Central Limit Theorem: Using the Central Limit Theorem*

Susan Dean
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the Creative Commons Attribution License †

It is important for you to understand when to use the CLT. If you are being asked to find the probability of an average or mean, use the CLT for means or averages. If you are being asked to find the probability of a sum or total, use the CLT for sums. This also applies to percentiles for averages and sums.

**NOTE:** If you are being asked to find the probability of an individual value, do **not** use the CLT. **Use the distribution of its random variable.**

1 Examples of the Central Limit Theorem

Law of Large Numbers

The **Law of Large Numbers** says that if you take samples of larger and larger size from any population, then the mean \( \bar{x} \) of the sample gets closer and closer to \( \mu \). From the Central Limit Theorem, we know that as \( n \) gets larger and larger, the sample averages follow a normal distribution. The larger \( n \) gets, the smaller the standard deviation gets. (Remember that the standard deviation for \( \bar{X} \) is \( \frac{\sigma}{\sqrt{n}} \).) This means that the sample mean \( \bar{x} \) must be close to the population mean \( \mu \). We can say that \( \mu \) is the value that the sample averages approach as \( n \) gets larger. The Central Limit Theorem illustrates the Law of Large Numbers.

Central Limit Theorem for the Mean (Average) and Sum Examples

**Example 1**

A study involving stress is done on a college campus among the students. **The stress scores follow a uniform distribution** with the lowest stress score equal to 1 and the highest equal to 5. Using a sample of 75 students, find:

1. The probability that the **average stress score** for the 75 students is less than 2.
2. The 90th percentile for the **average stress score** for the 75 students.
3. The probability that the **total of the 75 stress scores** is less than 200.
4. The 90th percentile for the **total stress score** for the 75 students.

Let \( X \) = one stress score.

Problems 1. and 2. ask you to find a probability or a percentile for an **average or mean**. Problems 3 and 4 ask you to find a probability or a percentile for a **total or sum**. The sample size, \( n \), is equal to 75.

---

*Version 1.15: Mar 8, 2010 2:48 pm US/Central
†http://creativecommons.org/licenses/by/2.0/
Source URL: http://cnx.org/content/col10522/latest/
Saylor URL: http://www.saylor.org/courses/ma121/
http://cnx.org/content/m16958/1.15/
Attributed to: Barbara Illowsky and Susan Dean*
Since the individual stress scores follow a uniform distribution, $X \sim U(1, 5)$ where $a = 1$ and $b = 5$ (See Continuous Random Variables\(^1\) for the uniform).

\[
\mu_X = \frac{a + b}{2} = \frac{1 + 5}{2} = 3
\]

\[
\sigma_X = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(5-1)^2}{12}} = 1.15
\]

For problems 1. and 2., let $\bar{X}$ = the average stress score for the 75 students. Then, $\bar{X} \sim N\left(3, \frac{1.15}{\sqrt{75}}\right)$ where $n = 75$.

**Problem 1**

Find $P(\bar{X} < 2)$. Draw the graph.

**Solution**

$P(\bar{X} < 2) = 0$

The probability that the average stress score is less than 2 is about 0.

\[
\text{normalcdf}\left(1, 2, 3, \frac{1.15}{\sqrt{75}}\right) = 0
\]

**Reminder:** The smallest stress score is 1. Therefore, the smallest average for 75 stress scores is 1.

---

**Problem 2**

Find the 90th percentile for the average of 75 stress scores. Draw a graph.

**Solution**

Let $k$ = the 90th percentile.

Find $k$ where $P(\bar{X} < k) = 0.90$.

$k = 3.2$

\(^{1}\)Continuous Random Variables: Introduction" <http://cnx.org/content/m16808/latest/>
The 90th percentile for the average of 75 scores is about 3.2. This means that 90% of all the averages of 75 stress scores are at most 3.2 and 10% are at least 3.2.

$$\text{invNorm} \left( .90, 3, \frac{1.15}{\sqrt{75}} \right) = 3.2$$

For problems c and d, let $\Sigma X$ = the sum of the 75 stress scores. Then, $\Sigma X \sim N \left( (75)(3), \sqrt{75} \cdot 1.15 \right)$

**Problem 3**
Find $P(\Sigma X < 200)$. Draw the graph.

**Solution**
The mean of the sum of 75 stress scores is $75 \cdot 3 = 225$
The standard deviation of the sum of 75 stress scores is $\sqrt{75} \cdot 1.15 = 9.96$

$$P(\Sigma X < 200) = 0$$

The probability that the total of 75 scores is less than 200 is about 0.

$$\text{normalcdf} \left( 75, 200, 75 \cdot 3, \sqrt{75} \cdot 1.15 \right) = 0.$$  

**Reminder:** The smallest total of 75 stress scores is 75 since the smallest single score is 1.

**Problem 4**
Find the 90th percentile for the total of 75 stress scores. Draw a graph.

**Solution**
Let $k$ = the 90th percentile.
Find $k$ where $P(\Sigma X < k) = 0.90$. 

Source URL: http://cnx.org/content/col10522/latest/
Saylor URL: http://www.saylor.org/courses/ma121/
http://cnx.org/content/m16958/1.15/
Attributed to: Barbara Illowsky and Susan Dean
The 90th percentile for the sum of 75 scores is about 237.8. This means that 90% of all the sums of 75 scores are no more than 237.8 and 10% are no less than 237.8.

\[ \text{invNorm (.90, 75 \cdot 3, \sqrt{75} \cdot 1.15)} = 237.8 \]

**Example 2**

Suppose that a market research analyst for a cell phone company conducts a study of their customers who exceed the time allowance included in their basic cell phone contract. The analyst finds that for those customers who exceed the time included in their basic contract, the **excess time used** follows an exponential distribution with a mean of 22 minutes. Consider a random sample of 80 customers. Find

1. The probability that the **average excess time** used by the 80 customers in the sample is longer than 20 minutes. Draw a graph.
2. The 95th percentile for the **average excess time** for samples of 80 customers who exceed their basic contract time allowances. Draw a graph.

Let \( X \) = the excess time used by one individual cell phone customer who exceeds his contracted time allowance. Then \( X \sim \text{Exp} \left( \frac{1}{22} \right) \) (see Continuous Random Variables\(^2\) for the exponential). Because \( X \) is exponential, \( \mu = 22 \) and \( \sigma = 22 \). The sample size is \( n = 80 \).

Let \( \bar{X} \) = the **average** excess time used by a sample of \( n = 80 \) customers who exceed their contracted time allowances. Then

\( \bar{X} \sim N \left( 22, \frac{22}{\sqrt{80}} \right) \) by the CLT for Sample Means or Averages

**Problem 1**

Find \( P \left( \bar{X} > 20 \right) \). Draw the graph.

**Solution**

\( P \left( \bar{X} > 20 \right) = 0.7919 \)

The probability that the average excess time used by a sample of 80 customers is longer than 20 minutes is 0.7919.

\(^2\) Continuos Random Variables: Introduction <http://cnx.org/content/m16808/latest/>
Problem 2
Find the 95th percentile for the average excess time for samples of 80 customers who exceed their basic contract time allowances. Draw a graph.

Solution
Let \( k = \) the 95th percentile for the average excess time.

Find \( k \) where \( P(\bar{X} < k) = 0.95 \)

\[
\bar{X} \sim \text{Normal}(22, \sqrt{80}/80)
\]

\[
\text{invNorm}(0.95, 22, \sqrt{80}/80) = 26.0
\]

The 95th percentile for the average excess time for samples of 80 customers who exceed their basic contract time allowances is about 26 minutes. This means that 95% of the average excess times are at most 26 minutes and 10% are at least 26 minutes.

\[
\text{invNorm}(0.95, 22, \sqrt{80}/80) = 26.0
\]
Historically, being able to compute binomial probabilities was one of the most important applications of the Central Limit Theorem. Binomial probabilities were displayed in a table in a book with a small value for \(n\) (say, 20). To calculate the probabilities with large values of \(n\), you had to use the binomial formula which could be very complicated. Using the Normal Approximation to the Binomial simplified the process. To compute the Normal Approximation to the Binomial, take a simple random sample from a population. You must meet the conditions for a binomial distribution:

- there are a certain number \(n\) of independent trials
- the outcomes of any trial are success or failure
- each trial has the same probability of a success \(p\)

Recall that if \(X\) is the binomial random variable, then \(X \sim B(n, p)\). The shape of the binomial distribution needs to be similar to the shape of the normal distribution. To ensure this, the quantities \(np\) and \(nq\) must both be greater than five \((np > 5 \text{ and } nq > 5)\); the approximation is better if they are both greater than or equal to 10. Then the binomial can be approximated by the normal distribution with mean \(\mu = np\) and standard deviation \(\sigma = \sqrt{npq}\). Remember that \(q = 1 - p\). In order to get the best approximation, add 0.5 to \(X\) or subtract 0.5 from \(X\) (use \(X + 0.5\) or \(X - 0.5\). The number 0.5 is called the continuity correction factor.

**Example 3**

Suppose in a local Kindergarten through 12th grade (K-12) school district, 53 percent of the population favor a charter school for grades K-5. A simple random sample of 300 is surveyed.

1. Find the probability that at least 150 favor a charter school.
2. Find the probability that at most 160 favor a charter school.
3. Find the probability that more than 155 favor a charter school.
4. Find the probability that less than 147 favor a charter school.
5. Find the probability that exactly 175 favor a charter school.

Let \(X\) = the number that favor a charter school for grades K-5. \(X \sim B(n, p)\) where \(n = 300\) and \(p = 0.53\). Since \(np > 5\) and \(nq > 5\), use the normal approximation to the binomial. The formulas for the mean and standard deviation are \(\mu = np\) and \(\sigma = \sqrt{npq}\). The mean is 159 and the standard deviation is 8.6447. The random variable for the normal distribution is \(Y\). \(Y \sim N(159, 8.6447)\).

See the Normal Distribution for help with calculator instructions.

For Problem 1., you include 150 so \(P(X \geq 150)\) has normal approximation \(P(Y \geq 149.5) = 0.8641\).

\[
\text{normalcdf}(149.5, 10^{-99}, 159, 8.6447) = 0.8641
\]

For Problem 2., you include 160 so \(P(X \leq 160)\) has normal approximation \(P(Y \leq 160.5) = 0.5689\).

\[
\text{normalcdf}(0, 160.5, 159, 8.6447) = 0.5689
\]

For Problem 3., you exclude 155 so \(P(X > 155)\) has normal approximation \(P(Y > 155.5) = 0.6572\).

\[
\text{normalcdf}(155.5, 10^{-99}, 159, 8.6447) = 0.6572
\]

For Problem 4., you exclude 147 so \(P(X < 147)\) has normal approximation \(P(Y < 146.5) = 0.0741\).

\[
\text{normalcdf}(0, 146.5, 159, 8.6447) = 0.0741
\]

For Problem 5., \(P(X = 175)\) has normal approximation \(P(174.5 < Y < 175.5) = 0.0083\).

\[
\text{normalcdf}(174.5, 175.5, 159, 8.6447) = 0.0083
\]

**Because of calculators and computer software** that easily let you calculate binomial probabilities for large values of \(n\), it is not necessary to use the the Normal Approximation to the Binomial provided you have access to these technology tools. Most school labs have Microsoft Excel, an example of computer software that calculates binomial probabilities. Many students have access to the TI-83 or 84 series calculators and they easily calculate probabilities for the binomial.
In an Internet browser, if you type in "binomial probability distribution calculation," you can find at least one online calculator for the binomial.

For Example 3, the probabilities are calculated using the binomial \( n = 300 \) and \( p = 0.53 \) below. Compare the binomial and normal distribution answers. See Discrete Random Variables for help with calculator instructions for the binomial.

\[
\begin{align*}
\Pr(X \geq 150) &= 1 - \text{binomialcdf}(300, 0.53, 149) = 0.8641 \\
\Pr(X \leq 160) &= \text{binomialcdf}(300, 0.53, 160) = 0.5684 \\
\Pr(X > 155) &= 1 - \text{binomialcdf}(300, 0.53, 155) = 0.6576 \\
\Pr(X < 147) &= \text{binomialcdf}(300, 0.53, 146) = 0.0742 \\
\Pr(X = 175) &= \text{binomialpdf}(175, 0.53, 146) = 0.0083
\end{align*}
\]

Glossary

**Definition 1: Average**
A number that describes the central tendency of the data. There are a number of specialized averages, including the arithmetic mean, weighted mean, median, mode, and geometric mean.

**Definition 2: Central Limit Theorem**
Given a random variable (RV) with known mean \( \mu \) and known standard deviation \( \sigma \), we are sampling with size \( n \) and we are interested in two new RVs - called the sample mean, \( \bar{X} \), and called the sample sum, \( \Sigma X \). If the size \( n \) of the sample is sufficiently large, then \( \bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \) and \( \Sigma X \sim N(n\mu, \sqrt{n}\sigma) \). In words, if the size \( n \) of the sample is sufficiently large, then the distribution of the sample means and the distribution of the sample sums will approximate a normal distribution regardless of the shape of the population. And even more, the mean of the sampling distribution will equal the population mean and the mean of sampling sums will equal \( n \) times the population mean. The standard deviation of the distribution of the sample means, \( \frac{\sigma}{\sqrt{n}} \), is called the standard error of the mean.

**Definition 3: Exponential Distribution**
Continuous random variable (RV) that appears when we are interested in intervals of time between some random events, for example, the length of time between emergency arrivals at a hospital. Notation: \( X \sim \text{Exp}(m) \); the mean is \( \mu = \frac{1}{m} \), and the standard deviation is \( \sigma = \frac{1}{m} \), the probability density function is \( f(x) = me^{-mx}, x \geq 0 \) and the cumulative distribution is \( P(X \leq x) = 1 - e^{-mx} \).

**Definition 4: Mean**
A number to measure the central tendency (average), shortened from the arithmetic mean. By definition, the mean for a sample (denoted by \( \bar{x} \)) is \( \bar{x} = \frac{\text{Sum of all values in the sample}}{\text{Number of values in the sample}} \), and the mean for a population (denoted by \( \mu \)) is \( \mu = \frac{\text{Sum of all values in the population}}{\text{Number of values in the population}} \).