Rockefeller MPA Excel Workshop: Clinton Impeachment Data Example

This exercise is a follow-up to the MPA admissions example used in the Excel Workshop. The basic Excel skills in this exercise are prerequisite for MPA courses. As a graduate student in the MPA program, you are responsible for mastering this material prior to the fall term.

Download the Excel dataset (Welcome Week Impeachment Excel Exercise.xlsx) for this exercise from Welcome Week Central at: http://www.albany.edu/rockefeller/welcomeweek

The data:

For each U.S. Senator, his or her votes on whether to remove President Clinton on each of the two articles of impeachment are provided, as well as each Senator's values on several variables that could be predictive of his or her vote (e.g., Senator's degree of conservatism, how well Clinton did in the Senator's state in the 1996 Presidential election).

1. Check the data. There is an error in the variable “Clinton 1996 vote in state (%).” For one senator, the two digits are reversed. Find this error and correct it.

Hint – use data filter

SOLUTION

This is how the data should look when you open the file:

<table>
<thead>
<tr>
<th>Name</th>
<th>State</th>
<th>Region</th>
<th>Vote on Article I</th>
<th>Vote on Article II</th>
<th>Party</th>
<th>Conservatism</th>
<th>Clinton 1996 vote in state (%)</th>
<th>Year of next election</th>
<th>First-term?</th>
</tr>
</thead>
<tbody>
<tr>
<td>sessions</td>
<td>Alabama</td>
<td>South</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>43</td>
<td>2002</td>
<td>1</td>
</tr>
<tr>
<td>shelby</td>
<td>Alabama</td>
<td>South</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>92</td>
<td>43</td>
<td>2004</td>
<td>0</td>
</tr>
<tr>
<td>murkowski</td>
<td>Alaska</td>
<td>West</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>68</td>
<td>34</td>
<td>2004</td>
<td>0</td>
</tr>
<tr>
<td>stevens</td>
<td>Alaska</td>
<td>West</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>58</td>
<td>34</td>
<td>2002</td>
<td>0</td>
</tr>
<tr>
<td>kyl</td>
<td>Arizona</td>
<td>West</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>96</td>
<td>47</td>
<td>2000</td>
<td>1</td>
</tr>
<tr>
<td>mccain</td>
<td>Arizona</td>
<td>West</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>80</td>
<td>47</td>
<td>2004</td>
<td>0</td>
</tr>
<tr>
<td>lincoln</td>
<td>Arkansas</td>
<td>South</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>36</td>
<td>54</td>
<td>2004</td>
<td>1</td>
</tr>
<tr>
<td>timhutch</td>
<td>Arkansas</td>
<td>South</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>54</td>
<td>2002</td>
<td>1</td>
</tr>
<tr>
<td>boxer</td>
<td>California</td>
<td>West</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>51</td>
<td>51</td>
<td>2004</td>
<td>0</td>
</tr>
<tr>
<td>feinstei</td>
<td>California</td>
<td>West</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>51</td>
<td>2000</td>
<td>1</td>
</tr>
</tbody>
</table>
Click the Data tab and then click “Filter.” You should now see little triangles in the lower right corner of the cells in the top row.

Click on the arrow in the cell for “Clinton 1996 vote in state (%).” A dialog box will appear. Click on “Sort Smallest to Largest.”

The box in the lower right corner for the “Clinton 1996 vote in state (%)” now has a small arrow pointing up. This indicates that all the rows of data are sorted from the smallest to the largest values in that column.

The smallest number for “Clinton 1996 vote in state (%)” is 15. That is much smaller than any of the other percentages which indicates that it may be an error. This is data for Senator Murray from the state of Washington. We can check the data for the other senator to see if it is correct. When we do that we see that the number for Senator Gorton of Washington is 51%, which confirms the data entry error for Senator Murray.
Type in the correct value of 51% for Senator Murphy.

<table>
<thead>
<tr>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conservatism</td>
<td>Clinton 1996 vote in state (%)</td>
<td>Year of next election</td>
</tr>
<tr>
<td>0</td>
<td>51</td>
<td>2004</td>
</tr>
<tr>
<td>68</td>
<td>33</td>
<td>2000</td>
</tr>
<tr>
<td>68</td>
<td>33</td>
<td>2000</td>
</tr>
<tr>
<td>68</td>
<td>34</td>
<td>2004</td>
</tr>
</tbody>
</table>

2. Use conditional formatting to highlight all the cells for Republicans (“Party” = 1) and all the cells that have conservatism scores greater than 50.

SOLUTION

First select column that contains the codes for political party. Do this by clicking on the letter F at the top of the column.

Remember that 1 stands for Republicans and 0 stands for Democrats.

1. On the “Home” tab click on “Conditional Formatting.”
2. In the dialog box that appears, click “Highlight Cells Rules.”
3. In the dialog box to the right of that, click “Equal to…”
In the dialog box that appears, enter 1 (Republican) in the box labeled “Format cells that are EQUAL TO:” and choose your favorite color scheme in the other box. Click OK. Republicans are now highlighted.

For the second part of the problem first select the variable “Conservatism” by clicking on the letter G. at the top of the column.
On the “home” tab, click on “Conditional Formatting,” then “Highlight Cells Rules,” and then “Greater Than…”

In the dialog box that appears enter “50” and your favorite color scheme for the cells.

Conservatism scores above 50 are now highlighted.

Do you notice any pattern in the two columns that you have conditionally formatted?
3. Create a new variable that is the total number of guilty votes by each senator for the two articles of impeachment.

**SOLUTION**

First create a blank column for the new variable. Right-click on column F and then click on the insert in the menu that appears.

Type a heading for the new variable “Number of guilty votes.”
Enter the formula in the cell directly under the variable heading (Cell F2).

The formula is \( =D2+E2 \)

You can add either type the formula in directly, or you can select the cells that are included in the formula and their addresses will appear in the formula.

Press enter to complete the formula.

The total now appears in the cell where you typed the formula. You can copy and paste the formula into the remaining cells in the column. Since you used relative cell references (no “$”), each formula will use the data from the adjacent cells to the left.

A shortcut to copying and pasting the formula is to select the cell and click on the on the small square handle on the lower right corner of this cell and drag it down.

Another shortcut is to select the cell containing the formula and double click on the small square handle on the lower right corner of this cell. The formula will be copied down as long as there are adjacent cells containing data.
Do you see any relation between the senators’ party and their number of guilty votes?

Complete the table below and paste it into a Word document.

Table 1. Number of states by region.

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of states</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
</tr>
</tbody>
</table>

SOLUTION

As with almost anything in Excel, there is more than one way to do this.

First, insert a new worksheet and create the table on that worksheet.

Near the bottom left corner of your screen is a small tab for creating a new worksheet. Click on that tab.
The new worksheet will be given a name like “Sheet1.” You should rename it “Tables.” To do this, right click on the tab with the worksheet name and then click on “Rename” on the menu that appears.

The name of the worksheet will be highlighted and you can type in the new name.

Press enter after you type in the name.

In the new worksheet, enter a blank table matching the one in the exercise. Format the borders and shading to taste.

A keyboard shortcut for formatting cells is to select the cells you want to format and then press Ctrl-1. The cell formatting dialog comes up and you can change borders, fill, etc. to your heart’s desire.
Use the “countif” function to fill in the table.

1. Select cell B2 – the first data cell in the table.
2. Click the “Formulas” tab.
3. Click “Insert Function.”

1. Type “COUNTIF” in the box under “Search for a function.”
2. Then click “Go”

A brief description of the COUNTIF function appears below the box. If you are not sure what function will serve your needs, you can browse them using this dialog box.

3. Click “OK.”

A dialog box for choosing the arguments for the COUNTIF function appears. You can type in the arguments directly, or you can select a range for each argument by clicking on the box and selecting the range by dragging your mouse.

In this example, the “Range” (range of data to be counted) contains the values of the “Region” variable in the data. It is an absolute cell reference ($) because you
don’t want the range to change when you copy and paste the formula into another cell.

The “Criteria” box specifies which cells you want to be counted. In this case, cell A2 contains “Northeast” so all instances of “Northeast” in the specified range will be counted.

When you click “OK” the number of values of “Northeast” will appear in the cell.

But remember that there are two senators from each state, so the number has to be divided by 2. You can press F2 to edit the formula in the cell, or you can click in the formula box and add “/2” to the end of the formula. This will divide the result of the COUNTIF function by 2, giving the correct number of states.

Now copy the formula down to the cells below it by double clicking on the “handle.” The number of states in each region will appear in the cells.

As a check, use the sum function to make sure that there are a total of 50 states.

First, select the cell with the total in it (B6).

Then, select the sum function
from right end of the ribbon ("Home" tab).

The sum function guesses which numbers you want to sum and surrounds them with a dashed “marquee.” If you are satisfied with the guess, click Enter.

The total number of states is correct. The table is done. (except for the mixed fonts, which should be fixed).

Now the challenge is to copy it to a Word document. Fortunately, that’s very easy.

Select the entire table by dragging the mouse over it. Press Ctrl-c to copy it. The marquee appears, indicating that the table has been copied.

Open a Word document (or create a new one) and move the cursor to the place you want the table to be. Press Ctrl-v to paste the table.
4. Create a table of basic descriptive statistics for “Conservatism” and “Clinton 1996 vote in state” and paste it into a Word document.

Table 2. Description of Conservatism and Percent of vote for Clinton in the senator’s state (1996) for 100 senators.

<table>
<thead>
<tr>
<th></th>
<th>Conservatism</th>
<th>Vote for Clinton in 1996 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SOLUTION**

As in the previous question, create the blank table in your Excel worksheet with row and column headings.

Select the cell that will contain the mean for “Conservatism.”

Then click “Insert Function” on the “Formulas” tab.

Select the category “Statistical.”

Click on “Average.” This is the function that computes the mean.
The dialog box for the “average” function will appear. Make sure the cursor is in the box for “Number 1” and then select the data for “Conservatism” on the “Data” worksheet.

The marquee shows that the data have been selected.

Hit enter and then click “OK.”

The mean for “Conservatism” now appears in the cell.

Follow the same pattern for entering the functions for Median (“MEDIAN”), Minimum (“MIN”), and Maximum (“MAX”).

The table is complete for “Conservatism.”

Since the column of data for “Clinton 1996 vote in state (%)” is just to the left of the column for “Conservatism,” there is a shortcut to entering the formulas for that variable.
1. Select the cells containing statistics for “Conservatism.”

2. Copy and paste them to “Vote for Clinton,”

….or…

2. Drag the handle at the lower right corner of the selection one column to the right.

The table is done. Select the cells; copy and paste into a Word document as before.

5. Calculate the mean level of conservatism for Republicans and Democrats in each of the 4 regions with numbers formatted to two decimal places.

Table 3. Mean conservatism by region and party.
SOLUTION

The completed table looks like this:

<table>
<thead>
<tr>
<th>Region</th>
<th>Democrat</th>
<th>Republican</th>
<th>Overall mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>8.80</td>
<td>53.13</td>
<td>28.50</td>
</tr>
<tr>
<td>Midwest</td>
<td>11.67</td>
<td>79.33</td>
<td>45.50</td>
</tr>
<tr>
<td>South</td>
<td>14.15</td>
<td>87.79</td>
<td>57.88</td>
</tr>
<tr>
<td>West</td>
<td>3.60</td>
<td>77.38</td>
<td>49.00</td>
</tr>
</tbody>
</table>

| Overall mean | 9.96  | 77.87 | 47.31 |

One way to do this is to use the sorting function to isolate the subsamples and compute their means directly.

Another way is to use the “AVERAGEIF” and “AVERAGEIFS” functions to compute the average for subsamples that are identified for a variable. The procedures for entering functions have been show above.

Below is the AVERAGEIFS dialog box for obtaining the mean for Democrats from the Northeast.
**Average_range** is the range of data to be averaged. The data in Column H – “Conservatism” are specified. $’s are used to make this an absolute reference because we will want to copy this formula to other cells without changing the data to be averaged.

**Criteria_range1** is the variable containing the first set of criteria for selecting a subsample of data. The data in column G for “Party” are specified because we want to select Democrats.

Criteria1. Cases will be selected from **Criteria_range1** based on this value. 0 is entered to select Democrats. When the formula is copies to the Republicans, you will have to remember to change this to a “1.”

**Criteria_range2** is the variable containing the second set of criteria for selecting a subsample of data. The data in column C for “Region” are specified because we want to select Northeast.

Criteria2. Cases will be selected from **Criteria_range2** based on this value. $A3 is entered to select Northeast. The “$A” is used because when we copy the formula to the right, we don’t want the column to change. But when we copy the formula down, we do want the row to change.

To get the overall means for rows and columns, we use the AVERAGEIF function.

Below is the AVERAGEIF dialog box for obtaining the overall mean for all senators from the Northeast.

![AVERAGEIF dialog box](image)

**Range** is the variable for selecting cases. In this case it is “Region” in Column C.

**Criteria.** Cases will be selected from **Range** based on this value. $A3 is entered to select Northeast.

**Average_range** is the range of data to be averaged. The data in Column H – “Conservatism” are specified.
Below are the complete formulas (rearranged to fit on a page).

<table>
<thead>
<tr>
<th>Region</th>
<th>Democrat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>=AVERAGEIFS(Data!H$2:H$101,Data!S$2:S$101,$A3,Data!S$2:S$101,0)</td>
</tr>
<tr>
<td>Midwest</td>
<td>=AVERAGEIFS(Data!H$2:H$101,Data!S$2:S$101,$A4,Data!S$2:S$101,0)</td>
</tr>
<tr>
<td>South</td>
<td>=AVERAGEIFS(Data!H$2:H$101,Data!S$2:S$101,$A5,Data!S$2:S$101,0)</td>
</tr>
<tr>
<td>West</td>
<td>=AVERAGEIFS(Data!H$2:H$101,Data!S$2:S$101,$A6,Data!S$2:S$101,0)</td>
</tr>
<tr>
<td>Overall mean</td>
<td>=AVERAGEIF(Data!S$2:S$101,0,Data!H$2:H$101)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>=AVERAGEIFS(Data!H$2:H$101,Data!S$2:S$101,$A9,Data!S$2:S$101,1)</td>
</tr>
<tr>
<td>Midwest</td>
<td>=AVERAGEIFS(Data!H$2:H$101,Data!S$2:S$101,$A10,Data!S$2:S$101,1)</td>
</tr>
<tr>
<td>South</td>
<td>=AVERAGEIFS(Data!H$2:H$101,Data!S$2:S$101,$A11,Data!S$2:S$101,1)</td>
</tr>
<tr>
<td>West</td>
<td>=AVERAGEIFS(Data!H$2:H$101,Data!S$2:S$101,$A12,Data!S$2:S$101,1)</td>
</tr>
<tr>
<td>Overall mean</td>
<td>=AVERAGEIF(Data!S$2:S$101,1,Data!H$2:H$101)</td>
</tr>
</tbody>
</table>

(A third, and probably the best way is to use pivot tables, but we don’t have time to study that in the workshop.)

6. Plot mean conservatism by party and region as a bar chart. Include a legend, axis titles, and a chart title.

**SOLUTION**

This is very easy to do once you have the table from the previous problem.

Select the cells containing the row and column headings and the means.
On the “Insert” tab, in the “Charts” group, click “Column.”

There are many bar chart formats. The first one is simple and often the best, but it depends on the data and your preferences.

Click on the icon.

The chart appears on the sheet. Your colors may differ. Excel selected red for Republicans and blue for Democrats!

But this chart is missing a title and labels for the axes. The title is optional. If the chart is going into a report, a figure caption will be added in the report and the title will be
redundant. But all axes should be labeled.

Click on the chart to select it and the “Chart Tools” will appear in the ribbon.

Click on “Add Chart Element.”

1. Click on “Chart Title.”
2. Click on “Above Chart.”

Now click on “Axis Titles” and insert Horizontal and Vertical axis titles.
Here is what the chart should look like after you insert chart and axis titles.

The title and axis labels are changed by simply clicking on them and editing them. Often the fonts have to be changed because Excel’s default fonts are too small. The tick labels for Mean Conservatism have been changed to integers because the decimal places took up space without communicating any information. Colors, gridlines and other chart features can be edited to produce clean, professional charts.
7. Plot the percent of impeachment vote by region as a pie chart.

Below is a table of the percent of the impeachment vote on Article II by region.

<table>
<thead>
<tr>
<th>Region</th>
<th>Percent of impeachment vote on Article II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>6.0%</td>
</tr>
<tr>
<td>Midwest</td>
<td>24.0%</td>
</tr>
<tr>
<td>South</td>
<td>38.0%</td>
</tr>
<tr>
<td>West</td>
<td>32.0%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Plot this table as a pie chart. Include a legend.

**SOLUTION**

Enter the table into an Excel worksheet.

Select the cells shown. Don’t select the row containing the total or it will appear on the pie chart.

On the “Insert” tab, select “Pie” chart.
From the drop down menu, select your favorite pie chart design.

Here is a basic pie chart. You can change fonts and colors to your taste.

8. Make a scatterplot with Conservatism on the X-axis and the Clinton 1996 vote in state (%) on the Y-axis. Include axis labels and title. Since there is only one series, you don’t need a legend.
Select the data columns for Conservatism and Clinton 1996 vote in state (%) (columns H and I).

You can chart non-adjacent columns of data, but it is easier if they are adjacent. The first one will end up on the X-axis. If that is not what you want, you can fix it by editing the chart.

Click “Scatter” in the charts group on the “Insert” tab.

There are many options for scatterplots. Choose the first one (just points, no lines) by clicking on it.
Here is the default chart. It is not acceptable for any paper or report or presentation. It has no axis labels and it has an unnecessary legend.

Add axis titles as shown above (Problem 7).

To get rid of the legend, just click on it to select it and press Delete.

Click on the words “Axis Title” and enter the correct titles.

Notice that the scale for conservatism goes to 120, but the actual data stop at 100.
To change the scale, right click on the axis and click “Format Axis…”

In the “Format Axis” dialog box, under “Axis Options,” Click “Fixed” for the Maximum and enter the value 100.

Click Close
Here is the chart after changing the maximum value on the X-axis and making some other cosmetic changes, including:

--larger fonts for the axis labels
--changing to round black markers
--dragging the corner of the chart to resize and reshape it.

Do you see any relation between Senators’ conservatism and the vote for Clinton in their state?

9. Recommended book for learning more about Excel:


If you run into specific issues, Google is also a great resource.

There are many other handbooks, so ultimately the best book is the one you will use. Please let the Welcome Week Excel Team know if you have other recommendations!