

**Math 263-007 Excel Assignment # 4**  
Due 3/26/15 in class or to my office before 5:30 PM.

**Tip: read through the entire assignment before starting to get an idea of what we're doing.**

**Instructions** The purpose of this assignment is to investigate the sampling distribution of a sample mean and the central limit theorem. The data set (263S15HW4Dat.xlsx) has 4 tabs/sheets:

- *Original* show the values of the pdf of the random variable  $X$ . The mean and standard deviation of this distribution are 10 and 5.77, respectively.
- *Samples\_4* shows a live simulation of 1000 samples of size 4. Their means are calculated.
- *Samples\_25* has the start of a simulation for 1000 samples of size 25. The first line is completed for you.
- *Samples\_50* has the outline for a simulation for 1000 samples of size 50, but has not been completed.

As usual, please turn in a nicely formatted hand-written or typed document, copying values and graphs over from excel. Do not turn in an excel spreadsheet with answers typed into cells.

1. Is the random variable  $X$  quantitative or categorical? Discrete or continuous? How do you know? (Look in the *Original* sheet. You could try adding the probabilities, and remember that discrete random variables must have the sum of their probabilities equal to 1.
2. Using the information in the *Original* sheet:
  - (a) Draw a graph of the pdf of  $X$ , using either Excel (a scatterplot is fine) or by hand. Label the axes with a scale (Excel will do this automatically).
  - (b) What kind of distribution is shown?
  - (c) Given the shape of the graph, why does the pdf take the value 0.05 at each given  $x$  value?

## Interpreting the Simulations

- 3.** In the sheet *Samples\_4*, Column F shows the means of the 1000 samples of size 4. This is an approximation to the sampling distribution of sample means. In class, we learned that the mean and standard deviation of Column F should be, respectively  $\mu_X$  and  $\sigma_X/\sqrt{N}$  where  $N$  is the sample size (here,  $N = 4$ ).
- (a) In the orange box, calculate the mean of the sample means using the Excel command =AVERAGE(array).
  - (b) Calculate the standard deviation of the sample means using the other orange box and the Excel command =STDEV.S(array) (the .S stands for ‘sample’). If your version of Excel is older you can just use =STDEV(array).
  - (c) Compare the mean you just calculated to the mean of the original distribution, which is 10. Are they close or not? Is this expected?
  - (d) Compare the standard deviation you just calculated to the standard deviation of the original distribution, which is 5.77. Are they close or not? Is this expected?

## Extending a Simulation

- 4.** In the sheet *Samples\_25*, the simulation has been started in Row 6. When it is finished, there will be 1000 samples of size 25.
- (a) Finish the simulation by copying down Columns B to AA. The easiest way to do this is to select cells B6:AA6 by first clicking B6, then holding shift and clicking AA6. Then, *double click* the lower left corner of the blue selection box (the corner you would ‘drag down’). This should fill all the cells from row 7 to row 1005. If this doesn’t work, you can also accomplish this by selecting cells B6:AA1005 by first clicking B6, then holding shift and clicking AA1005; then, go to Edit:Fill:Series. Select the buttons for ‘Columns’ and ‘AutoFill’, then click OK. Lastly, you can also select row 6, copy it with Edit:Copy or ctrl-C, then select the remaining cells (B7:AA1005) and paste. You do not need to turn anything in for this question.
  - (b) In the orange box, calculate the mean of all the sample means.
  - (c) In the other orange box, calculate the standard deviation of the sample means.
  - (d) Compare the means you calculated with the mean of the original distribution. Are they close? Is this expected?
  - (e) Compare the standard deviation you calculated with the standard deviation of the original distribution. Are they close or not? Is this expected?

## Creating a Simulation

The simulations in *Samples\_4* and *Samples\_25* use the Excel function =RAND() which generates a random number between 0 and 1. You will now use this function to create a simulation from scratch.

5. In the sheet *Samples\_50*, the simulation has not been started, but a template is laid out.
  - (a) Start the simulation in Cell B6 by typing =20\*RAND(). What does this command do?
  - (b) Copy to the right from cell B6 to cell AY6. The easiest way to do this is with copy/paste i.e. select B6, copy it, then select C6:AY6 and paste. There is nothing to hand in for this question.
  - (c) Calculate the mean of each sample using =AVERAGE(array). First do this for row 6, then copy the formula down by selecting AZ6 and double-clicking the lower right corner.
  - (d) In the orange box, calculate the mean of the sample means.
  - (e) In the other orange box, calculate the standard deviation of the sample means.
  - (f) Compare this mean to the mean of the original distribution. Are they close? Is this expected?
  - (g) Compare this standard deviation to the standard deviation of original distribution. Is this expected?

### Summary Observations: Effect of Sample Size

6. Consider the means of all samples of a fixed size,  $N$ , of this random variable. If  $N$  increases, what happens to:
  - (a) The mean of these sample means? Does it increase, decrease, or stay the same?
  - (b) The standard deviation of the sample means? Does it increase, decrease, or stay the same?