Assignment A03: Tables, Descriptive Statistics, Histograms, and Bell Curves

Test scores are provided for two tests taken by students in one course and the data (rows listing first name, score in Test #1, score in Test #2) are provided online (DOC file).

(a) Create a Data Table in Excel. Organize the table by alphabetizing the order variable (the names) using the "sort" function in Excel. Include a table title and table column headers.

(b) Using Excel, compute the following descriptive statistics for both sets of the legitimate variables (test scores): Count, average, standard deviation, minimum, maximum, median, and mode. These statistical data should appear in the Table below the data rows.

(c) Using Excel, create one marked scatter plot that shows the data of Tests #1 (red squares) and #2 (blue triangles). Include horizontal lines at the respective average value and using the appropriate color. Do the data clump?

(d) Create histograms for Tests #1 and #2. Full range of possible values, bin range ≥ 3, no gaps. Show both histograms with the same height (i.e., 3 inches) and the same width (i.e., 6 inches).

(e) Compute values of the functions $f(x) = (2\pi\sigma^2)^{-0.5} \exp(-(x-a)^2/2\sigma^2)$ using the averages a and standard deviations σ for Tests #1 and #2 and create unmarked line plots of the functions for $0 \leq x \leq 100$ together in one graph. Show the graph with the same height and width settings as you used for the histograms. Are the test score distributions well described by a normal distribution?

Complete the assignment with MS Excel. Submit one Excel file "A03_'your_last_names'.xlsx" with **(a)-(c)** on sheet #1 (sheet label "Data Table and Scatter Plots"), **(d)** on sheet #2 (sheet label "Histograms"), and **(e)** on sheet #3 (sheet label "Gaussian Plots").

Deadlines: Submit electronic file on Tuesday, 02/21/17 by midnight. Bring one (stapled) hardcopy to class on Wednesday, 02/22/17.

See back for comments regarding items (d) and (e).

RE (d) creating histograms using the FREQUENCY array formula: Older versions of Excel contained the "Analysis ToolPak" which included a histogram tool. Beginning with SP14, we no longer rely on the "Analysis ToolPak." Instead, you are required to determine the frequency distribution for a given bin range with the FREQUENCY function and to plot the histogram as a column graph of the frequencies.

Microsoft Support describes the FREQUENCY function online at http://support.microsoft.com/kb/100122. Note that one MUST press CONTROL+SHIFT+ENTER (in Microsoft Excel for Windows) or COMMAND+ENTER (in Microsoft Excel for the Macintosh) to enter the formula as an array formula. Instructions for the creation of a histogram using the FREQUENCY function can be found online at various places; for example, at URL http://www.ncsu.edu/labwrite/res/gt/gt-bar-home.html#ith. To create the histogram, first plot the frequency data as a column graph and then click the plot area and "Select Data" to modify the "Category (X) axis labels" using the bin values.

RE (e) creating the Gaussian function: To plot the Gaussian function $y = f(x) = (2\pi\sigma^2)^{-0.5} \exp(-(x-a)^2/2\sigma^2)$ using the averages a and standard deviations σ in the range $0 \le x \le 100$, one first needs to compute pairs (x, y).

Start by entering integers 1, 2, …, 100 in column A in cells A1:A100. This is best done by (a) entering "1" in A1, (b) entering "=A1+1" in A2, and (c) copying cell A2 to cells A3:A100.

There are several ways to compute the y values in columns B1:B100. The most direct way is to enter "=(2*PI()*mysigma^2)^-0.5*EXP((A1-myaverage)^2/(-2*mysigma^2))" in B1, where "mysigma" and "myaverage" are the values of the two parameters of your Gaussian function. For example, you can type the mysigma value in C1 and the myaverage value in C2, and then use "=(2*PI()*C1^2)^-0.5*EXP((A1-C2)^2/(-2*C1^2))". Excel has the value of π build in as the argument-free function PI(). Then copy B1 to cells B2:B100. Another way to compute the y values uses the function NORM.DIST. Enter "=NORM.DIST(A1,myaverage,mysigma, FALSE)" in B1 and then proceed as before.