Applications of the Central Limit Theorem

**Application 1:** Assume that the systolic blood pressure of 30-year-old males is normally distributed with mean $\mu = 122$ mmHg and standard deviation $\sigma = 10$ mmHg.

a) What is the probability that a randomly selected 30-year-old male will have a systolic blood pressure of 127 mmHg or more? (We are randomly selecting a single 30-year-old male from the parent population.)

**Solution:** Let the random variable $x$ = the systolic blood pressure of a 30-year old male.

$$P( x \geq 127 ) = \text{normalcdf}(127, 9999, 122, 10) = 0.309.$$

b) What is the probability that a randomly selected 30-year-old male will have a systolic blood pressure between 115 mmHg and 130 mmHg? (We are randomly selecting a single 30-year-old male.)

**Solution:** Let the random variable $x$ = the systolic blood pressure of a 30-year-old male.

$$P( 115 \leq x \leq 130 ) = \text{normalcdf}(115, 130, 122, 10) = 0.546.$$

c) What is the probability that a sample of 16 randomly selected 30-year-old males will have a sample mean systolic blood pressure of 127 mmHg or more? (We are selecting a sample of males from the parent population. Compare this problem and solution to the problem in part (a) above.)

**Solution:** Let the random variable $\bar{x}$ = the mean systolic blood pressure of a random sample of 30-year-old males where the sample size = 16.

Since the pd of $x$ is normal, the pd of $\bar{x}$ is normal with $\mu = 122$ mmHg and $\sigma = 10 / \sqrt{16} = 2.5$; regardless of the sample size. Therefore $P( \bar{x} \geq 127 ) = \text{normalcdf}(127, 9999, 122, 2.5 ) = 0.023.$

d) What is the probability that a sample of 100 randomly selected 30-year-old males will have a mean systolic blood pressure between 121 mmHg and 123 mmHg? (We are selecting a sample of 100 males.)

**Solution:** Let the random variable $\bar{x}$ = the mean systolic blood pressure of a simple random sample of 30-year-old males where the sample size = 100.

Since the pd of $x$ is normal, the pd of $\bar{x}$ is normal with $\mu = 122$ mmHg and $\sigma = 10 / \sqrt{100} = 1$.

$$P( 121 \leq \bar{x} \leq 123 ) = \text{normalcdf}( 121, 123, 122, 1 ) = 0.683.$$
e) Suppose the probability distribution of systolic blood pressures for 30-year old males is continuous, but **not normal** with \( \mu = 120 \) mmHg and \( \sigma = 12 \) mmHg. Find the probability that 36 randomly selected 30-year-old males have a sample mean systolic blood pressure between 117 and 123? (We are selecting a simple random sample of 30-year-old males.)

**Solution:** Let the random variable \( \bar{x} \) = the mean systolic blood pressure of a random sample of 30-year-old males where the sample size = 36.

Since the sample size is \( \geq 30 \), the probability distribution of \( \bar{x} \) is approximately normal with \( \mu = 120 \) and \( \sigma = \frac{12}{\sqrt{36}} = 2 \); regardless of the probability distribution of the random variable \( x \).

\[
P( 117 < \bar{x} < 123 ) = \text{normalcdf}(117, 123, 120, 2) = 0.866 .
\]

f) Same situation as described in part (e) above, except that the population consists of 425 30-year-old males who live in a Chicago suburb.

**Solution:** Let the random variable \( \bar{x} \) = the mean systolic blood pressure of a simple random sample of 30-year-old males taken from a finite parent population.

Since we are not sampling with replacement and the sample size of 36 is larger than 5% of the population size of 425, we should adjust the standard deviation of the \( \bar{x} \) sample statistic. We do this by multiplying \( \sigma = 2 \) by a standard deviation correction factor = \( \frac{N - n}{N - 1} = \frac{425 - 36}{425 - 1} = 0.91745 \). Therefore the adjusted standard deviation \( \sigma = 2 \times 0.91745 = 1.8349 \).

\[
P( 117 < \bar{x} < 123 ) = \text{normalcdf}(117, 123, 120, 1.8349) = 0.898 .
\]

**Application 2:** In the 1936 presidential election between Alf Landon and Franklin D. Roosevelt, the *Literary Digest* predicted that Landon would capture 57% of the popular vote. The Digest sent out 10,000,000 ballots and received 2,266,566 responses. Because the Digest correctly predicted the outcome of the 1916, 1920, 1924, 1928 and 1932 presidential elections, we will assume that the parent population proportion \( p = 0.57 \).

a) Find the probability that less than 50% of a random sample of 100 voters will vote for Landon.

**Solution:** Let the random variable \( \hat{p} \) equal the percent of voters in a sample who will vote for Landon. We need to find \( P( \hat{p} < 0.50 ) \) where the sample size \( n = 100 \).

The probability distribution of the \( \hat{p} \) statistic is approximately normal with \( \mu = 0.57 \) and \( \sigma = \sqrt{ \frac{0.57 \times 0.43}{100} } = 0.0495 \).

\[
P( \hat{p} < 0.50 ) = P( \hat{p} \leq 0.50 ) = \text{normalcdf}(-999, 0.50, 0.57, 0.0495) = 0.0787 .
\]
b) Find the probability that more than 60% of a random sample of 400 voters will vote for Landon.

**Solution:** Let the random variable $\hat{p}$ equal the percent of voters in a sample who will vote for Landon. We need to find $P(\hat{p} > 0.60)$ where the sample size $n = 400$.

The probability distribution of the $\hat{p}$ statistic is approximately normal with $\mu = 0.57$ and $\sigma = \sqrt{\frac{0.57 \times 0.43}{400}} = 0.02475$.

\[
P(\hat{p} > 0.60) = P(\hat{p} \geq 0.60) = \text{normacdf}(0.60, 999, 0.57, 0.02475) = 0.113.
\]

c) Find the probability that of a random sample of 2,500 voters will have between 55% to 60% of the votes for Landon.

**Solution:** Let the random variable $\hat{p}$ equal the percent of voters in a sample who will vote for Landon. We need to find $P(0.55 \leq \hat{p} \leq 0.60)$ where the sample size $n = 2,500$.

The probability distribution of the $\hat{p}$ statistic is approximately normal with $\mu = 0.57$ and $\sigma = \sqrt{\frac{0.57 \times 0.43}{2,500}} = 0.009902$.

\[
P(0.55 \leq \hat{p} \leq 0.60) = \text{normacdf}(0.55, 0.60, 0.57, 0.009902) = 0.977.
\]

d) Suppose the *New York Times* took a simple random sample of 5,000 voters and found that 2,480 of the voters were for Landon. If the *Literary Digest* is correct, is it likely that the *Time’s* poll would show only 49.6% for Landon?

**Solution:** Let the random variable $\hat{p}$ equal the percent of voters in a sample who will vote for Landon. We need to find $P(\hat{p} \leq 0.496)$ where the sample size $n = 5,000$.

The probability distribution of the $\hat{p}$ statistic is approximately normal with $\mu = 0.57$ and $\sigma = \sqrt{\frac{0.57 \times 0.43}{5,000}} = 0.007001$.

\[
\hat{p} = \frac{2,480}{5,000} = 0.496.
\]

\[
P(\hat{p} \leq 0.496) = \text{normacdf}(-999, 0.496, 0.57, 0.007001) = 0.007001.
\]

The number crunchers at the *Literary Digest* committed a fundamental sampling error. If the population proportion $p = 0.57$ as the *Literary Digest* claims, there is no way on earth that the *Times* could have gotten a simple random sample with $\hat{p} = 0.496$. 
e) George Gallup used a simple random sample of 50,000 voters and predicted that Roosevelt would capture 57% of the vote. It turned out that Roosevelt captured 61% of the popular vote and Landon captured 37%. How is it possible that the Digest's prediction was so bad and Gallup's prediction was relatively close even though the Digest had a much larger sample?

**Comment:** The confidence interval that Gallup calculated most likely did not contain the true population proportion $p = 0.61$. Recall that a confidence interval for a population $p$ is an interval on the real number line that is highly likely to contain $p$, but there is **no** guarantee the confidence actually contains $p$. Prior to election day, no human knew or could know the true value of $p = 0.61$.

The Digest's prediction was based on a voluntary sample which is naturally biased. Statistical sampling theory is based on the assumption that the sample is a simple random sample of the parent population. In a simple random sample, every member of the population has an equal chance of being included in the sample. When doing various statistical tests and calculations, we need to continually remind ourselves to pay attention to **all requirements** applicable to the test or calculation. If you ignore any statistical requirements, your conclusions which are based on various statistical calculations will most likely be **spurious**!

**Application 3:** The Gallup Organization found that 65% of adult Americans favor the death penalty for individuals convicted of murder. The 65% figure is an estimate of the true population proportion $p$ of American adults who favor the death penalty. Let us pretend that the true population proportion $p$ of American adults who favor the death penalty = 0.65 and consider a simple random sample of 1,000 American adults. Suppose only 630 adults in the sample approved of the death penalty. What is the probability that no more than 630 adults in a sample of 1,000 American adults would favor the death penalty? Based on this sample alone, can we claim that the true population proportion $p$ is less than 65%? We will use two different methods to answer these questions.

### Binomial Probability Problem

- **n** = number of Bernoulli trials = 1,000
- Probability of success = 0.65 where a success equals an adult who favors the death penalty.
- $x$ = the number of successes in 1,000 trials

\[
P( x \leq 630 ) = \text{binomcdf}( 1000, 0.65, 630 ) = 0.0984
\]

### Probability Distribution of $\hat{p}$ Statistic Problem

From the Central Limit Theorem, it follows that the pd of the $p$ statistic is approximately normal.

- $n$ = sample size = 1,000, $p = 0.65$ and $q = 0.35$
- $np = 650 \geq 10$ and $nq = 350 \geq 10$.
- $\mu$ of $\hat{p}$ statistic = 0.65
- $\sigma$ of $\hat{p}$ statistic = $\sqrt{ ( 0.65 \times 0.35 / 1,000 ) } = 0.015083$

\[
P( \hat{p} \leq 0.63 ) \approx \text{normedcdf}( -0.999, 0.63, 0.65, 0.015083 ) = 0.0924
\]

**Conclusion:** The above results indicate that the event of finding no more that 630 adults who favor the death penalty is not statistically significant. An unusual event in this case should be $2\sigma$ or more below $\mu$ or have an event probability less than 0.05. Therefore there is not enough evidence to reject Gallup's claim that 65% of adult Americans approve of the death penalty. Keep in mind that $\hat{p}$ is a random variable. If Gallup is correct, we should expect that other random samples would yield $\hat{p}$ values above and below 0.65.
**Application 4:** A person arrives at a certain bus stop each morning. The waiting time, in minutes, for a bus to arrive is uniformly distributed on the interval $[0, 15]$ . Let the random variable $x$ equal the waiting time, in minutes, for the bus to arrive.

a) Parent population parameters: $\mu = 15/2 = 7.5$ minutes and $\sigma = 15/\sqrt{12} = 4.33$ minutes.

b) Probability the waiting time is between 11 and 15 minutes $= P(11 < x < 15) = 4/15 = 0.267$.

c) Suppose waiting times on different mornings are independent. Consider the event $W$ that the waiting time is less than 5 minutes. Find the probability that event $W$ occurs 10 or more times in a 21-day period for some rider who boards the bus once each day in the 21-day period.

**Solution:** We are now dealing with a binomial probability distribution with 21 Bernoulli trials, where a success equals the event $W$ and the probability of success $= 5/15 = 0.3333$. Let the random variable $x$ equal the number of successes in a 21-day period.

$$P(x \geq 10) = 1 - P(x < 9) = 1 - \text{binomcdf}(21, 0.3333, 9) = 0.125$$

d) Find the probability the mean of a sample of 60 waiting times is less than 7 minutes. Also find the probability the mean of a sample of 60 waiting times is between 7 and 10 minutes.

**Solution:**

Let the random variable $\bar{x}$ = the sample mean of 60 waiting times. Even though the parent population is uniform, according to the Central Limit Theorem the probability distribution of $\bar{x}$ is approximately normal since the sample size is greater than 30.

Probability distribution parameters for $\bar{x}$: $\mu = 7.50$ minutes and $\sigma = 4.33/\sqrt{60} = 0.559$ min.

$$P(\bar{x} < 7) = \text{normalcdf}( -99999, 7, 7.5, 0.559) = 0.186$$
$$P(7 < \bar{x} < 10) = \text{normalcdf}(7, 10, 7.5, 0.559) = 0.814$$

e) Find the probability the mean of a sample of 100 waiting times is less than 7 minutes. Also find the probability the mean of a sample of 100 waiting times is between 7 and 10 minutes.

**Solution:** Refer to the explanation given in part (d) above.

Probability distribution parameters for $\bar{x}$: $\mu = 7.50$ minutes and $\sigma = 4.33/\sqrt{100} = 0.433$ min.

$$P(\bar{x} < 7) = \text{normalcdf}( -99999, 7, 7.5, 0.433) = 0.124$$
$$P(7 < \bar{x} < 10) = \text{normalcdf}(7, 10, 7.5, 0.433) = 0.876$$

**Comment:** Increasing the sample size from 60 to 100 caused the standard deviation of the random variable $\bar{x}$ to decrease and the two probabilities to change.
**Application 5:** In a certain large city, the number of potholes on city streets follows a Poisson distribution with 3 potholes per mile. Let the random variable \( x \) equal the number of potholes found in a five-mile stretch of road. A Poisson interval equals any five-mile stretch of road.

a) Parent population parameters: \( \lambda = \mu = 3 \times 5 = 15 \) and \( \sigma = \sqrt{\lambda} = \sqrt{15} = 3.873 \)

b) Find the probability that a five-mile stretch of road has less than 10 potholes.

Solution: \( P( x < 10) = P( x \leq 9) = \text{poissoncdf}(15, 9) = 0.0699 \)

c) Find the probability that a five-mile stretch of road has between 12 and 18 potholes inclusive.

Solution: \( P(12 \leq x \leq 18) = P(x \leq 18) - P(x \leq 11) \)
\[ = \text{poissoncdf}(15, 18) - \text{poissoncdf}(15, 11) = 0.635 \]

d) A city road inspector inspects a random sample of 30 five-mile stretches of road and records the number of potholes in each five-mile stretch of road. Find the probability that mean of the inspector's sample is less than 14.25 potholes. Also find the probability that the mean is between 14.5 and 16.5 potholes.

Solution:

Let the random variable \( \bar{x} \) equal the mean number of potholes found in a random sample of 30 five-mile stretches of road. Even though the parent population is Poisson and therefore discrete, according to the Central Limit Theorem, the probability distribution of \( \bar{x} \) is approximately normal since the sample size is 30. Because the parent population is discrete, the inspector should have obtained a larger sample size, however, a sample size of 30 should give satisfactory approximations.

Population parameters for \( \bar{x} \): \( \mu = 15 \) potholes and \( \sigma = 3.873 / \sqrt{30} = 0.7071 \) potholes

\[ P( \bar{x} < 14.25) = \text{normalcdf}( -99999, 14.25, 15, 0.7071) = 0.144 \]
\[ P(14.5 < \bar{x} < 16.5) = \text{normalcdf}( 14.5, 16.5, 15, 0.7071) = 0.743 \]

e) Refer to part (d) above. Suppose the inspector increased his sample size to 50. Recalculate the probabilities in part (d) above.

Solution:

New population parameters for \( \bar{x} \): \( \mu = 15 \) potholes and \( \sigma = 3.873 / \sqrt{50} = 0.5477 \) potholes

\[ P( \bar{x} < 14.25) = \text{normalcdf}( -99999, 14.25, 15, 0.5477) = 0.0854 \]
\[ P(14.5 < \bar{x} < 16.5) = \text{normalcdf}( 14.5, 16.5, 15, 0.5477) = 0.816 \]