Intermediate Excel Workshop

Empirical Reasoning Center

June 7, 2016

1 Nested Formulas

Create an 'Attractive' variable defined by having a beauty value greater than or equal to the average beauty in the sample. To do this use nested functions and if statements. In cell I1 label the column 'Attractive'.

This variable should equal 1 if that observation's beauty value is greater than the average beauty of the sample and 0 otherwise. An if statement requires three inputs: the first is a comparison, the second is what value the cell should display if the comparison is true, and what value the cell should display if the comparison is false. In this case the inputs are: the beauty value is greater than or equal to the average beauty, the true value is one, and the false value is zero. So in cell I2 enter the formula '=if(E2>=average($E$2:$E$464),1,0)'. Apply this formula to the entire column.

2 PivotTables

A PivotTable can help summarize data based on different categories.

What is the average course evaluation score by gender and attractiveness? PivotTables automate these calculations. To create a PivotTable, select the 'Insert' tab, within the 'Tables' group select 'PivotTable'. In the menu box, for 'Table/Range' select the entire dataset, for 'Location' select the cell where you want the PivotTable to begin (cell K1), and select 'OK'. You will create and edit the PivotTable using the PivotTable Fields menu. Select the variables you want to analyze in the PivotTable – 'female', 'course eval', and 'Attractive'. The variables will appear in the 2×2 table below. You can drag each variable to the other quadrants. Drag the female box to the 'Columns' quadrant, drag the course eval box to the 'Values' quadrant, and leave the Attractive box in the 'Rows' quadrant. Notice that the PivotTable updates with every change. To display the average course evaluation values, select the course eval box in the 'Values' quadrant and select 'Value Field Settings'. In the menu box, select 'Average'. Edit the 'Custom Name' to 'Average Course Evaluation Score' and select 'OK'.

Now the PivotTable should be displaying the correct statistics. The PivotTable still requires formatting for the information to be clear. Delete the row labeled 'blank'. Change the column and row headers from zeroes and ones to 'female', 'male', 'unattractive', and 'attractive'. Change the variable titles by double clicking the cells and editing the 'Custom Name' – change 'Attractive' in cell K2 to 'Attractiveness' and change 'female' in cell L1 to 'Gender'.

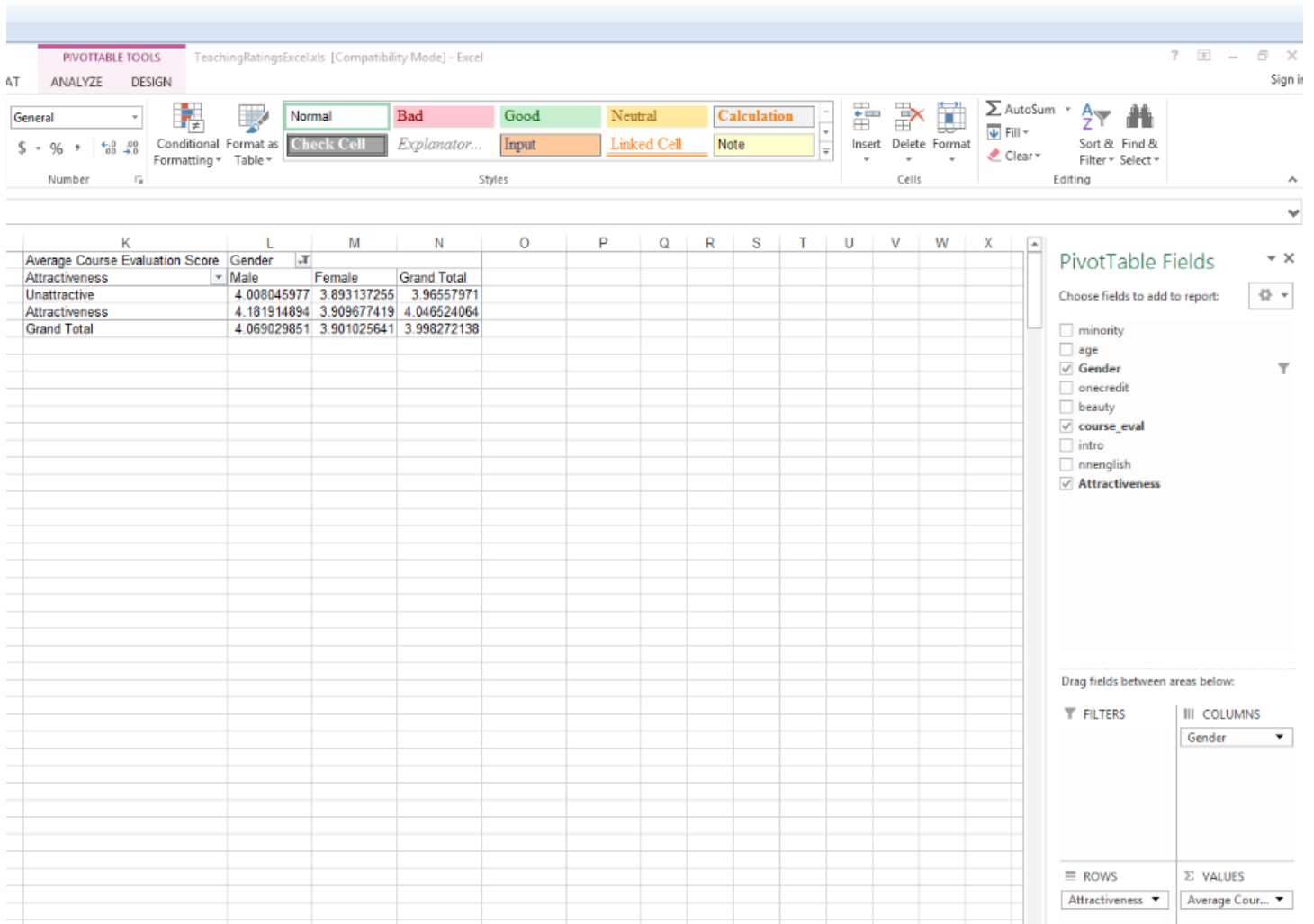The final PivotTable should resemble Figure 1.

Figure 1: PivotTable

3 Data Analysis ToolPak I

The Data Analysis ToolPak is used for complex, statistical analyses. It is an Add-In that you will need to install. To install it, select 'File', then 'Options', then select the 'Add-Ins' tab. Under 'Inactive Application Add-ins' select 'Analysis ToolPak'. For the 'Manage' dropdown menu select 'Excel Add-ins' and select 'Go'. Select 'Analysis ToolPak' and select 'OK'.

If you select the 'Data' tab, the 'Analysis' group will now show a 'Data Analysis' feature.

3.1 Descriptive Statistics

Select the 'Data Analysis' feature. Select 'Descriptive Statistics' and then 'OK'. For the 'Input Range' select the variables you want analyzed. For this example, select all of the variables (columns A to I). This dataset is grouped by 'Columns'. Select 'Labels in the first row' and for 'Output Options' select 'New Worksheet Ply'. In the textbox label the new worksheet 'Desc. Stats'. Then select the statistics you want reported: 'Summary statistics' and 'Confidence Level for Mean'. Select 'OK'.

A new worksheet will appear with the descriptive statistics.

3.2 Histograms

Histograms are used to display frequency data.

Select the 'Data Analysis' feature, select 'Histogram', and then select 'OK'. For the 'Input Range' select one column of data. For this example, analyze the course evaluation values, so type '=f1:f464'. Select 'Labels' if the input range contains a column label (like course eval). For 'Output Options' select 'New Worksheet Ply' and label the worksheet 'Histogram'. Then select 'Cumulative Percentage' and 'Chart Output', then select 'OK'.

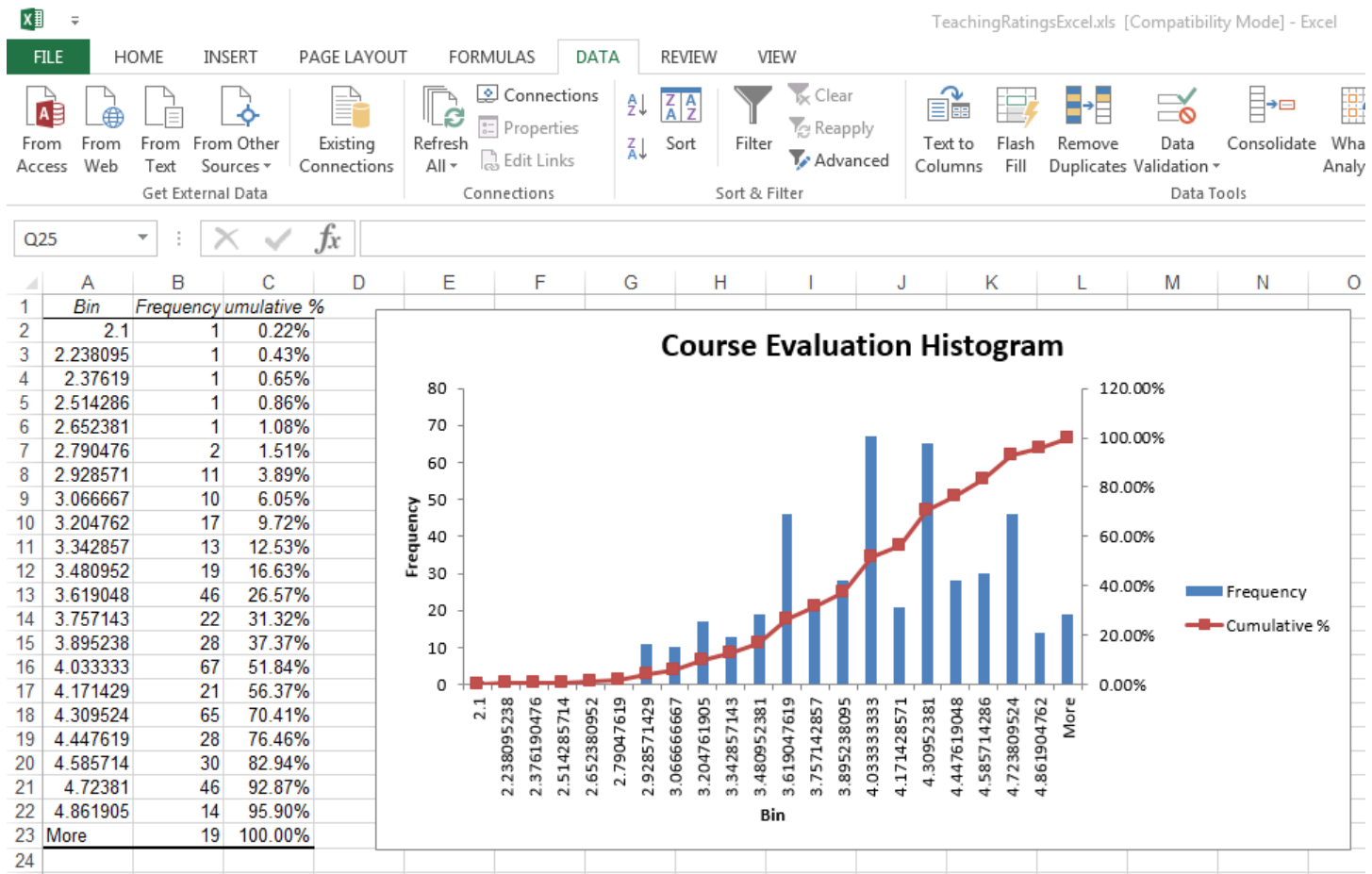A new worksheet will appear with the histogram table and chart as in Figure 2.

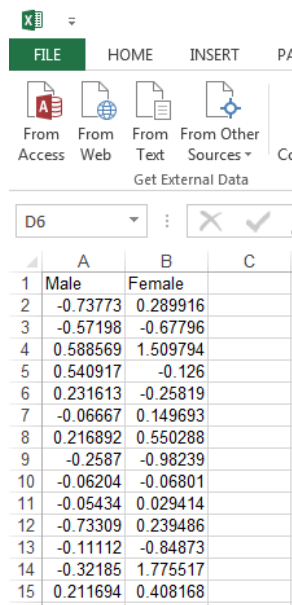Figure 2: Histogram

4 Sort & Filter

4.1 Sorting

Sort the dataset by gender. Select all of your data. Select the 'Home' tab, then within the Editing group select the funnel icon labeled 'Sort & Filter', and select 'Custom Sort'. In the top right corner of the menu box select 'My data has headers'. For the 'Sort by' dropdown, select 'female'. The 'Sort On' and 'Order' options allow you to set how the data is sorted. They should be set to 'Values' and 'Smallest to Largest', respectively. Now add a second level of sorting: within each gender, sort by beauty. Select 'Add Level' and in the 'Then by' dropdown select 'beauty'. Change the 'Order' dropdown to 'Largest to

Smallest' for the beauty sorting. Select 'OK'.

4.2   Filtering

Isolate the male observations. Select the entire dataset, within the Editing group, select the 'Sort & Filter' dropdown, and select 'Filter'. Dropdown arrows will appear in every column header. Select the dropdown for the 'female' column and unselect the value one and select 'OK'. All of the observations that are women are hidden. Copy and paste the beauty column to a new Worksheet. Relabel the column 'Male'. Label that worksheet 'Beauty/Gender'. Then go back to the female drop down and select the one value and unselect the zero value. Now male observations are excluded. Copy and paste the beauty column to the Beauty/Gender Worksheet and relabel the column 'Female'. To turn the filter off, select the 'Sort & Filter' dropdown, select 'Clear', and unselect 'Filter'.

The Beauty/Gender worksheet should look like Figure 3.

Figure 3: Sorted & Filtered Data

5 Data Analysis ToolPak II

5.1 Single Factor ANOVA

You can determine whether the difference between sample means is significant or not using a Single Factor ANOVA.

Use the Beauty/Gender Worksheet to determine if the average beauty is statistically different across genders. Select the 'Data Analysis' feature, select 'ANOVA: Single Factor', and select 'OK'. For the 'Input Range' select columns A and B, including the headers. Select 'Labels in First Row' and 'New Worksheet Ply'. Label the new worksheet 'ANOVA' and select 'OK'. A new worksheet will appear with the ANOVA table.

Interpreting the results: If the F value is larger than F crit, you must reject the null hypothesis. That is, there is a significant difference between the population means. Also, note that the p-value is much smaller than our alpha value .05, which is another indication that we must reject the null hypothesis.

5.2 Regressions

Linear regression analysis uses the "least squares" method to fit a line to the set of observations. You can analyze how a dependent variable is affected by one or more independent variables.

Univariate Regressions   To run a univariate regression, select the 'Data Analysis' feature, select 'Regression', and then select 'OK'. For 'Input Y Range' select the dependent variable. For this example select the course eval column (F1:F424). For 'Input X Range' select the independent variable. For this example select the beauty column (E1:E424). Select 'Labels' and 'New Worksheet Ply'. Label the new worksheet 'Univariate'. Then select 'Line Fit Plots' then 'OK'. The Univariate Worksheet will contain the output and a chart.

Multivariate Regressions   To run a multivariate regression, select the 'Data Analysis' feature, select 'Regression', and then select 'OK'. For 'Input Y Range' select the dependent variable. For this example select the course eval column (F1:F424). For 'Input X Range' select the independent variables. For this example select the female, one credit, and beauty columns (C1:E424). Select 'Labels' and 'New Worksheet Ply'. Label the new worksheet 'Multivariate'. Then unselect 'Line Fit Plots' then 'OK'. The Multivariate Worksheet will contain the output and the chart.