

## CHAPTER 7: Central Limit Theorem: CLT for Averages (Means)

$X$  = the number obtained when rolling one six sided die once.

If we roll a six sided die once, the mean of the probability distribution is \_\_\_\_\_

$X$	$P(X = x)$

**Simulation:** We simulated rolling a six sided die 100 times using our calculators.

$\bar{X}$  = the sample average for a sample of 100 rolls of one die

When we each simulated rolling a six sided die 100 times in class and found the average for each students' sample of 100 rolls, the mean of the sample averages was \_\_\_\_\_

Compare the spreads of the probability distribution for rolling one die once to the probability distribution of averages from samples of rolling the die 100 times.

Which has more spread?

Which is more concentrated about the mean?

**Suppose that we have a large population with mean  $\mu$  and standard deviation  $\sigma$ .**

**Suppose that we select random samples of size  $n$  items this population.**

**Each sample taken from the population has its own average  $\bar{X}$ .**

**The sample average for any specific sample may not equal the population average exactly.**

- **The sample averages  $\bar{X}$  follow a probability distribution of their own.**

- **The average of the sample averages is the population average:**

$$\mu_{\bar{X}} = \mu$$

- **The standard deviation of the sample averages equals the population standard deviation divided by the square root of the sample size**

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{\text{sample size}}}$$

- **The shape of the distribution of the sample averages  $\bar{X}$  is normally distributed**

**IF the sample size is large enough**

**OR**

**IF the original population is normally distributed**

- **The larger the sample size, the closer the shape of the distribution of sample averages becomes to the normal distribution.**

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

**This is called the \_\_\_\_\_**

Amazingly, this means that even if we don't know the distribution of individuals in the original population, as the sample size grows large we can assume that the sample average follows a normal distribution.

We can find probabilities for sample averages using the normal distribution,

- even if the original population is not normally distributed.
- even if we don't know the shape of the distribution of the original population.

- **CHAPTER 7: Central Limit Theorem: CLT for Averages (Means)**

*Class examples selected from those below; some but not all problems or parts will be done in class.*

**EXAMPLE 1. Biology:** A biologist finds that the lengths of adult fish in a species of fish he is studying follow a normal distribution with a mean of 20 inches and a standard deviation of 2 inches.

- Find the probability that an individual adult fish is between 19 and 21 inches long.
- Find the probability that for a sample of 4 adult fish, the average length is between 19 and 21 inches.
- Find the probability that for a sample of 16 adult fish, the average length is between 19 and 21 inches.
- Sketch the graphs of the probability distributions for a, b, and c on the same axes showing how the shape of the distribution changes as the sample size changes.

**OPTIONAL EXAMPLE 2. Percentiles for sample means:** A biologist finds that the lengths of adult fish in a species of fish he is studying follow a normal distribution with a mean of 20 inches and a standard deviation of 2 inches.

- Find the 80<sup>th</sup> percentile of individual adult fish lengths and write a sentence interpreting the 80<sup>th</sup> percentile.
- Find the 80<sup>th</sup> percentile of average fish lengths for samples of 16 adult fish and complete the interpretation.

**Interpretation:** If we were to take repeated samples of 16 fish, 80% of all possible samples of 16 fish would have average lengths of less than \_\_\_\_\_ inches.

## CHAPTER 7: Central Limit Theorem: CLT for Averages (Means)

*Class examples selected from those below; some but not all problems or parts will be done in class.*

### EXAMPLE 3. Central Limit Theorem with Exponential Distribution:

Emergency services such as "911" monitor the time interval between calls received. Suppose that in a city, the time interval between calls to "911" has an exponential distribution, with an average of 5 minutes.

- a. Find the probability that the time interval until the next call is between 4 and 6 minutes.
  
  
  
  
  
  
  
  
  
  
- b. Find the probability that the sample average time interval is between 4 and 6 minutes, for sample size  $n = 36$ .
  
  
  
  
  
  
  
  
  
  
- c. Find the probability that the sample average time interval is between 4 and 6 minutes, for sample size  $n = 64$ .
  
  
  
  
  
  
  
  
  
  
- d. Sketch the graphs of the probability distributions for a, b, and c on the same axes showing how the shape of the distribution changes as the sample size changes.

**OPTIONAL EXAMPLE 4.** Suppose that in a certain city, the time interval between "911" calls has an exponential distribution, with an average of 5 minutes.

- a. Find the probability that the time interval until the next call is less than 3 minutes.
  
  
  
  
  
  
  
  
  
  
- b. Find the probability that the sample average time interval is less than 3 minutes for sample size  $n = 64$ .  
*Round the probability to 5 decimal places.*

## CHAPTER 7: Central Limit Theorem: CLT for Averages (Means)

*Class examples selected from those below; some but not all problems or parts will be done in class.*

**EXAMPLE 5. Central Limit Theorem with Uniform Distribution:**

The ages of students riding school buses in a large city are uniformly distributed between 6 and 16 years old.

- Find the probability that one randomly selected student who rides the school bus is between 10 and 12 years old.
- Find the probability that the average age is between 10 and 12 years for a random sample of 30 students who ride school buses.
- Sketch the graphs of the distributions for a and b on one set of axes.

**EXAMPLE 6. Environmental Science:** Power plants and industrial processes use water from sources such as rivers to regulate temperature. Water is taken from the river, run through pipes to cool the power or production process, and then released (clean) back into the river. The temperature of released water is monitored closely. Fish and plants living in the river are very sensitive to the water temperature; small differences can affect survival.

Suppose that water used to cool an industrial process is released into a river; the temperature of the released water has an unknown skewed right distribution with average temperature  $14.1^{\circ}\text{C}$  and standard deviation of  $2.5^{\circ}\text{C}$ .

- Explain why you can't find the probability that the water released into the river is more than  $15^{\circ}\text{C}$
- Find the probability that for a sample of 42 days, the average temperature of released water is more than  $15^{\circ}\text{C}$ .

## CHAPTER 7: Central Limit Theorem: CLT for Averages (Means)

### Conceptual Questions about the Central Limit Theorem

Central Limit Theorem questions can take the form of calculation questions or concept questions. As we did some of the calculations in Examples 1 – 6, we discussed the concepts. Try to answer these questions that ask about understanding the concepts but do not need calculations.

**EXAMPLE 7.** Explain what happens to the mean of  $\bar{X}$  as the sample size increases.

**EXAMPLE 8a.** Explain what happens to the standard deviation of  $\bar{X}$  as the sample size increases.

**8b.** Explain how this shows in the graph of the sample averages  $\bar{X}$  as the sample size increases.

**EXAMPLE 9.** What happens to the shape of the distribution when you look at sample averages instead of individuals?

**EXAMPLE 10.** Refers to Example 6 (power plants): Suppose that water used to cool an industrial process is released into a river; the temperature of the released water follows an unknown distribution that is skewed to the right with an average temperature of  $14.1^{\circ}\text{C}$  with a standard deviation of  $2.5^{\circ}\text{C}$ . For samples of 60 days, would the shape of the probability distribution for sample averages be skewed to the right, like the original distribution of temperatures on individual days? Explain why or why not.

#### How large does the sample size need to be in order to use the Central Limit Theorem?

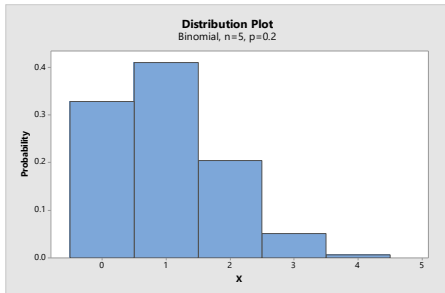
The value of  $n$  needed to be a "large enough" sample size depends on the shape of the original distribution of the individuals in the population

- If the individuals in the original population follow a normal distribution, then the sample averages will have a normal distribution, no matter how small or large the sample size is.
- If the individuals in the original population ( $X$ ) do not follow a normal distribution, then the sample averages  $\bar{X}$  become more normally distributed as the sample size grows larger. In this case the sample averages  $\bar{X}$  do not follow the same distribution as the original population.
- The more skewed the original distribution of individual values, the larger the sample size needed. If the original distribution is symmetric, the sample size needed can be smaller.
- Many statistics textbooks use the rule of thumb  $n \geq 30$ , considering 30 as the minimum sample size to use the Central Limit Theorem. But in reality there is not a universal minimum sample size that works for all distributions; the sample size needed depends on the shape of the original distribution.
- In your homework in chapter 7, assume the sample size is large enough for the Central Limit Theorem to be used to find probabilities for  $\bar{X}$ .

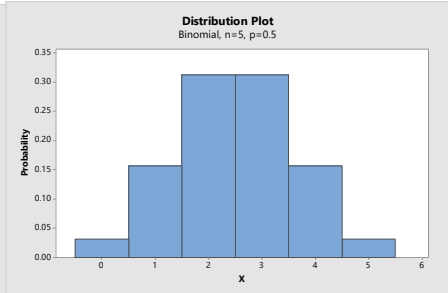
## CHAPTER 7: Central Limit Theorem for Proportions

The shape of the binomial distribution begins to approach a normal distribution as the sample size gets larger, with a mean of  $\mu = np$  and  $\sigma = \sqrt{npq}$

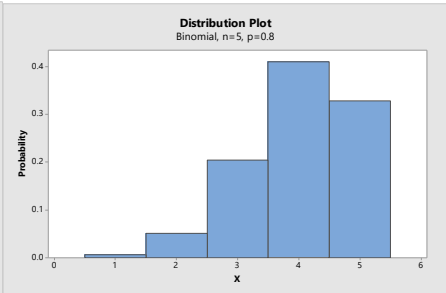
**B(5, 0.2)**



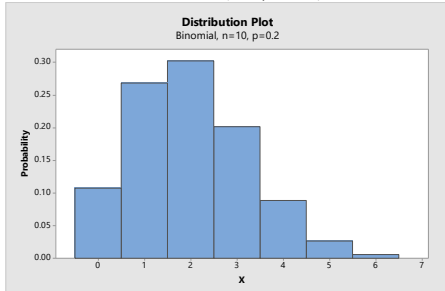
**B(5, 0.5)**



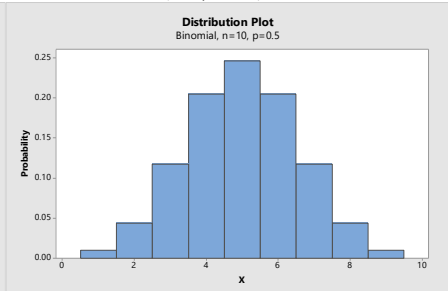
**B(5, 0.8)**



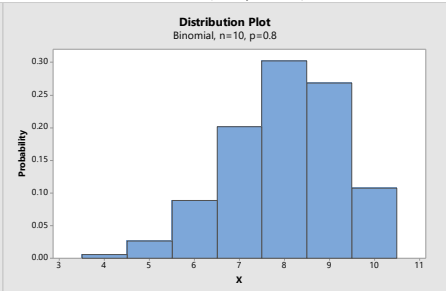
**B(10, 0.2)**



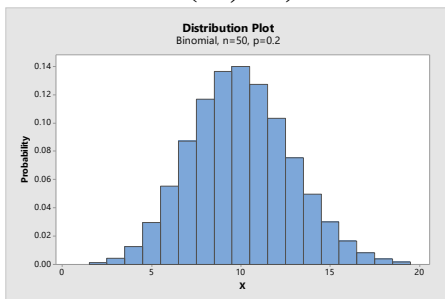
**B(10, 0.5)**



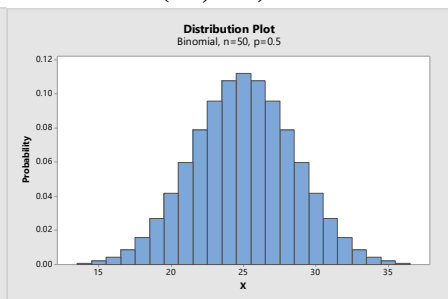
**B(10, 0.8)**



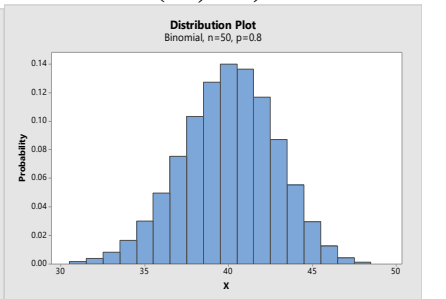
**B(50, 0.2)**



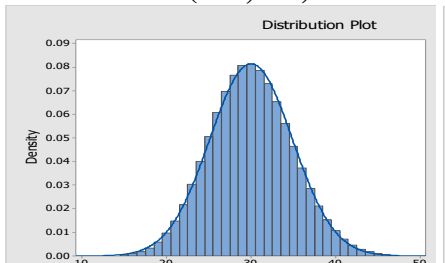
**B(50, 0.5)**



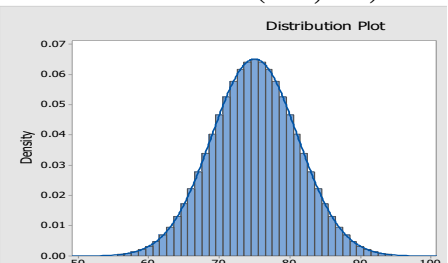
**B(50, 0.8)**



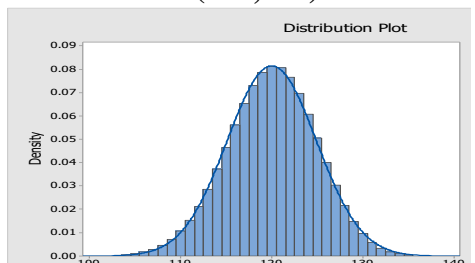
**B(150, 0.2)**



**B(150, 0.5)**



**B(150, 0.8)**



## CHAPTER 7: Central Limit Theorem for Proportions

If we have an infinite (or very large) population that can be divided into two categories:

- “success” (the thing we are interested in counting)
- “failure” (anything that is not a success).

then any particular sample selected from this population will have a sample proportion:

$$p' = \hat{p} = \frac{x}{n} = \frac{\text{number of successes in the sample}}{\text{number of items or individuals in the sample}}$$

(The symbols for the sample proportion are:  $p'$ , said as “p prime”; or  $\hat{p}$ , said as “p hat”.)

- The sample proportions are different for different samples due to randomness.
- The sample proportions from all the possible samples form a probability distribution, since some values for the sample proportion are more likely to occur than others.
- This probability distribution is called the sampling distribution for proportions.

### Central Limit Theorem for Proportions

Let  $p$  be the probability of success,  $q$  be the probability of failure in the population

$$p' = \hat{p} = \frac{x}{n} = \frac{\text{number of successes in the sample}}{\text{number of items or individuals in the sample}} = \text{sample proportion}$$

**For sufficiently large samples of size  $n$  drawn from an infinite or a very large population compared to the sample size, then the sample proportion  $p'$  or  $\hat{p}$  is approximately**

**normally distributed with mean  $\mu_{\hat{p}} = p$  and standard deviation  $\sigma_{\hat{p}} = \sqrt{pq/n}$**

$$\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right) \quad p' \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

We'll use this fact when using data from a sample to estimate a proportion for a population in chapter 8.

*(Technical Note: If the population is not very large in comparison to the sample size then if we are sampling without replacement we would need use a slightly different normal distribution involving a “finite population correction factor” when calculating the standard deviation.)*

**EXAMPLE 11:** Suppose that at a large community college, 40% of students are full-time students. Random samples of 100 students are selected. For each sample we determine the proportion of students in the sample who are full-time students. This proportion will vary depending on the sample picked.

For samples of  $n = 100$  students, what is the probability distribution for the sample proportions of full time students?

**EXAMPLE 12:** Tony's Pizza & Pasta Place sells 200 pizzas every week; 30% of pizzas sold are delivered to customers and 70% are either eaten or picked up in the store by the customer. For samples of 200 pizzas we determine the proportion of pizzas that are delivered. This proportion will vary depending on the.

For samples of  $n = 200$  pizzas, what is the probability distribution for the sample proportion of pizzas that are delivered to customers?