

**Math 263: Excel Assignment 4 on Sampling Distributions of Means**

***Due Thursday February 20, at the start of class. Hand in a paper print-out of your work.***

*Everyone must do these simulations individually, although you are welcome to talk to Adrian or I about them.*

*Suggestion: Read the assignment all the way through before starting, so you can see what questions it answers.*

**Purpose of Assignment: Sampling Distributions**

This assignment is about the *sampling distribution of means* and how the size of the sample affects the sampling distribution.

We consider a random variable  $X$ . In the spreadsheet *CLTmeans.xlsx*, there are four pages:

- **Original** shows 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 line for Sample 1 is complete.
- **Samples\_50** has the outline for a simulation for 1000 samples of size 50, but has not been started.

1. Is the random variable  $X$  quantitative or categorical? Discrete or continuous? How do you know? (Look at what is in the *Original* sheet. You could try adding the probabilities.)
  
2. Using the information on the *Original* sheet:
  - (a) Draw a graph of the pdf of  $X$ , either using Excel (a scatterplot will work) or by hand. Label the axes with a scale. (Excel will do this automatically.)
  - (b) What is the name of this distribution?
  - (c) Given the shape of the graph, why does the pdf take the value of 0.05 at each  $x$  value?

**Interpreting the Simulation**

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 means. In class, we will compare the mean and standard deviation of these simulated means with what you would expect from theory.
  - (a) Calculate the mean of all the sample means. (Orange box; use the command =AVERAGE(array))
  - (b) Calculate the standard deviation of all the sample means. (Orange box; use the command =STDEV(array) or =STDEV.S(array))
  - (c) Compare the mean you calculated with the mean of the original distribution, which is 10. Are they close or not?

- (d) Compare the standard deviation you calculated with the standard deviation of the original distribution, which is 5.77. Are they close or not?

### 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. (Copy is under the edit menu. There is nothing to hand in for this part.)
  - (b) Calculate the mean of all the sample means. (Orange box)
  - (c) Calculate the standard deviation of all the sample means. (Orange box)
  - (d) Compare the mean you calculated with the mean of the original distribution. Are they close or not?
  - (e) Compare the standard deviation you calculated with the standard deviation of the original distribution. Are they close or not?

### Creating a Simulation

The simulation uses the command =RAND() which generates a random number between 0 and 1.

5. In the sheet *Samples\_50*, the simulation has not been started, but the space has been laid out.
- (a) Start the simulation in Cell B6 by typing =20\*RAND(). What does this command do? (Look at the explanation of =RAND() in the box above. What does the 20\* do?)
  - (b) Copy to the right from B6 to Column AY. Finish the simulation by copying down Columns B to AY. (Nothing to hand in here.)
  - (c) Calculate the mean of each sample. Use =AVERAGE(array). (Nothing to hand in here.)
  - (d) Calculate the mean of all the sample means. (Orange box)
  - (e) Calculate the standard deviation of all the sample means. (Orange box)
  - (f) Compare the mean you calculated in part (d) with the mean of the original distribution. Are they close?
  - (g) Compare the standard deviation you calculated in part (e) with the standard deviation of the original distribution. Are they close?

### 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 these sample means? Does it increase, decrease, or stay the same?