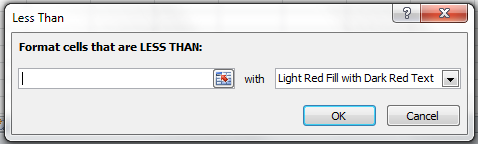
1. **Conditional Formatting**

When we are dealing with complex data, it’s incredibly helpful to be able to spot patterns and trends quickly. Excel’s conditional formatting feature can help you do this by changing the format of individual cells based on criteria you specify. For example, I often use this feature to quickly identify negative correlations between items.

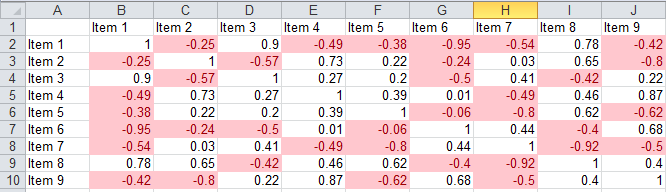
Conditional formatting is right on the “Home” toolbar. To highlight all of the negative correlations in a matrix, first, select all of the cells in the matrix. Then click “Conditional Formatting,” “Highlight Cells Rules” and then “Less Than.”



In the dialog box that pops up, type “0” and choose the way you want the formatted cells to appear.

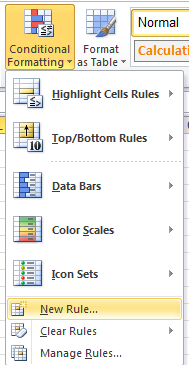


You then get something that looks like this:

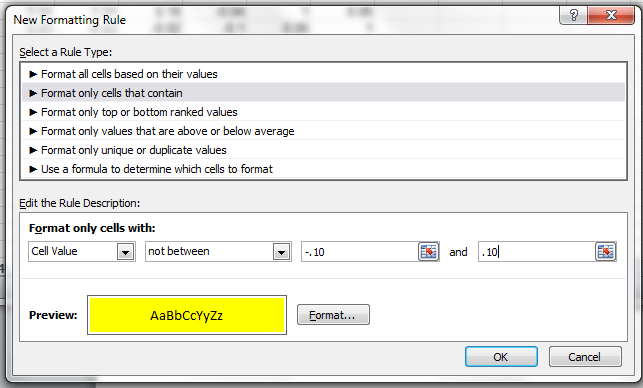


Now you can spot negative correlations at a glance. This may not seem like a big deal… until you have a correlation matrix for 50 items… !

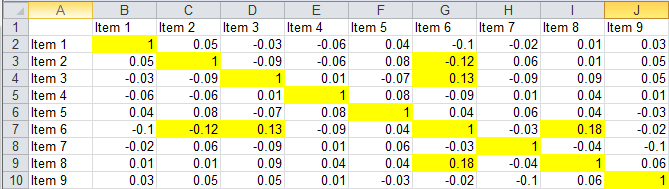
You can also use conditional formatting with more complex rules. For example, later on we’ll talk about looking at residuals with absolute values above .10. That means we want to see both values > .10 AND values < - .10 at the same time. To do that, you’ll use a slightly different approach. Instead of choosing “Highlight Cells Rules,” go to the bottom of the menu and click on “New Rule.”



Now you get a different dialog box with lots of choices:



In this example, I’ve told Excel to find cells that are NOT between -.10 and .10 and to format them with a yellow fill in the background. This gives me:

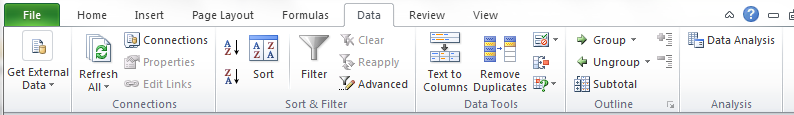


This makes it very easy to identify the handful of large discrepancies. Of course, you can use conditional formatting for all kinds of other things too.

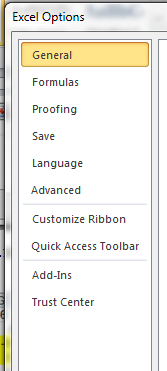
Note that the most common errors in conditional formatting are (a) not selecting (all of) the cells you wish to format before you start the conditional formatting process and (b) not specifying a format in the “New Formatting Rule” dialog box. If you run conditional formatting and nothing changes, one of these might be the reason why.

1. **The Analysis ToolPak**

This nifty thing is called an Excel add-in. It is already part of your Excel software – you don’t have to buy or download anything, this is a perfectly legitimate Microsoft product – but it hides until you tell Excel you need it. Once you install it, it will always be available on the “Data” section of the toolbar:

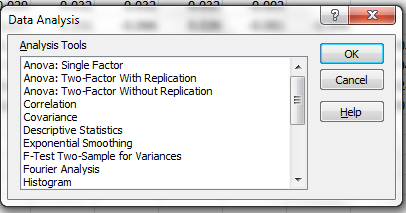


If you don’t see anything on that part of your toolbar, that means you need to install the ToolPak. Go to the “File” menu and click “Options.” That brings up a dialog box with a menu on the left:



Click “Add-Ins.” Select “Analysis TookPak” from the list of available add-ins and click “OK.” Excel will install the add-in, and when it’s done you can go check your toolbar again.

So what does it do? Why would you want to install it? It allows you to perform a bunch of basic statistical functions & tests with your data that are not available in plain ordinary Excel. For example:



You can use this feature to get a whole correlation or covariance matrix in one step. Choose “Correlation” from the dialog box, then select ALL of the variables you want to correlate in the “Input Range.”

It’s not as sophisticated as SPSS (much less R), and there is some concern about whether some of the calculations are exactly correct, but it can come in handy when you’re just exploring and checking the structure of your data.

1. **Creating Random Data**

This isn’t really specific to measurement, but I often find it useful to be able to create random data. Excel has a number of functions for this.

=RAND() (with nothing in the parentheses) will give you a random(ish) number between 0 and 1.

=RANDBETWEEN(x, y) will give you a random integer between the two values you specified in the parentheses (sub in any numbers you want for x and y).

The tricky thing about these formulas is that they aren’t stable – Excel will pick a new random number *every single time you make any kind of a change to the workbook.* So if you’re trying to perform calculations on the data (i.e., to test out an analysis), your results will change at every step. To avoid that, after you’ve generated your random data, select all of it and choose “Copy” and then “Paste Values.” That will save the current set of values so that you can use it in future calculations.

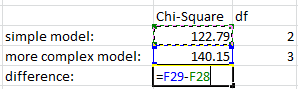
**8. Quick Statistical Tests**

Excel contains functions for most common statistical distributions – *F*, *t*, *Χ2*, etc. This comes in handy for comparing nested models using the Δ*Χ2* statistic (that is, the difference between the chi-square values of two different models).

To set this up, enter the two chi-square values and their accompanying degrees of freedom:

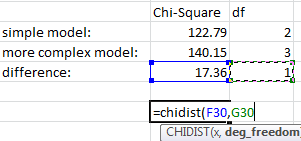


Then use a formula to find the difference between the chi-squares:



And repeat for the degrees of freedom (remember, you can just drag that formula over to the next cell). Remember to subtract the smaller value from the bigger value so that the result is positive.

Then, in a new cell, type the formula “=CHIDIST(“. Click on your Δ*Χ2* value first, then your Δ*df*, like this:



When you close the parentheses, the result of the formula will be the *p*-­value of this Δ*Χ2* in the chi-squared distribution. A value < .05 means the two models are significantly different in fit (usually implying that the more complex model should be preferred). You can do this for t-tests, etc., just as easily, but these are the tests we’ll most often be concerned with in measurement.

**Lab Exercise:**

Create, analyze, and format some data. Follow these steps, in order:

1. Generate random data for 100 participants on 10 items, using a Likert-type scale from 1-5.
2. Use sequences to create participant IDs and to label the items “Item 1”, “Item 2”, etc.
3. Create a column titled “Mean Score” and use a formula to find the mean score across all items for each participant.
4. Use a formula to find the overall mean score across all participants, and another to find the standard deviation.
5. Create another column titled “Z-score.” Using a formula, subtract the overall mean from each participant’s mean score, then divide by the standard deviation to find each participant’s Z-score.
6. Find the correlation matrix for all 10 items.
7. Use conditional formatting to highlight any correlations with an absolute value greater than .20.

Save your entire workbook and turn it in via Canvas.