



# Introduction

By:  
OpenStaxCollege



*If you want to figure out the distribution of the change people carry in their pockets, using the central limit theorem and assuming your sample is large enough, you will find that the distribution is normal and bell-shaped. (credit: John Lodder)*

## Chapter Objectives

By the end of this chapter, the student should be able to:

- Recognize central limit theorem problems.
- Classify continuous word problems by their distributions.
- Apply and interpret the central limit theorem for means.
- Apply and interpret the central limit theorem for sums.

Why are we so concerned with means? Two reasons are: they give us a middle ground for comparison, and they are easy to calculate. In this chapter, you will study means and the **central limit theorem**.

The central limit theorem (clt for short) is one of the most powerful and useful ideas in all of statistics. There are two alternative forms of the theorem, and both alternatives are

## Introduction

concerned with drawing finite samples size  $n$  from a population with a known mean,  $\mu$ , and a known standard deviation,  $\sigma$ . The first alternative says that if we collect samples of size  $n$  with a "large enough  $n$ ," calculate each sample's mean, and create a histogram of those means, then the resulting histogram will tend to have an approximate normal bell shape. The second alternative says that if we again collect samples of size  $n$  that are "large enough," calculate the sum of each sample and create a histogram, then the resulting histogram will again tend to have a normal bell-shape.

**In either case, it does not matter what the distribution of the original population is, or whether you even need to know it. The important fact is that the distribution of sample means and the sums tend to follow the normal distribution.**

The size of the sample,  $n$ , that is required in order to be "large enough" depends on the original population from which the samples are drawn (the sample size should be at least 30 or the data should come from a normal distribution). If the original population is far from normal, then more observations are needed for the sample means or sums to be normal. **Sampling is done with replacement.**

### Collaborative Classroom Activity

Suppose eight of you roll one fair die ten times, seven of you roll two fair dice ten times, nine of you roll five fair dice ten times, and 11 of you roll ten fair dice ten times.

Each time a person rolls more than one die, he or she calculates the sample mean of the faces showing. For example, one person might roll five fair dice and get 2, 2, 3, 4, 6 on one roll.

The mean is  $\frac{2+2+3+4+6}{5} = 3.4$ . The 3.4 is one mean when five fair dice are rolled. This same person would roll the five dice nine more times and calculate nine more means for a total of ten means.

Your instructor will pass out the dice to several people. Roll your dice ten times. For each roll, record the faces, and find the mean. Round to the nearest 0.5.

Your instructor (and possibly you) will produce one graph (it might be a histogram) for one die, one graph for two dice, one graph for five dice, and one graph for ten dice. Since the "mean" when you roll one die is just the face on the die, what distribution do these **means** appear to be representing?

**Draw the graph for the means using two dice.** Do the sample means show any kind of pattern?

**Draw the graph for the means using five dice.** Do you see any pattern emerging?

**Finally, draw the graph for the means using ten dice.** Do you see any pattern to the graph? What can you conclude as you increase the number of dice?

As the number of dice rolled increases from one to two to five to ten, the following is happening:

1. The mean of the sample means remains approximately the same.
2. The spread of the sample means (the standard deviation of the sample means) gets smaller.
3. The graph appears steeper and thinner.

You have just demonstrated the central limit theorem (clt).

The central limit theorem tells you that as you increase the number of dice, **the sample means tend toward a normal distribution (the sampling distribution).**