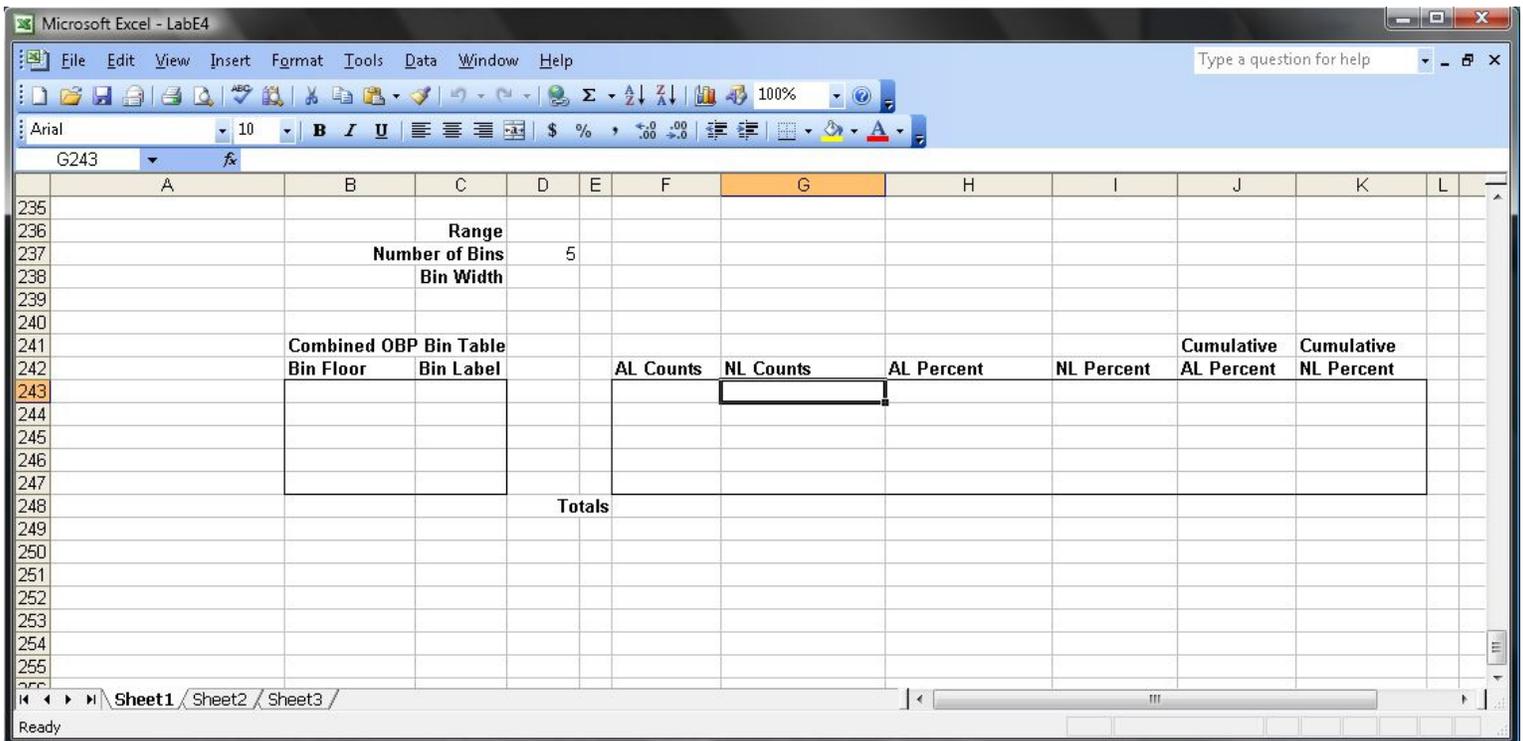


CSU101 Summer 2009 Lab Assignment E4

To complete this assignment you must submit an electronic copy to Blackboard by the due date.

Download the LabE4.xls file from Blackboard (or the course website), and save a copy of it. This is a solution to Lab E1. Using the data that was calculated previously, we wish to create a histogram, a frequency polygon, and a cumulative frequency polygon, that will allow us to compare the distributions of OBP for AL and NL players. A distribution tells you how many cases there are for various ranges of values of some characteristic.

Below the existing answers, cells have been added similar to the ones in the following picture.



1. In order to create a histogram, we need to group individual data points into groups that have similar values. We will call these groups of values “bins”. Each bin will contain a range of values, and we call the size of each range of values the bin width. Next to cell containing “Range” add an address to the cell in your previous assignment that gave the range for the **combined** population of AL and NL players. We are trying to divide the range for the combined population into 5 equal-sized bins. The bin width should be the range divided by the number of bins, *rounded* to 3 decimal places. In the example above, the lookup table is labeled Combined OBP Bin Table. Each bin has a minimum value, which we call the bin floor. The floor of the first bin must be the minimum OBP value of the combined population.

The floor of the next bin is the floor of the previous bin plus the width of the bin. Bin floors should be *formatted* to 3 decimal places (don't round). Place suitable bin labels in the second column of the table. For our purposes, suitable bin labels should be the mid-point of the bin, also *formatted* to 3 decimal places, and should be calculated from its corresponding bin floor.

2. Assign the name *BinTable* to your lookup table, and use it in a VLOOKUP to assign bin labels to each AL and NL player in our population. The AL values should appear in the column to the right of the AL players' OBPs, and the NL's in the column to the right of the NL's OBPs. Assign the range names *AL_OBPVal* and *NL_OBPVal*, to these two ranges, respectively. Format these ranges to 3 decimal places.

3. We will report our results in the table to the right of the bin lookup table. Each row in that table corresponds to the bin in the lookup table. The first column should contain the number of cases in *AL_OBPVal* of bin values matching the bin value for that row. The second column should contain the number of cases in *NL_OBPVal* of values matching the bin value for that row. In the totals row, give the total of each of these columns. At this point, add the appropriate percentages in the next two columns. The percentages should total to 100% in each column. In the last two columns, put the cumulative percentages. Cumulative percentages are the sum of the percentages for all the bins up to, and including, that bin. The final cumulative percentage should be 100%

4. You are now ready to plot the histogram. Construct a chart, **a column chart to be placed in the spreadsheet**, using the bin values as your category labels, and using the first two columns of your result table as your two data series. Provide the title "Histogram of AL and NL OBP", and label the data series "AL" and "NL", respectively. The x values should be the interval mid-points, i.e., the bin labels, and the y values should be the appropriate *counts*. Label the x axis "OBP Bin Midpoint", and the y axis "# of Players".

5. Now construct another chart: **a line chart with data points shown**. This time the chart title should be "Frequency Polygon of AL and NL OBP". The x values should be the interval mid-points, i.e., the bin labels, and the y values should be the appropriate *percentages*. Remember, your chart should show two graphs (lines). Label the x axis "OBP Bin Midpoint", and the y axis "% of Players".

6. Now construct another chart, a cumulative frequency polygon, which plots as its x values the interval **endpoints** (i.e., not the labels, see the note below), and as its y values the cumulative percentages. Your chart this time is **a line chart with data points shown**, and its title should be "Cumulative Frequency Polygon of AL and NL OBP". Label the x axis "On Base Percentage", and the y axis "Cum. % of Players".

Note: To get the endpoints for each bin, just use the floor of the next bin. You'll need to create a dummy 6th bin, so that you can get the endpoint of the last (5th) bin. This dummy bin may also be useful (but not necessary) to calculate the midpoints.