Using Excel’s PivotTable to look at
Bivariate Categorical Data

This handout briefly discusses MS Excel’s PivotTable routine. PivotTables are very
powerful tools to sort through data and make different summaries. They will prove
useful for other than statistical analyses as well.
Consider the Housing Data. Let’s suppose that we would like to examine whether
basements are more common in Cape Cod, two-story or Ranch houses (both variables
are clearly categorical). Our end result could look like the following contingency table:

<table>
<thead>
<tr>
<th>Count of basement</th>
<th>Basement</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Style</td>
<td>0</td>
<td>1</td>
<td>Grand</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>14</td>
<td>25</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>25</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>41</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Grand Total</td>
<td>17</td>
<td>91</td>
<td>108</td>
<td></td>
</tr>
</tbody>
</table>

We could interpret the results as follows. Of all 17 houses without basements, 14 (over
82%) were in Cape Cod houses, as compared to only a little over 27% of the houses
with basements in Cape Cod style. So, a Cape Cod style house is more likely to have
no basement than having a basement. In contrast, all of the two-story homes had a
basement. If the presence or absence of basements would be independent of the house
style, we would roughly see 1/3 of the houses without (and with) basements for each
house style, which is clearly not the case here. So, in conclusion, we can say that there
seems to be a relationship between the categorical variables house style and the
presence of basements.

We will also show how we could generate a contingency table with percentages rather
than absolute counts, such as the examples in Tables 2.15, 2.16 or 2.17 in the textbook,
and for the Housing Data may look like this:

<table>
<thead>
<tr>
<th>Style</th>
<th>Basement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Cape Cod</td>
<td>82.35%</td>
</tr>
<tr>
<td>Two-story</td>
<td>0.00%</td>
</tr>
<tr>
<td>Ranch</td>
<td>17.65%</td>
</tr>
<tr>
<td>Grand Total</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Again, we may interpret this table by saying that of all the houses without a basement,
over 82% are Cape Cod style houses, which again gives evidence that there is a
relationship between house style and the presence or absence of a basement.
Before we look at the *bivariate case*, it may be easier to start with a one-way table that summarizes the categorical data for just one variable. The output of this one-way table may then easily be used with graphical displays such as Pie Charts, Bar Charts or Pareto diagrams. Here are the steps to accomplish this. Press the “Insert” tab at the top of the Ribbon. Then choose “PivotTable” in the Tables group as shown below.

In the next Step we choose the data range we want to use for our PivotTable. Excel makes a clever guess and in most instances finds the range of your data automatically and all you have to do is click the *Next* button. In some exceptional cases you may have to enter the range yourself, or select with the mouse after pressing the *red arrow* button on the right. Do not worry that Excel chooses a range that is too big (with, e.g., numerical as well as categorical variables). It comes in handy later when you can do whatever analysis you want (and didn’t think of when creating the PivotTable) since all variables are available. Here we also need to specify where the output (our PivotTable) will go. It’s always a good idea to generate the results in a new worksheet.
After pressing OK, Excel gives us the following PivotTable template:
Now we can drag and drop the categorical variables from the Pivot Table field list on the right to either rows or columns, and then specify what we would like to compute for each value of this variable. Suppose we are interested in knowing how many houses are in each style, then we drag the “Style” field from the right onto the area “Drop Column Fields here” (for no reason did we choose “Column Fields: changing this to “Row Fields” will work as well). We would also want to count the number of observations in each style, so we drop the “Style” field in the “Drop Data Items here” area as well. At first, we get some nonsensical result (see below), and by right-clicking in the “Drop Data Items here” area it is revealed why.
When we click in the “Value Field settings,” we see that Excel **summed** all the values of style. Of course, we would like to just count them, so we choose “count” from the “Summarize value field by” list as follows.
You should now see the following screen:
The output should speak for itself. Try dragging the “style” button in cell A3 to B4. The table “pivots”! Hence the name PivotTable.

From the fields in the Pivot Table Field Lis on the right, we can generate new analyses and views of the data on-the-fly! For example, drag the “school” field from the “PivotTable Field list” to the cell A1 (where it says “Drop Page Fields here”). Now we can pick from the list in cell B1 whether we want the table to include the houses of the Plum Ridge school district (value of school = 0), the Apple Valley school district (value = 1) or both. Choosing the first option, the PivotTable looks like the screen snapshot below. It seems almost as if we can display multiple pages of tables, for different values of the "school" variable. If we don’t need this option any longer, we can just drag the “school” button from cell A1 back to “PivotTable Fields” list.
Now back to our initial problem. We wanted to count the number of houses per style (in the rows) versus the presence or absence of a basement. How to accomplish this? The answer is just too easy: drag the “basement” field to cell B3, where it previously said “Drop Column Fields here”! Your spreadsheet should then look like:
This is pretty much what we set out to do! Now, if we would like the data inside the table to be percentages (percent of grand total, percent of row total, percent of column total), we proceed as follows. Right-click inside the body of the table as before, and choose the “Value Field Settings” option from this list (see page 5). Then click the tab “Show values as” and the following dialog box appears:
Under the drop-down list “Show values as” the following options appear:

Choosing the “% of column” option then transforms our pivot table into the contingency table on the next page. This was the table we were after in the first place.
The reader should feel free to explore some more options and try new things. For example, note that we could even see the relationship between more than two variables. For example, we could see whether the presence of a basement may be explained not only by the style of the house, but by its location (i.e., the school district the house is located in) as well. I.e., we could divide the houses by school district within each house style. This is how we proceed. Drag the “school” button from “PivotTable Fields” list to column A, somewhere between row 4 and 8. Depending on how far you dragged it, your spreadsheet may look like:
The order may be “School”, followed by “Style” if you dragged “School” all the way to the left. Try dragging “School” or “Style” to the COLUMN section (row 3, columns C, D or E) and see what happens.

We can also do many more analyses that are related to statistics, but are extremely valuable for managers. Instead of just looking at the frequencies, we may be interested what the average sales price is as a function of house style and school district. This is how you do it. First, drag the field price from the “PivotTable Fields” to cell A3 (the upper left corner of the PivotTable). If the data displayed in the table isn't the average price (but, e.g., the sum of all sales), then right-click in the DATA section (i.e., the “body”) of the table, and a pop-up menu will appear. From there, choose the Field settings option and you'll again see the available computations for the field price. Choosing “Average” from that list, your final spreadsheet will look as follows:
PivotTables are a very powerful tool to sift through a lot of data from different dimensions quickly, allowing you to compute different things for each dimension rather easily. Although cumbersome at first, PivotTables become easy after just a little bit of practice.