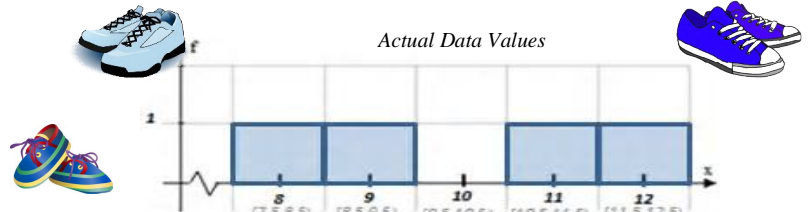


The Central Limit Theorem offers us the opportunity to make substantial statistical predictions about the population based on the sample. To better understand the relationship between samples and populations let's consider a controlled situation with a population that is very small and look at possible samples.

Consider the following shoe sizes of 4 students to represent an entire population: **8, 9, 11, 12**. The population mean (μ) of these four data points is 10. Consider looking at every possible sample of 2 from the group (allowing repetition of data points). The following samples of 2 are possible. The population distribution is shown at the right.



FIND σ :

x	μ	$x-\mu$	DEV^2
8	10	-2	4
9	10	-1	1
11	10	1	1
12	10	2	4

$\frac{4+1+1+4}{4} = \frac{10}{4} = 2.5 = \text{VARIANCE}$

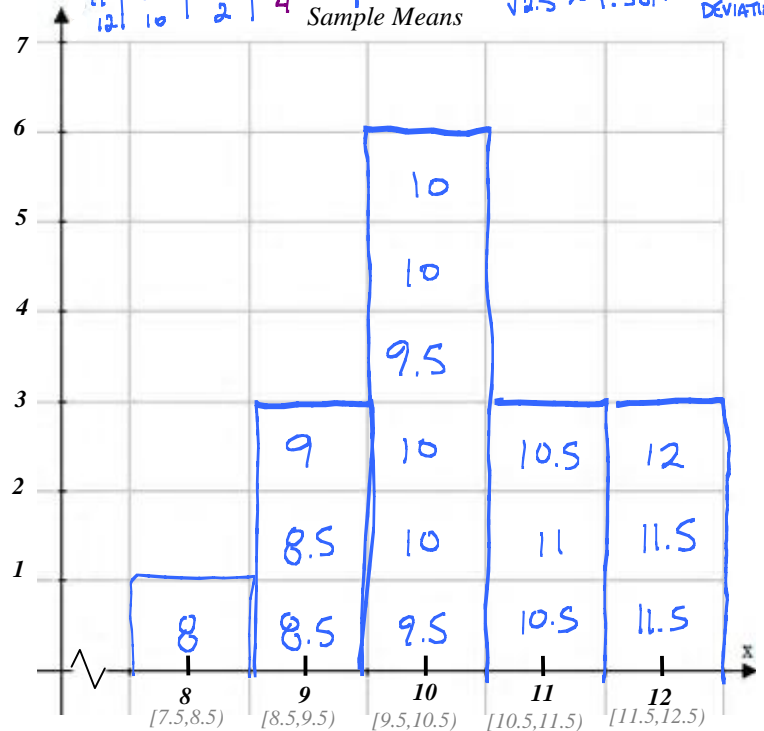
$\sqrt{2.5} \approx 1.581 \approx \text{STANDARD DEVIATION}$

Find the mean of each 2 number samples (i.e. $n = 2$)

1. {8,8}: $\bar{x} = 8$ X	9. {11,8}: $\bar{x} = 9.5$ ✓
2. {8,9}: $\bar{x} = 8.5$ X	10. {11,9}: $\bar{x} = 10$ ✓
3. {8,11}: $\bar{x} = 9.5$ ✓	11. {11,11}: $\bar{x} = 11$ ✓
4. {8,12}: $\bar{x} = 10$ ✓	12. {11,12}: $\bar{x} = 11.5$ X
5. {9,8}: $\bar{x} = 8.5$ X	13. {12,8}: $\bar{x} = 10$ ✓
6. {9,9}: $\bar{x} = 9$ ✓	14. {12,9}: $\bar{x} = 10.5$ ✓
7. {9,11}: $\bar{x} = 10$ ✓	15. {12,11}: $\bar{x} = 11.5$ X
8. {9,12}: $\bar{x} = 10.5$ ✓	16. {12,12}: $\bar{x} = 12$ X

Create a frequency histogram at the right, using data points from the table above.

(Reminder: In the histogram at the right the first class boundary would include the data points in the range $8.5 \leq x < 9.5$ which suggests 9.5 would not be included in the first bar of the histogram)



- Find the percentage of the 16 means from ~~problem #3~~ that are at most 1 unit from the mean (i.e. $9 \leq x \leq 11$). This would represent the percentage of the times a random representative sample would be within 1 unit from the mean (or the percentage of the times a sample of 2 would be reasonably accurate for our population).

$$\frac{10}{16} = .625 = \boxed{62.5\%}$$

- What is the mean ($\mu_{\bar{x}}$) and standard deviation ($\sigma_{\bar{x}}$) of the sample means of size 2 from problem

#3 and verify that these formulas hold true for this data set ($\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$)

$$\mu_{\bar{x}} = 10$$

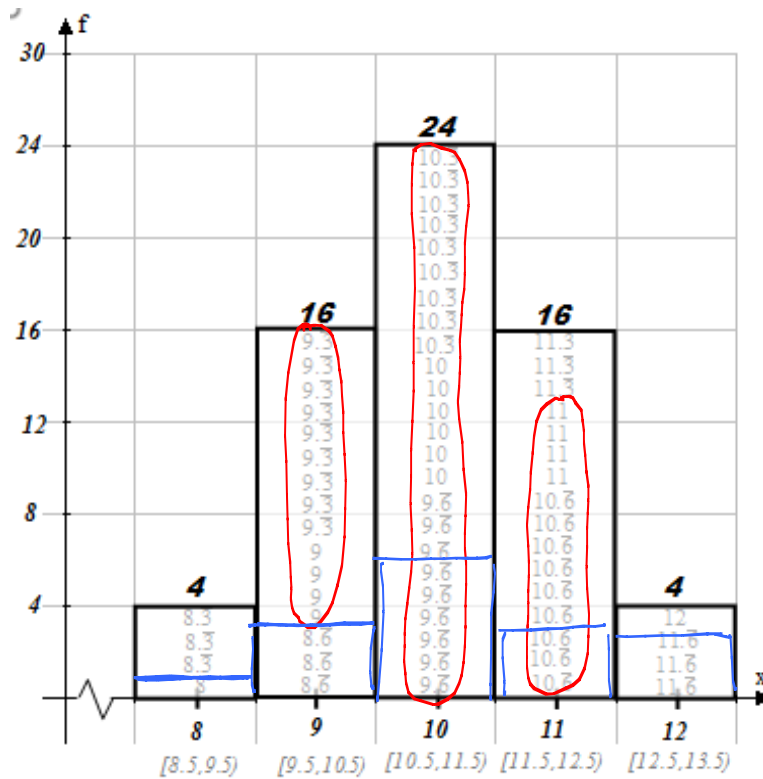
$$\sigma_{\bar{x}} \approx 1.118$$

$$\mu = 10$$

$$\sigma_{\bar{x}} = \frac{1.581}{\sqrt{2}} = \frac{1.581}{1.414} \approx 1.118$$

3. Consider using the same population of **8, 9, 11,12** but this time using all of the possible samples of size 3 would yield the following.

1. {8,8,8}: $\bar{x} = 8$	9. {8,11,8}: $\bar{x} = 9$	17. {9,8,8}: $\bar{x} = 8.\bar{3}$	25. {9,11,8}: $\bar{x} = 9.\bar{3}$	33. {11,8,8}: $\bar{x} = 9$	41. {11,11,8}: $\bar{x} = 10$	49. {12,8,8}: $\bar{x} = 9.\bar{3}$	57. {12,11,8}: $\bar{x} = 10.\bar{3}$
2. {8,8,9}: $\bar{x} = 8.\bar{3}$	10. {8,11,9}: $\bar{x} = 9.\bar{3}$	18. {9,8,9}: $\bar{x} = 8.\bar{6}$	26. {9,11,9}: $\bar{x} = 9.\bar{6}$	34. {11,8,9}: $\bar{x} = 9.\bar{3}$	42. {11,11,9}: $\bar{x} = 10.\bar{3}$	50. {12,8,9}: $\bar{x} = 9.\bar{6}$	58. {12,11,9}: $\bar{x} = 10.\bar{6}$
3. {8,8,11}: $\bar{x} = 9$	11. {8,11,11}: $\bar{x} = 10$	19. {9,8,11}: $\bar{x} = 9.\bar{3}$	27. {9,11,11}: $\bar{x} = 10.\bar{3}$	35. {11,8,11}: $\bar{x} = 10$	43. {11,11,11}: $\bar{x} = 11$	51. {12,8,11}: $\bar{x} = 10.\bar{3}$	59. {12,11,11}: $\bar{x} = 11.\bar{3}$
4. {8,8,12}: $\bar{x} = 9.\bar{3}$	12. {8,11,12}: $\bar{x} = 10.\bar{3}$	20. {9,8,12}: $\bar{x} = 9.\bar{6}$	28. {9,11,12}: $\bar{x} = 10.\bar{6}$	36. {11,8,12}: $\bar{x} = 10.\bar{3}$	44. {11,11,12}: $\bar{x} = 11.\bar{3}$	52. {12,8,12}: $\bar{x} = 10.\bar{6}$	60. {12,11,12}: $\bar{x} = 11.\bar{6}$
5. {8,9,8}: $\bar{x} = 8.\bar{3}$	13. {8,12,8}: $\bar{x} = 9.\bar{3}$	21. {9,9,8}: $\bar{x} = 8.\bar{6}$	29. {9,12,8}: $\bar{x} = 9.\bar{6}$	37. {11,8,8}: $\bar{x} = 9.\bar{3}$	45. {11,12,8}: $\bar{x} = 10.\bar{3}$	53. {12,9,8}: $\bar{x} = 9.\bar{6}$	61. {12,12,8}: $\bar{x} = 10.\bar{6}$
6. {8,9,9}: $\bar{x} = 8.\bar{6}$	14. {8,12,9}: $\bar{x} = 9.\bar{6}$	22. {9,9,9}: $\bar{x} = 9$	30. {9,12,9}: $\bar{x} = 10$	38. {11,8,9}: $\bar{x} = 9.\bar{6}$	46. {11,12,9}: $\bar{x} = 10.\bar{6}$	54. {12,9,9}: $\bar{x} = 10$	62. {12,12,9}: $\bar{x} = 11$
7. {8,9,11}: $\bar{x} = 9.\bar{3}$	15. {8,12,11}: $\bar{x} = 10.\bar{3}$	23. {9,9,11}: $\bar{x} = 9.\bar{6}$	31. {9,12,11}: $\bar{x} = 10.\bar{6}$	39. {11,8,11}: $\bar{x} = 10.\bar{3}$	47. {11,12,11}: $\bar{x} = 11.\bar{3}$	55. {12,9,11}: $\bar{x} = 10.\bar{6}$	63. {12,12,11}: $\bar{x} = 11.\bar{6}$
8. {8,9,12}: $\bar{x} = 9.\bar{6}$	16. {8,12,12}: $\bar{x} = 10.\bar{6}$	24. {9,9,12}: $\bar{x} = 10$	32. {9,12,12}: $\bar{x} = 11$	40. {11,8,12}: $\bar{x} = 10.\bar{6}$	48. {11,12,12}: $\bar{x} = 11.\bar{6}$	56. {12,9,12}: $\bar{x} = 11$	64. {12,12,12}: $\bar{x} = 12$



Recreate your histogram from problem #3 on top of the completed histogram above.

4. Find the percentage of the 64 means from problem 3 that are at most 1 unit from the mean (i.e. $9 \leq \bar{x} \leq 11$). This would represent the percentage of the times a random representative sample would be within 1 unit from the mean (or the percentage of the times a sample of 3 would be reasonably accurate for our population).

$$13 + 24 + 13 = 50 \quad \frac{50}{64} = .78125 \approx 78.1\%$$

5. How does question 4 compare with the question #4? AS WE INCREASED THE SAMPLE SIZE FROM $n=2$ TO $n=3$, WE INCREASED THE PROBABILITY OF OBTAINING A GOOD RANDOM SAMPLE FROM 63% TO 78%.
6. What is the mean ($\mu_{\bar{x}}$) and standard deviation ($\sigma_{\bar{x}}$) of the sample means of size 3 from problem 3 using the

formulas ($\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$)?

$$\mu = 10 \quad \sigma \approx 1.581$$

$$\mu_{\bar{x}} = 10 \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{1.581}{\sqrt{3}} \approx 0.91$$

7. Describe and compare the frequency distributions between the population, the sample means of size $n = 2$, the sample means of size $n = 3$.

AS WE INCREASED THE SAMPLE SIZE FROM $n=2$ TO $n=3$, THE DISTRIBUTION BECAME MORE NORMALLY DISTRIBUTED, TALLER, AND MORE NARROW.

THE CENTRAL LIMIT THEOREM

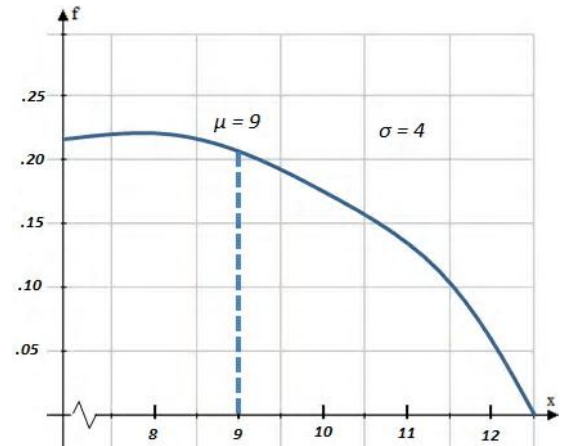
- If samples of size n , where $n \geq 30$, are drawn from any population with a mean (μ) and a standard deviation (σ), then the sampling distribution of sample means approximates a normal distribution. The greater the sample size, the better the approximation.
- If the population itself is normally distributed, the sampling distribution of the sample means is normally distributed for any sample size n .
- For all distributions
 - The mean of the sample means is equal to the population mean (i.e. $\mu_{\bar{x}} = \mu$)
 - The standard deviation of the sample means can be described by the formula ($\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$).

8. The following shows the population distribution of the number of hours of sleep elementary children were getting at one school. Assume sample size of 81 is drawn from the population. Decide which of the graphs would most closely resemble the sampling distribution of the sample means for the graph. Explain your reasoning.

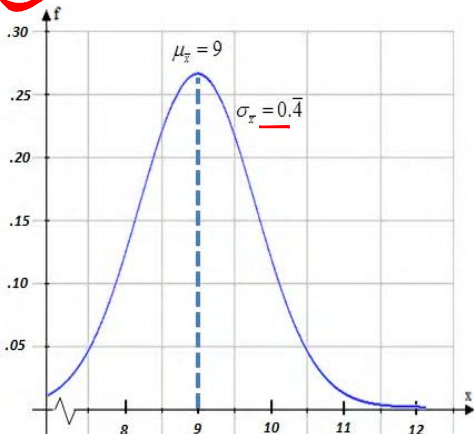
$n \geq 30$

$$\mu_{\bar{x}} = \mu = 9$$

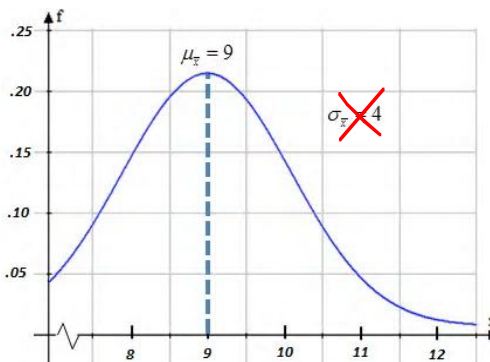
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{81}} = \frac{4}{9} = 0.44$$



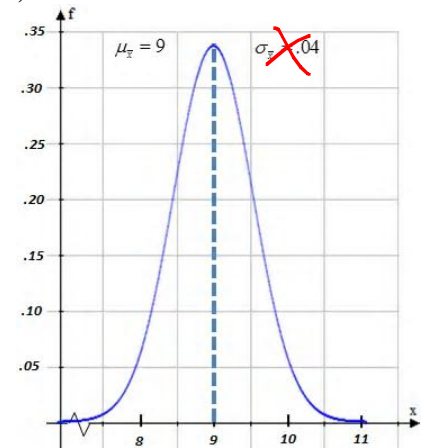
(A)



(B)



(C)



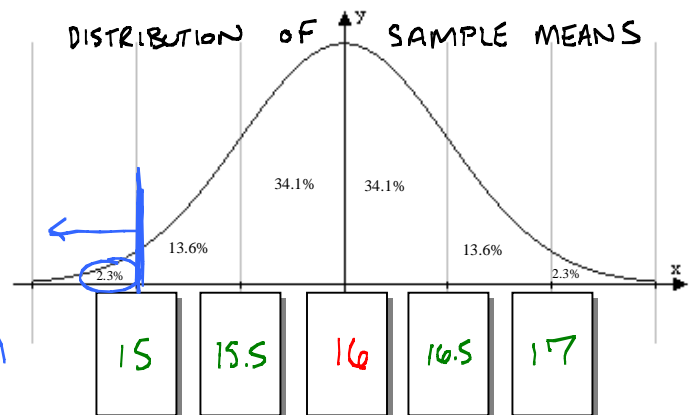
9. Use the central limit theorem if possible. For a sample of $n = 36$, find the probability of a sample mean being **less than 15** if $\mu = 16$ and $\sigma = 3$. Nothing is known about the population distribution.

$$\mu_{\bar{x}} = \mu = 16$$

$n \geq 30$ ✓

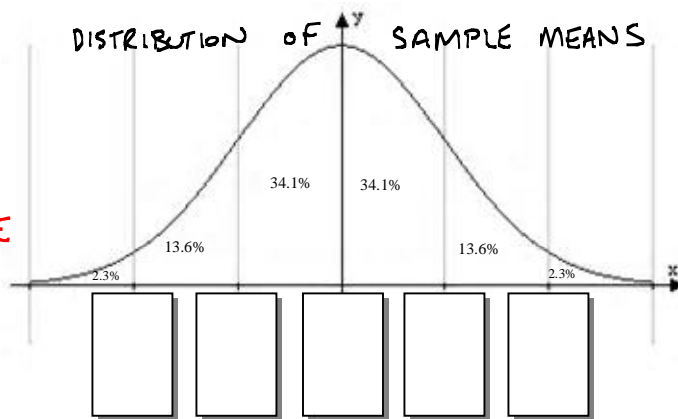
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = \frac{3}{6} = 0.5$$

THERE IS A 2.3% CHANCE OF A RANDOM SAMPLE OF 36 WILL HAVE A MEAN LESS THAN 15.



10. Use the central limit theorem if possible. For a sample of $n = 4$, find the probability of a sample mean being greater than 9.2 if $\mu = 8$ and $\sigma = 2$. The population distribution is uniformly distributed.

WE CANNOT APPLY THE CENTRAL LIMIT THEROM CONFIDENTLY BECAUSE THE SAMPLE IS LESS THAN 30 AND THE POPULATION DISTRIBUTION IS NOT NORMALLY DISTRIBUTED.

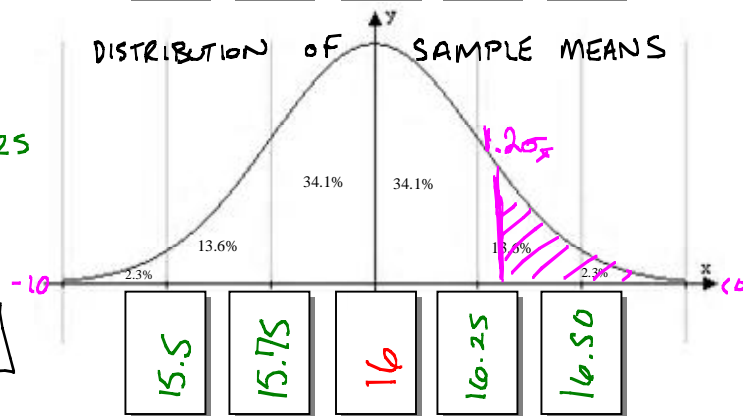


11. Use the central limit theorem if possible. For a sample of $n = 16$, find the probability of a sample mean being greater than 16.3 if $\mu = 16$ and $\sigma = 1$. The population distribution is normally distributed. ✓

$$\mu_{\bar{x}} = \mu = 16 \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{16}} = \frac{1}{4} = 0.25$$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{16.3 - 16}{.25} = 1.2$$

$$\text{NORMCDF}(1.2, 10) \approx 0.115 \text{ OR } \boxed{11.5\%}$$



12. True / False.

FALSE The distribution of sample means is always normally distributed. NOT UNLESS WE KNOW $n \geq 30$ OR THE POPULATION IS NORMAL.

TRUE Regardless of the sample size the mean of the sample means is always the same. ← AS LONG AS THE SAME POPULATION IS BEING SAMPLED
 $\mu_{\bar{x}} = \mu$ FOR ALL DISTRIBUTIONS & SAMPLE SIZES

13. Consider the following situations and using the **CENTRAL LIMIT THEOREM** find the MEAN of the sample means ($\mu_{\bar{x}}$) and the STANDARD DEVIATION of the sample means also known as Standard Error ($\sigma_{\bar{x}}$):

- I. Data for the parking lot of Phoenix H.S. showed that the mean year model of a car is 2004.3 with a standard deviation of 2.5



- a. If random samples of 30 cars at a time were taken, what would be the mean of all of the samples?

$$\mu_{\bar{x}} = \mu = 2004.3 \quad \boxed{\mu_{\bar{x}} = 2004.3}$$

- b. If random samples of 30 cars at a time were taken, what would be the standard deviation of all of the samples?

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{30}} \quad \boxed{\sigma_{\bar{x}} \approx 0.456}$$

- II. The mean GPA for seniors this year is 2.75 with a standard deviation of 0.42 and is approximately normally distributed.

- a. If random samples of 14 seniors at a time were taken, what would be the mean of all of the samples?

$$\mu_{\bar{x}} = \mu = 2.75 \quad \boxed{\mu_{\bar{x}} = 2.75}$$

- b. If random samples of 14 seniors at a time were taken, what would be the standard deviation of all of the samples?

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.42}{\sqrt{14}} \quad \boxed{\sigma_{\bar{x}} \approx 0.112}$$

- III. The mean speed of computer processors in the school is 2.8 ghz with standard deviation of 0.4 ghz. The data is NOT normally distributed.



- a. If random samples of 34 computers at a time were taken, what would be the mean of all of the samples?

$$\mu_{\bar{x}} = \mu = 2.8 \quad \boxed{\mu_{\bar{x}} = 2.8}$$

- b. If random samples of 34 computers at a time were taken, what would be the standard deviation of all of the samples?

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.4}{\sqrt{34}} \quad \boxed{\sigma_{\bar{x}} \approx 0.069}$$

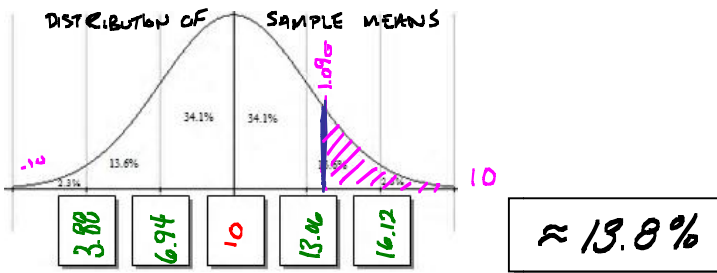
14. Consider the following situations and using the **CENTRAL LIMIT THEOREM** find the requested PROBABILITY.



I. The mean length of wait time to see a tech specialist at the APPLE store is 10 minutes with a standard deviation of 5.3 minutes. (Assuming the population data is normally distributed)

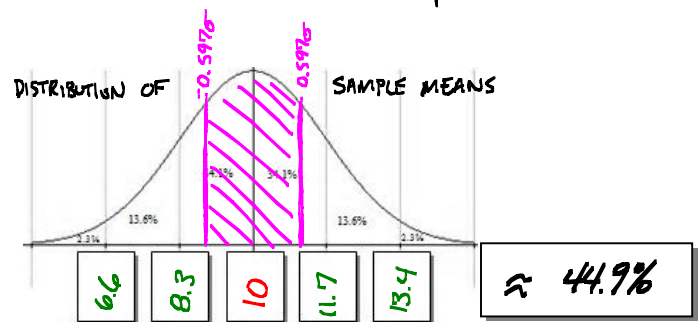
a. If those parameters are correct, what is the probability that 3 friends would have to wait more than a total of a 40 minutes to see a tech specialist? (hint: $n=3$ and their average wait time was $40/3$ minutes)

STEP #1	STEP #2	STEP #3
$\mu_{\bar{x}} = \mu = 10$	DATA POINT $\bar{x} = 13.\bar{3}$	$\text{NORMCDF}(1.09, 10)$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5.3}{\sqrt{3}}$	$Z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$	$\text{normalcdf}(1.09, 10)$
$\sigma_{\bar{x}} \approx 3.06$	$Z_{\bar{x}} = \frac{13.\bar{3} - 10}{3.06}$	≈ 0.13785
	$Z_{\bar{x}} \approx 1.09$	



b. If those parameters are correct, what is the probability of having a tech specialist log that 10 of his clients had an average wait time between 9 and 11 minutes?

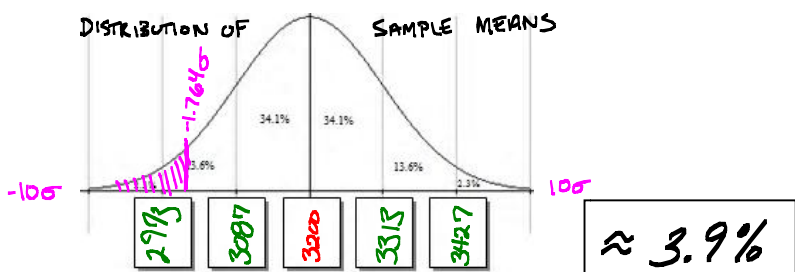
STEP #1	STEP #2	STEP #3
$\mu_{\bar{x}} = \mu = 10$	$Z_{\bar{x}_1} = \frac{9 - 10}{1.676}$	$\text{NORMCDF}(-0.597, 0.597)$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5.3}{\sqrt{10}}$	$Z_{\bar{x}_1} = -0.597$	≈ 0.449
$\sigma_{\bar{x}} \approx 1.676$	$Z_{\bar{x}_2} = \frac{11 - 10}{1.676}$	$\text{normalcdf}(-0.597, 0.597)$
	$Z_{\bar{x}_2} = 0.597$	≈ 0.449



II. Many times banks look at the mean value of a savings account over the periods of time. Over the past year, the mean value of a particular person's savings account is \$3200 with a standard deviation of \$300. (Assuming the population data is normally distributed)

a. If those parameters are correct, what is the probability that for 7 randomly selected days of the year, the average value over those 7 days is less than \$3000?

STEP #1	STEP #2	STEP #3
$\mu_{\bar{x}} = \mu = 3200$	$Z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$	$\text{NORMCDF}(-1.0, -1.764)$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{300}{\sqrt{7}}$	$Z_{\bar{x}} = \frac{3000 - 3200}{113.39}$	$\text{normalcdf}(-1.0, -1.7638)$
$\sigma_{\bar{x}} \approx 113.39$	$Z_{\bar{x}} \approx -1.764$	≈ 0.03888



b. If those parameters are correct, what is the probability that for 5 randomly selected days of the year, the average value over those 5 days is between \$3300 and \$3500?

STEP #1	STEP #2	STEP #3
$\mu_{\bar{x}} = \mu = 3200$	$Z_{\bar{x}_1} = \frac{3300 - 3200}{134.16}$	$\text{NORMCDF}(-0.745, 2.236)$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{300}{\sqrt{5}}$	$Z_{\bar{x}_1} \approx 0.745$	$\text{normalcdf}(-0.745, 2.236)$
$\sigma_{\bar{x}} \approx 134.16$	$Z_{\bar{x}_2} = \frac{3500 - 3200}{134.16}$	≈ 0.21546
	$Z_{\bar{x}_2} \approx 2.236$	

