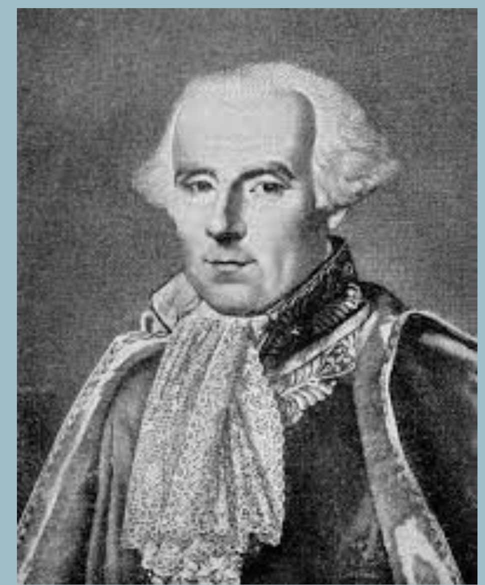


The standard version of the central limit theorem was first proved by the French mathematician **Pierre-Simon Laplace** in 1810.

Central Limit Theorem

Yufei Huang & Megan Melzer

Virginia Polytechnic Institute and State University

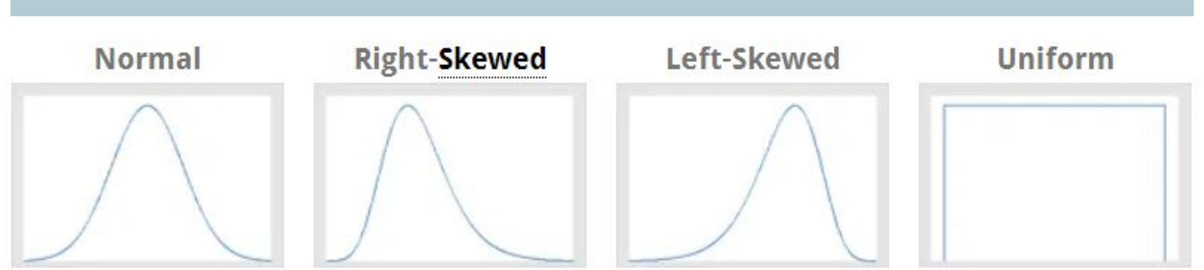
Introduction

The central limit theorem (CLT) states that if you have a population with **mean μ** and **standard deviation σ**, and a sufficiently large **sample** is randomly drawn from the population with replacement, then the distribution of the sample means will be approximately normal.[3]

$$z = \frac{\bar{x} - \mu_x}{\left(\frac{\sigma_x}{\sqrt{n}}\right)}$$

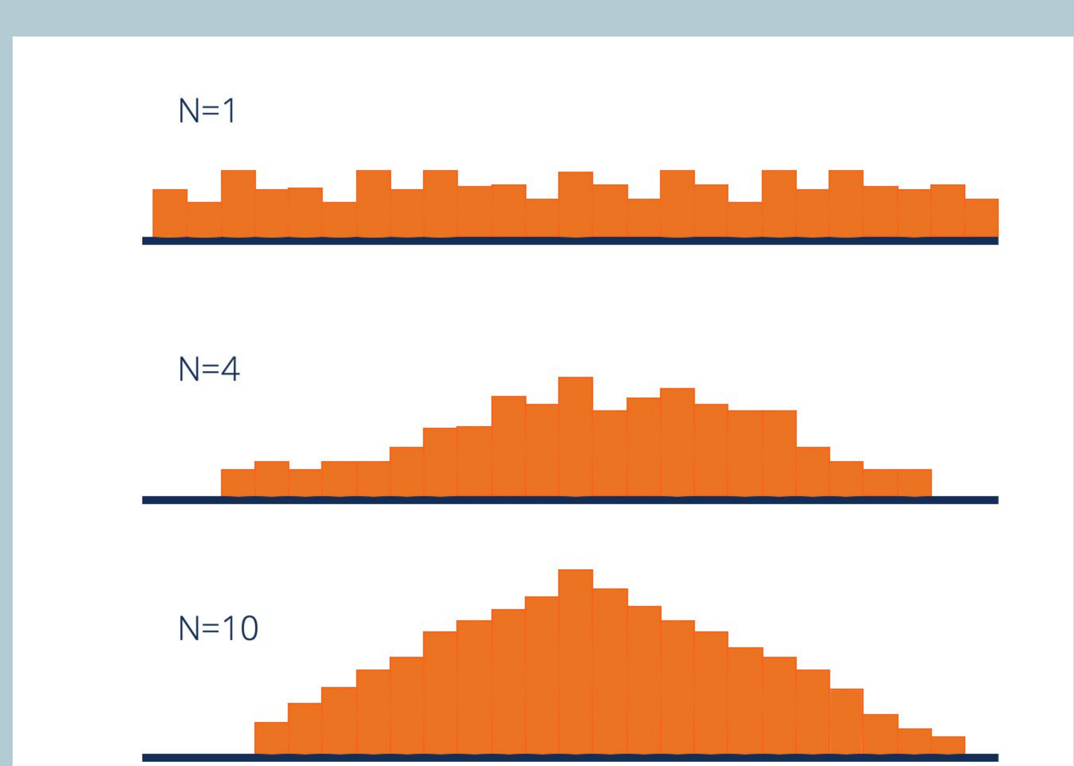
- μ is the mean
- σ is the standard deviation
- n is the sample size

The values of a variable in a population can follow different probability distributions. [6]



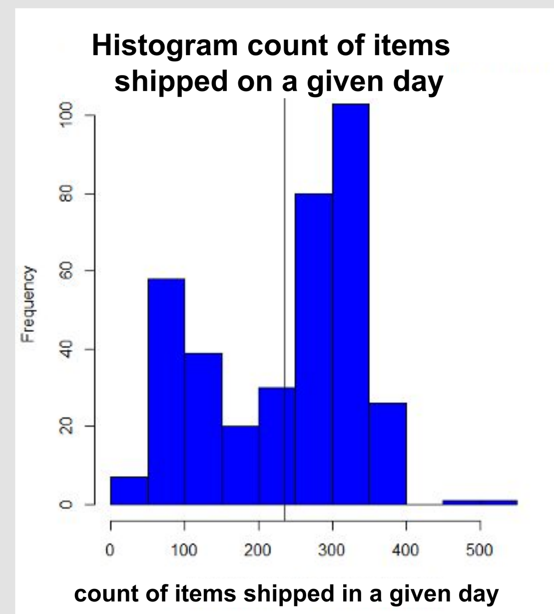
As the sample size increases, the mean distribution of the repeated samples tend to normalize, and resemble a normal distribution. [4]

The result is the same regardless of the original shape of the distribution. It can be illustrated in the following figure:

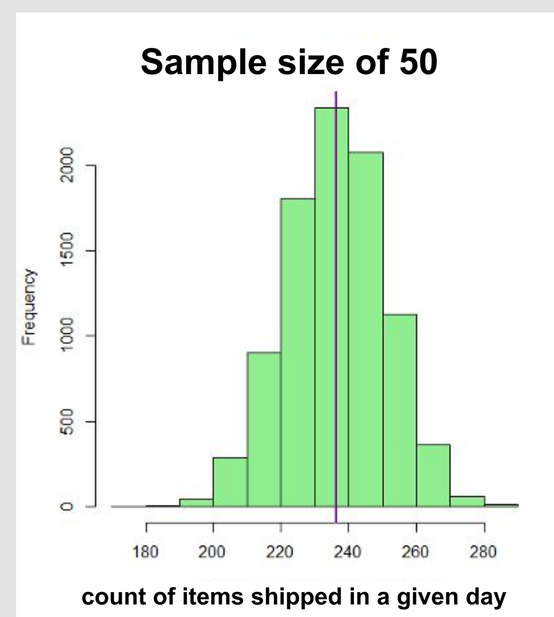


Example

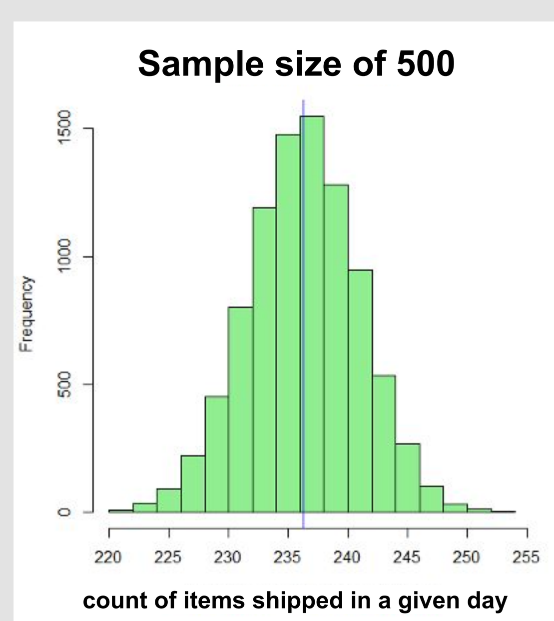
Let's take a look at a real-life, live, example! We will explore a shipment data set of a pharmaceutical company, which has hundreds of thousands of observations.



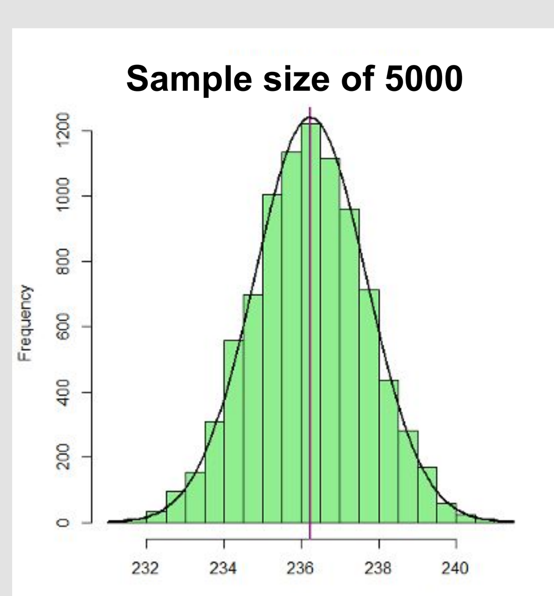
- Population distribution has a mean of about 236 and a variance of about 102
- There is no definite distribution and the spread is way too large
- The distribution is nowhere close to a normal distribution
- Let's call CLT to the rescue!



- Start with a relatively small sample size, say n = 50
- Distribution appears to begin to have characteristics of a normal distribution
- Still not perfect by any means



- Increase sample size significantly to n = 500
- Much more normally distributed than when sample size was only 50
- Still not a perfect distribution yet, as we expect when the CLT is in its fullest effect



- Increase sample size significantly to n = 5000
- A nearly perfectly normal sampling distribution is achieved

TAKE AWAY: Each time we take a sample that is sufficiently greater than the previous ones, the spread becomes tighter thereby increasing the precision of the estimate of the sample.[1]

Applications

The application of CLT is extremely broad. In many cases, our ultimate goal in the application of CLT is **to identify the characteristics of a population**. A population is the group of individuals we are studying, but the elements that make up a population are not necessarily people.[2]

For instance, if a manager of a grocery chain is trying to increase the efficiency of replenishing seltzer in each store each week so that he or she can sell as much seltzer as possible and avoid large unsold inventory. In this case, all the seltzers sold in that store represent the population.

The Law of Large Numbers

- Very closely related to the CLT, the Law of Large Numbers states that if the same experiment or study is repeated independently a large number of times, the average of the results of the trials must be close to the **expected value** (a long-run average value of random variables). [5]
- The result becomes closer to the expected value as the number of trials is increased.[5]
- The simplest example of the law of large numbers is rolling the dice. The expected value of the dice events is:

$$\mathrm{EV} = \frac{1+2+3+4+5+6}{6} = 3.5$$

- If we roll the dice only three times, and say we roll a 6,6, and 3, the average obtained is 5, which is quite far from our expected value. According to the law of the large numbers, if we roll the dice a large number of times, the average result will be closer to the expected value of 3.5.

Further Research

- Can CLT be applied in all cases?
- What conditions are needed for CLT to work?
- What other applications are there for CLT?
- Does the CLT apply to dependent random variables?
- What are the differences between CLT and the law of large numbers?
- What is it about the number 30 that makes it so special in the CLT?

Works Cited

[1] Bento, C. (2020, August 12). *Understanding central limit theorem with an example | AnalytixLabs*. Blogs & Updates on Data Science, Business Analytics, AI Machine Learning. https://www.analytixlabs.co.in/blog/central-limit-theorem-with-example/

[2] Bento, C. (2020, October 15). *Central limit theorem: A real-life application*. towards data science. https://towardsdatascience.com/central-limit-theorem-a-real-life-application-f638657686e1

[3] *Central limit theorem*. (2016, July 24). https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/BS704_Probability12.html

[4] Corporate Finance Institute. (2020, May 15). *Central limit theorem*. https://corporatefinanceinstitute.com/resources/knowledge/other/central-limit-theorem/

[5] Corporate Finance Institute. (2020, May 26). *Law of large numbers*. https://corporatefinanceinstitute.com/resources/knowledge/other/law-of-large-numbers/

[6] Frost, J. (2020, August 12). *Central limit theorem explained*. Statistics By Jim. https://statisticsbyjim.com/basics/central-limit-theorem/