Connexions module: m16953

CENTRAL LIMIT THEOREM: INTRODUCTION*

Susan Dean Barbara Illowsky, Ph.D.

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

Abstract

This module provides a brief introduction to the Central Limit Theorem.

1 Student Learning Outcomes

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

- Recognize the 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.

2 Introduction

Why are we so concerned with means? Two reasons are that 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. Both alternatives are concerned with drawing finite samples of size n from a population with a known mean, μ , and a known standard deviation, σ . The first alternative says that if we collect samples of size n and n is "large enough," 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 sample means and the sums tend to follow the normal distribution. And, the rest you will learn in this chapter.

The size of the sample, n, that is required in order to be to be 'large enough' depends on the original population from which the samples are drawn. If the original population is far from normal then more observations are needed for the sample means or the sample sums to be normal. Sampling is done with replacement.

^{*}Version 1.17: Jun 5, 2012 6:56 am -0500

[†]http://creativecommons.org/licenses/by/3.0/

Connexions module: m16953 2

Optional Collaborative Classroom Activity

Do the following example in class: Suppose 8 of you roll 1 fair die 10 times, 7 of you roll 2 fair dice 10 times, 9 of you roll 5 fair dice 10 times, and 11 of you roll 10 fair dice 10 times.

Each time a person rolls more than one die, he/she calculates the sample **mean** of the faces showing. For example, one person might roll 5 fair dice and get a 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 5 fair dice are rolled. This same person would roll the 5 dice 9 more times and calculate 9 more means for a total of 10 means.

Your instructor will pass out the dice to several people as described above. Roll your dice 10 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 1 die, one graph for 2 dice, one graph for 5 dice, and one graph for 10 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 2 dice. Do the sample means show any kind of pattern?

Draw the graph for the means using 5 dice. Do you see any pattern emerging?

Finally, draw the graph for the means using 10 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 1 to 2 to 5 to 10, 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).

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 μ and known standard deviation σ . We are sampling with size n and we are interested in two new RVs - the sample mean, \overline{X} , and the sample sum, ΣX . If the size n of the sample is sufficiently large, then $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ and $\Sigma X \sim N\left(n\mu, \sqrt{n}\sigma\right)$. 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. The mean of the sample means will equal the population mean and the mean of the sample 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.