Central Limit Theorem for Averages (Chapter 7)

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.
$\overline{X}$ = 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 with mean μ and standard deviation σ.
Suppose that we select random samples of size n items this population.
Each sample taken from the population has its own average $\overline{X}$.
The sample average for any specific sample may not equal the population average exactly.

- **The sample averages $\overline{X}$ follow a probability distribution of their own.**
- **The average of the sample averages is the population average:** $\boxed{\mu_{\overline{X}} = \mu}$
- **The standard deviation of the sample averages equals the population standard deviation divided by the square root of the sample size**

$$\boxed{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{\textit{sample size}}}}$$

- **The shape of the distribution of the sample averages $\overline{x}$ 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.**

$$\boxed{\overline{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

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.

Calculating Probabilities using the Central Limit Theorem

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

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.

a. Find the probability that an individual adult fish is between 19 and 21 inches long.

b. Find the probability that for a sample of 4 adult fish, the average length is between 19 and 21 inches

c. Find the probability that for a sample of 16 adult fish, the average length is between 19 and 21 inches.

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.

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.

a. Find the 80^{th} percentile of individual adult fish lengths and write a sentence interpreting the 80^{th} percentile.

b. Find the 80^{th} 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.

Calculating Probabilities using the Central Limit Theorem

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

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.

4. Suppose that 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.

Calculating Probabilities using the Central Limit Theorem

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

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.

a. Find the probability that one randomly selected student who rides the school bus is between 10 and 12 years old.

b. Find the probability that the average age is between 10 and 12 years for a random sample of 30 students who ride school buses.

c. 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.

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 temperature 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°C and standard deviation of 2.5°C.

a. Explain why you can't find the probability that the water released into the river is more than 15°C

b. Find the probability that for a sample of 42 days, the average temperature of released water is more than 15°C.

Conceptual Questions about the Central Limit Theorem
Practice Problems – Try these at home.

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

7. a. Explain what happens to the standard deviation of $\overline{X}$ as the sample size increases.

b. Explain in words how this shows up in the graph of the sample averages $\bar{x}$ as the sample size increases.

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

9. Refers to question 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°C with a standard deviation of 2.5°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.

<u>How large does the sample size need to be in order to use the Central Limit Theorem?</u>

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 $\overline{X}$ become more normally distributed as the sample size grows larger. In this case the sample averages $\overline{X}$ 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 $\overline{X}$.