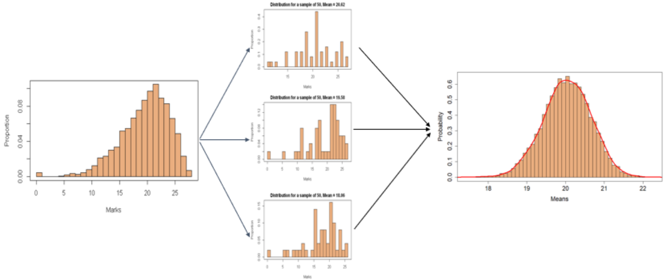
Central limit theorem



Outline

Learning Objectives

By the end of this video, we hope that you will be able to:

- Understand what is Central Limit Theorem (CLT) and how it works
- Examine some practical uses of CLT

Central limit theorem

Central Limit Theorem

The Central Limit Theorem (CLT) is a fundamental concept in probability theory. It states the following:

If you have a population

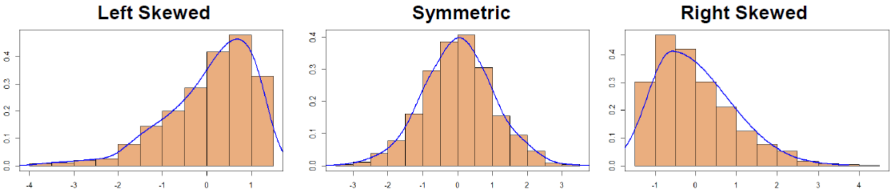
$$X_1, X_2, X_3, \cdots, X_n$$

Mean: $\mu_{\bar{X}} = \mu$

Standard deviation: $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

Central Limit Theorem

This original distribution could be discrete or continuous, left-skewed or right-skewed, and even multi-modal!



Central Limit Theorem

There are 2 important conditions to note:

- The samples are independent.
- The number of observations is large enough.

If the conditions are met, we can use the normal probability model to make inferences about the distribution mean, based on the arithmetic mean of observed values.

Computing CLT-based Confidence Intervals

Computing CLT-based Confidence Intervals

The CLT is what we use to compute Normal-based Confidence intervals.

The template formula

$$\bar{x} \pm (\text{margin of error})$$

When applying the CLT in this context, the margin of error will be computed as

$$\text{Margin of error} = z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

Computing CLT-based Confidence Intervals in R

Suppose we have a vector of observations X and wish to compute a 90% confidence interval for it. The necessary code is:

```
n <- length(X)
lower_limit <- mean(X) + qnorm(0.05)*sd(X)/sqrt(n)
upper_limit <- mean(X) - qnorm(0.05)*sd(X)/sqrt(n)
```

CLT Examples

Here are some real life examples of the application of the CLT

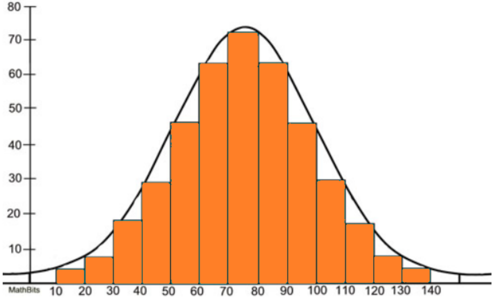
- In finance, an economist may collect a random sample of 100 individuals in country X and use the average annual income of the sampled individuals to estimate the average annual income of individuals in country X.
- In the sciences, a botanist may measure the height of 50 randomly selected trees and use the sample mean height to estimate the population mean height.
- In education, a teacher can use CLT to estimate the results of the student population.
- In medicine, a doctor can use CLT to estimate the effect of certain medication.

Checking Normality

Checking Normality

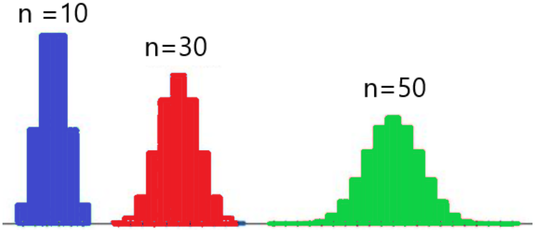
In probability theory, the central limit theorem (CLT) states that the distribution of a sample variable approximates a normal distribution (i.e., a "bell curve") as the sample size becomes larger, assuming that all samples are identical in size, and regardless of the population's actual distribution shape.

- We construct histograms of the original data and check for Normality.
- If the original distribution is close to Normal, we do not need such a large sample size in order for the CLT to be valid.



What is the best sample size?

- The shape of the sampling distribution depends on the sample size.
- If you perform the study using the same procedure and only change the sample size, the shape of the sampling distribution will differ for each sample size.

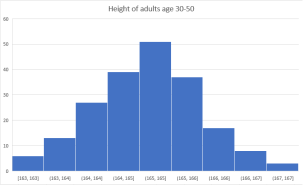


What is the best sample size to use?

Real Life Examples

For example, we want to know the height of adults aged 30-50.

- Suppose we take 200 simple random samples (with replacement) of size n=50 from the population and compute the mean for each of the samples.
- The Poisson distribution is another probability model that is useful for modelling discrete variables.



For example, a service helpdesk staff receives about 7 phone calls a day. What is the probability of receiving 10 phone calls?

- Similarly, if we take 200 simple random samples (with replacement) from the population and compute the mean for each of the samples.
- The sample means distribution will be normal.

References

Rossetti, M. (2015). *Simulation Modelling and Arena.* John Wiley Sons.