Excel: Sort and Filter

In this handout, we will consider the North Carolina Birth data. A portion of this data set is shown below.



Before we discuss sorting and filtering, let’s identify some features of this spreadsheet that you should always keep in mind when creating a database in Excel.

* The column headings (known as field names) identify the variables in the data set, and each observation (or record) is placed in its own row.
* The column heading should reside in only one cell.
* Do not leave a blank row between the column headers and the first record, and be sure that there are no blank rows between records (when data points are missing, it is okay to have some blank cells).

**SORTING DATA IN EXCEL**

Open the NC Birth dataset. Suppose that you want to sort the data by *Mother Age*. Sorting the data will change the order in which the records are displayed, but each row will be kept together so that the data is correctly maintained.

To sort based on this column, select “Sort” from the Data ribbon.



The following dialog box appears. Tell Excel to Sort by *Mother Age* by selecting this variable from the drop-down menu list, and also note that you are asked whether you wish to sort in an ascending or descending order. Here, we will sort from the youngest to the oldest mothers.



Click OK, and Excel returns the data sorted by the age of the mother (a portion is shown below).



Note that you can sort by more than one variable. For example, suppose that you wanted to also sort by *Father Age*. Once again, select “Sort” from the Data ribbon. In the upper-left-hand corner of the dialog box that appears, click “Add Level” and then choose the following:



The data should now appear as follows:



**Custom Sorting**

When you sort data with Excel, it is done by default alphabetically, numerically, or by date. Sometimes, you may have items in your list that you would like sorted in a different way.

For example, let’s start with the original NC Birth dataset (i.e., before sorting by mother’s age or father’s age). Suppose you want to sort by the variable *Kotelchuk Index for Adequacy of Prenatal Care.* Furthermore, you want to sort the variables in this order: Inadequate, Intermediate, Adequate, AdequatePlus, and NotReported.

From the Sort dialog box, select “Custom List.”



Enter the following in the List Entries section of the Custom Lists box:



Finally, click “Add” to add this to your custom lists, select it, and then the data should sort in the order you have requested.

**FILTERING DATA IN EXCEL**

A filter allows you to quickly and easily subset your data. In other words, you can easily view only those records that meet certain criteria. To invoke a filter in Excel, highlight all of the variables (columns) in your dataset and select “Filter” from the Data ribbon.



Each variable in your dataset now has a drop-down arrow, indicating that the filter has been activated. This drop-down arrow can be used to apply a filter for that variable.



For example, if you’d like to observe only those observations for which the father was Nonwhite, click on the *Father Minority* drop-down arrow and select only “Nonwhite.” This is shown here.



Note that once a filter has been applied, the drop-down arrow changes to a “funnel.” Also, note that several rows are missing and the row labels have changed color.



You can easily apply a filter on multiple variables. Each additional variable you add will narrow down your list, as records have to meet each set of criteria in order to be displayed. For example, consider the subset *Father Minority* = Nonwhite and *Mother Minority* = Nonwhite.



Notice the funnel icon on both *Father Minority* and *Mother Minority*.



To **remove** a filter, simply select “(Select All)” from the drop-down arrow for that variable. To deactivate the filters for all columns, click “Filter” once again from the Data ribbon. To deactivate a filter for just one column, click on the Filter button for that column and select “Clear Filter.”

**Filtering Numbers**

Unfortunately, Excel does not automatically differentiate between categorical data and numerical data. For example, suppose we want to see only the observations for which *Mother Age* > 40. When the drop-down arrow for *Mother Age* is selected, you obtain a selection box for each observation.



Excel does, however, provide a special menu for Number Filters. There are several options under the Number Filters submenu. If the intent is to subset based on Mother Age > 40, select “Greater Than…” from this submenu.



In the Custom AutoFilter window, specify that you want *Mother Age* > 40. This is shown below.



A small portion of this filter’s result is given here.



**More Custom Filters**

There are several filters that can be applied to text variables, as well.



For example, suppose we’d like to display all observations for which *Kotelchuck Index* = Inadequate, Intermediate, or Adequate. The easiest way to do this is to simply check the boxes for these three categories.

Also note, however, that each of these variable **ends with “e”** and that this rule can be used to create a custom filter.



After specifying this in the Custom AutoFilter box as shown above, you should obtain output similar to what is presented below.



**Using Wildcards When Filtering**

You can use a wildcard such as the asterisk (\*) in a filter to represent any character. For example, suppose you want to subset those observations with a *Kotelchuck Index* of “Adequate” or “AdequatePlus.” Once again, this could be accomplished by checking the box next to both of these categories from the drop-down arrow. However, we can also do this with a Custom Filter using wildcards as follows:



A portion of the records displayed by Excel is given below.



**Using Advanced Filters**

In some instances, the AutoFilter can’t be used to obtain the desired result. For example, suppose you wanted to view all records where either *Mother Age* > 40 **or** *Father Age > 40.*

In this case, you would have to use an Advanced Filter. In a blank part of the spreadsheet, type in the exact column headings that appear in the original dataset. Then, type in the desired criteria below these variable names. For this example, the following should be entered:



Note that when the criteria are listed on separate lines such as they are above, this indicates that we desire one **or** the other criterion to be true.

Next, click on the “Advanced” option in the Sort & Filter menu.



Enter the following in the Advanced Filter dialog box.



A portion of the filtered data is shown below.



Note that we when the criteria are listed on the same line, this indicates that we desire both criteria to be true. For example, suppose we had started with the following:



This tells Excel to include records where *Mother Age* > 40 **and** *Father Age* > 40. Carry out this Advanced Filter in Excel and verify that this is the case.