David Youngberg
BSAD 210—Montgomery College

LECTURE 12: THE NORMAL DISTRIBUTION AND THE CLT II

I. Overview
   a. What percent of women are "average height?" What percent are "short?" What definition of "short" would result in 10 percent of women being "short?" We can answer these questions with Excel.
   b. Excel has two main functions that explicitly deal with the normal distribution: "=NORM.DIST" and "=NORM.INV".
   c. The first tells us the area under a normal distribution at a particular value. We will call this α (alpha) though Excel calls this probability.
   d. The second tells us the value that would result at a particular α.
   e. In both cases, you must provide the standard deviation and average.
      i. Height follows a normal distribution. Women are on average about 64 inches tall with a standard deviation of about 2 inches.

II. "=NORM.DIST"
   a. This function tells you the area under the curve tells you the portion of the population is that value. It has a cumulative option; right now, leave it off (0, or FALSE).
      i. What percent of women are of average height? Type "=NORM.DIST(64,64,2,0)" and press ENTER. You should get about 19.95%.[1]
      ii. Let's see how it changes as you change standard deviation. Set up the cells as you see below, with the standard deviation referencing another cell rather than being a number. (You can also click the appropriate cell while inputting information for the NORM.DIST command.)

[1] For the more mathematically inclined, this would seem like a mistake. To find this kind of information, you'd take the integral for the area under a curve (using the equation for a normal distribution). That's indeed exactly what happens when we determine the percent of observations for a particular range (e.g. percent of the population with a height from 60 to 65). But it can't work here because there's no range. It would be like finding the area of a rectangle when the width is zero. So what's going on? The answer is that Excel is programmed well. The programmers know that a precise estimate, when cumulative is set to FALSE or 0, doesn't apply here. So when you set cumulative as FALSE or 0, Excel uses something called a probability mass function instead of its regular cumulative distribution function. (If it didn't the result would always be zero, which is quite boring!) A probability mass function is built to handle discrete (read: non-range) values like what we're doing when we try to figure out the percent of the population who has exactly average height.

	A	B	C	D
1	Std Dev	% at 64		
2	1	=NORM.DIST(64,64,A2,0)		
3	2			
4	3			
5	4			

NORM.DIST(x, mean, standard_dev, cumulative)

iii. Once you press ENTER, you can copy/paste the cell (or double-click the black square in the lower right-hand corner) and the cell reference will update.

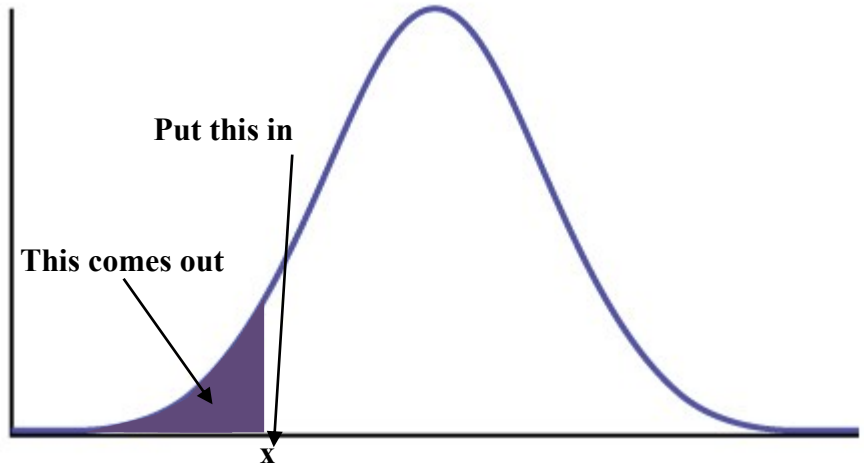
	A	B
1	Std Dev	% at 64
2	1	0.398942
3	2	
4	3	
5	4	

	A	B
1	Std Dev	% at 64
2	1	0.39894
3	2	0.19947
4	3	0.13298
5	4	0.09974

iv. Notice that as standard deviation increases, the portion of the population that's exactly the average decreases. That's because as standard deviation falls, observations are reallocated to the center to keep that bell-shaped curve. If you increase standard deviation, observations must leave the average value.

b. Cumulative Option

i. If you select the cumulative option (1, or TRUE), it will tell you the α at that value or lower.



IMPORTANT: This function will ***always*** display for the value AND LOWER. If you want to find the area for higher than the x value, subtract the result from 1.

c. We can use this technique to answer our original question: what percent of women are "short?" To answer that question, we need a definition of "short." Suppose we consider any women whose 60 inches or shorter to be "short." What percent are 60 inches tall or less?
   i. "=NORM.DIST(60,64,2,1)" does this job. Note the "1" at the end; that's because we turned cumulative "on" and Excel will now look at what percent of the curve is at 60 inches or less. You should get about 2.275%, or a little over 2 percent of women are "short."
   ii. If you think more than just over 2 percent of women are short, then that means 60 inches is too low. Let's try different cutoffs:

	A	B
1	**Cutoff**	**%**
2	60	=NORM.DIST(A2,64,2,1)
3	61	=NORM.DIST(A3,64,2,1)
4	62	=NORM.DIST(A4,64,2,1)
5	63	=NORM.DIST(A5,64,2,1)

	A	B
1	**Cutoff**	**%**
2	60	0.0227501
3	61	0.0668072
4	62	0.1586553
5	63	0.3085375

   iii. If you think something like 15 percent of women are short, then that means your definition of "short" is about 62 inches. If you think it's closer to 30 percent, your definition is about 63 inches.

d. We can also use this function to derive the Empirical Rule.
   i. We begin with what's called a standard normal distribution: a mean of zero and a standard deviation of one. Here's a table of αs for various standard deviations:

-3	-2	-1	1	2	3
0.00135	0.02275	0.158655	0.841345	0.97725	0.99865

   Remember: Excel always displays the α at the value and below.
   ii. Here is the same thing with the formulas shown:

-3	-2	-1	1	2	3
=NORM.S.DIST(-3,1)	=NORM.S.DIST(-2,1)	=NORM.S.DIST(-1,1)	=NORM.S.DIST(1,1)	=NORM.S.DIST(2,1)	=NORM.S.DIST(3,1)

   iii. Now subtract the left-side boundary from the right-side boundary (e.g. 0.841345 – 0.158655) and round:

-3	-2	-1	1	2	3
0.00135	0.02275	0.158655	0.841345	0.97725	0.99865

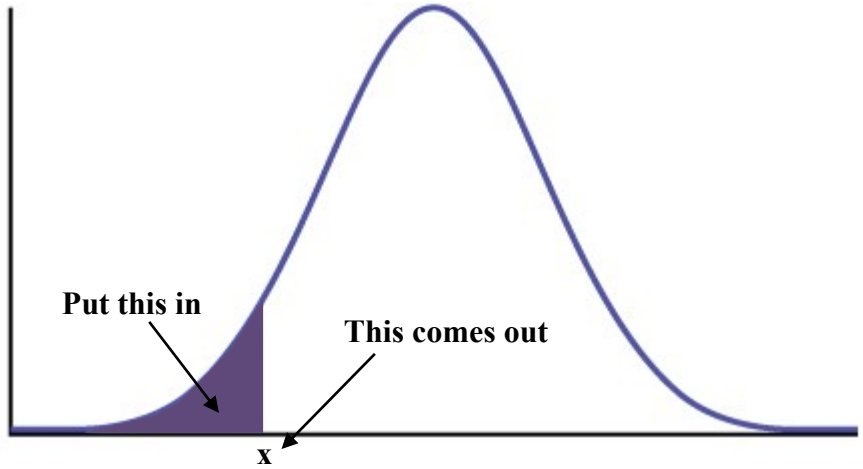
68%

95%

99.7%

III. "=NORM.INV"

a. The second function inverts the input and output; rather than putting in the x value and getting out the area under the curve, you put in the area under the curve and get out the x value.
   i. You still have to provide the mean and standard deviation.
   ii. As always, express your alpha (the area under the curve) as a decimal.



b. Suppose you think 25 percent of women are "short." What cutoff would you need to get 0.25? NORM.INV can do this:
   i. Type: =NORM.INV(0.25,64,2) You should get *about* 62.65 inches.
   ii. Note that putting in =NORM.DIST(62.65,64,2,1) gives you something very close to exactly 25%: 0.24984. If you reference the cell with the INV output, you'll get exactly 25%.

c. Imagine you work at company that makes kitchen appliances. Like all firms in this industry, you offer a warranty with your product. If it breaks within the warranty time frame, the company will replace it. But it's tricky determining how long the warranty should be:
   i. If you make the warranty too short, few people will buy the appliance. Many will think the appliance is of terrible quality.
   ii. If you make the warranty too long, you'll have to replace too many appliances and will make the entire line unprofitable.

Lifetime warranties are only good for products that last a really long time.

iii. Imagine a dishwasher lasts an average of 10 years with a standard deviation of 1.5 years. The distribution of the dishwasher lifespan is normal. The accounting department says the firm can afford to replace 4% of dishwashers sold. How long should the warranty last?

iv. We are looking for the area under a normal distribution such that the α is 0.04. Where should the x be? (Note that if a dishwasher lasts less than x, we replace it; thus we are concerned with the left-handed side of the distribution.)

v. Type "=NORM.INV(0.04,10,1.5)" and press ENTER. You should get about 7.37. You might round that down to a 7 year warranty. (But don't round it up! Then you'll be replacing more than 4% of dishwashers!)

d. Suppose you manage a sales team and you want to offer a bonus to the best salespeople. If you set the threshold for a bonus too high, no one will get the bonus and will resent you for setting a goal no one could make. If you set it too low, it will be too easy and people won't work any harder. Indeed, the naturally hard workers will resent you for giving their lazier colleagues a bonus as well.

i. Imagine the average salesperson makes $40,000 in sales per month with a standard deviation of $5,000. Sales follow a normal distribution.

ii. Imagine you want to give 20% of your workforce a bonus. What's the sales target to get the bonus?

iii. Now you want to award the TOP salespeople; you're curious about the area under the curve but on the right-handed tail. If you want to reward the top 20%, then you want to not reward the bottom 80%. Remember, Excel always outputs the bottom of the distribution, even if you define the bottom to be large.

iv. Type "=NORM.INV(0.8,40000,5000)" and press ENTER. You should get about 44,208.11; best to round this up to $45,000; if you round down, you'll give bonuses to more than 20%.

v. Note that in this case, you might still give bonuses to more than 20%, even after rounding up. That's because the promise of a bonus will encourage people to work harder. Maybe you'd want make the cut-off $46,000 (though if there are more sales, you can afford more bonuses), but this gives you a starting point.