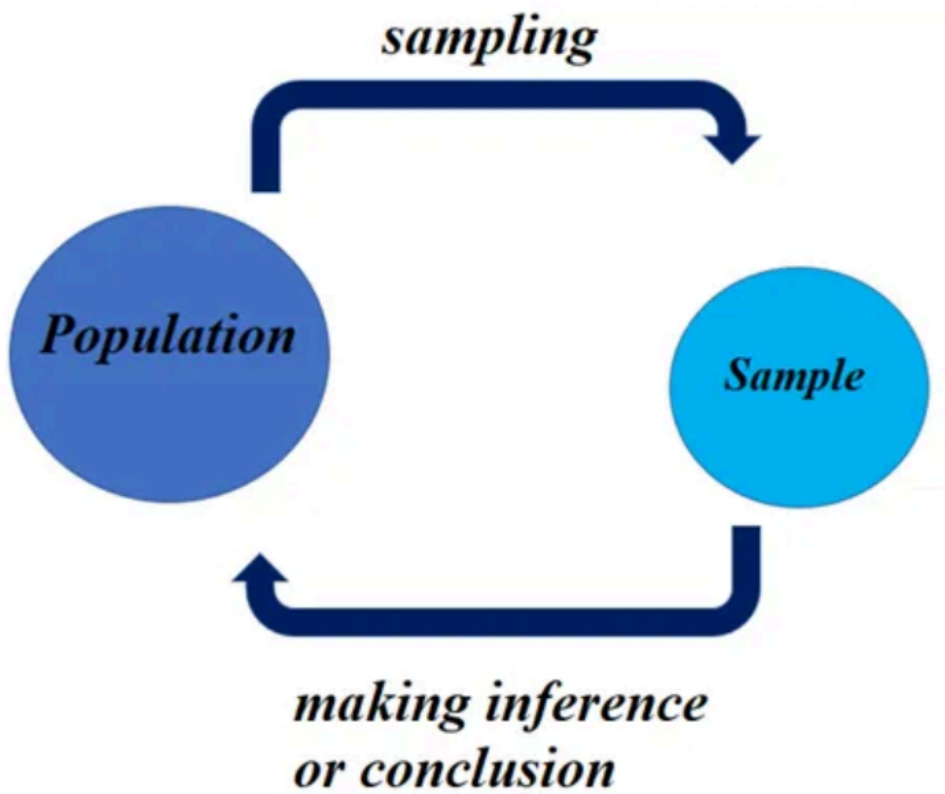
Statistical Inference

Adejumo Ridwan Suleiman

What is Statistical Inference?

Definition

This is the process of making conclusions about a population based on information derived from the sample of that population



Properties of a good sample

- When the sample is homogeneous
- When the sample is representative
- When the sample is choosen at random

Definition of Terms

Parameter:

This is a value that describes the whole population

Statistic:

This is a value that describes the sample.

Categories of Statistical Inference

- **Estimation:**
  - Interest in estimating the value of a parameter.
  - For example response rate of a cancer investigative drug.
- **Hypothesis testing**:
  - Compare the parameters for two samples using statistical tests of significance.
  - For example investigating whether a particular racial group have higher blood sugar level on average when compared to another racial group.

Point Estimate

Estimation

- One of the uses of biostatistics in health is to investigate certain characteristics of a population.
- Characteristics of a population are either:
  - **Quantitative** (average blood sugar level)
  - **Qualitative** (proportion of adult men who are hypertensive)

Types of Estimate

- Point estimate
- Interval estimate

Point Estimate

- Single numerical value used to estimate the corresponding population parameter
- Examples:
  - **sample mean**() is a point estimate for the **population mean**(). $\bar{x}$ μ
  - **sample proportion** is a point estimate for the **population proportion**.
  - **sample deviation**() is a point estimate for the **population standard deviation**(). s σ

Sampling Distribution

- Point estimate might not always equal to population parameter.
- Different samples generated from a population, will give different point estimate.
- This distribution of point estimates generated from a given population is called sampling distribution.

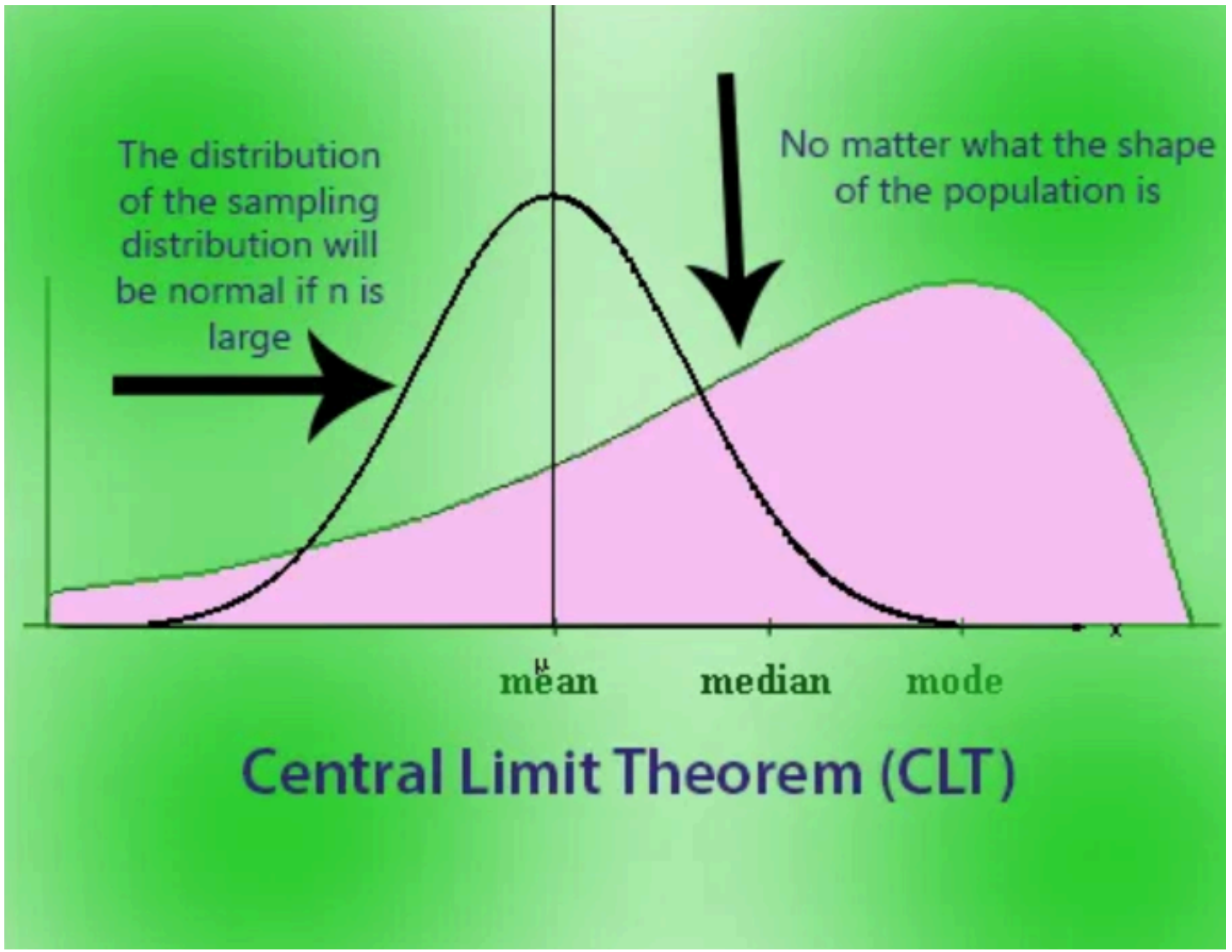
Central Limit Theorem

Regardless the distribution of a variable in the population, the sampling distribution of a statistic is always approximately normally distributed.

- Sampling distribution is normally distributed with parameters **population mean()** and **standard error(**.
- **Mean of sampling distribution** = **True population mean** μ
- Standard deviation of sampling distribution is referred to as standard error.
- The standard deviation of these sample means will be approximately equal to the standard error calculated from a single sample.

$\left(\frac{\sigma}{\sqrt{n}}\right)$

Sampling distribution



Source. Medium

Standard Error

- Measures the reliability and precision of an estimate

$$SE = \frac{\sigma}{\sqrt{n}}$$

Standard error for mean

$$SE(\bar{X}) = \frac{s}{\sqrt{n}}$$

Standard Error of the Proportion

Standard deviation vs standard error

Standard deviation	Standard Error
Used with individual measurements	Used with deviation of a statistic
Describes variation or dispersion in the dataset	Describes the precision of the sample mean estimate
Does not change significantly with sample size	Decreases as sample size increases

Interval Estimate

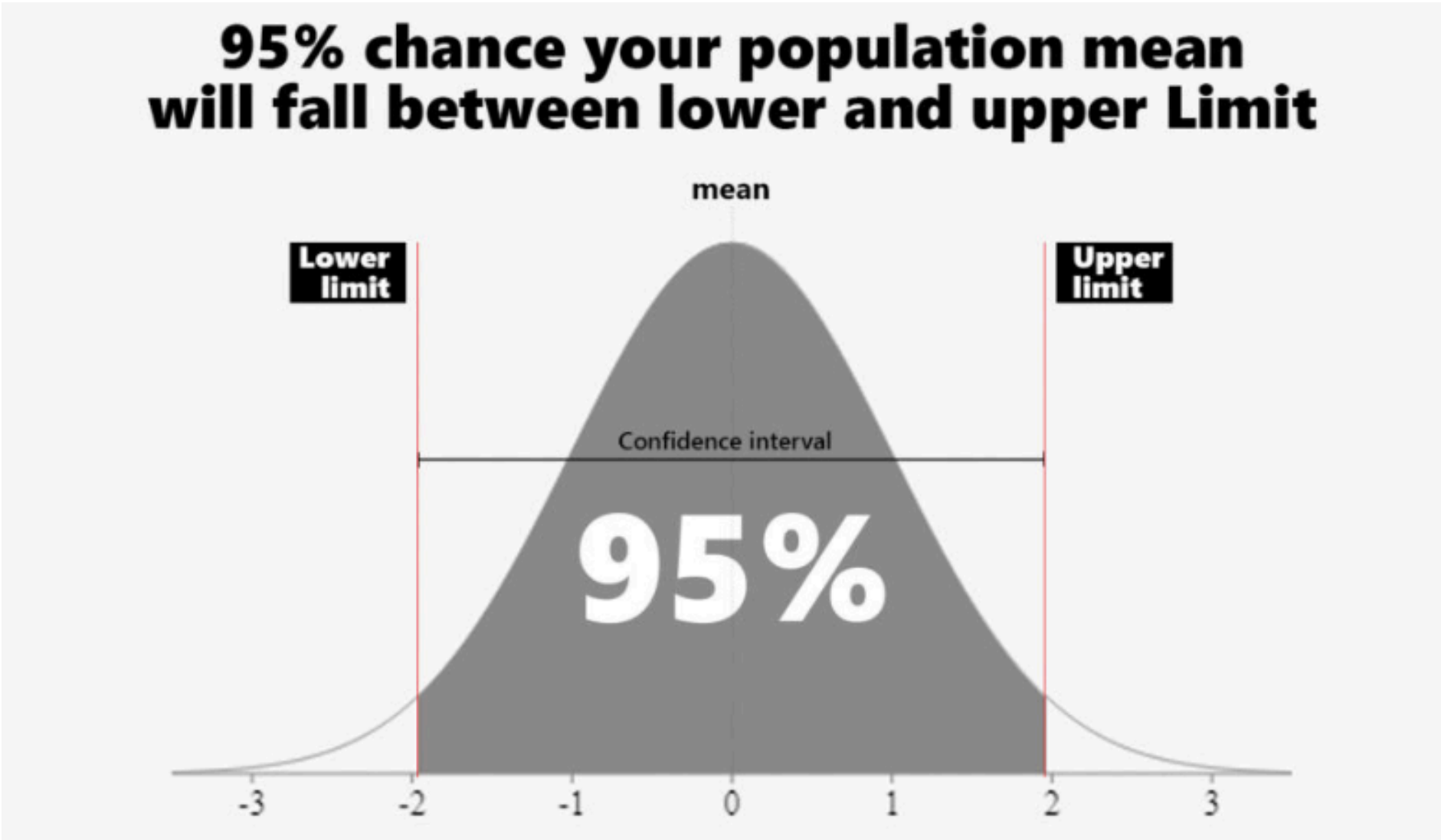
Interval Estimate

- Combination of the point estimate and its standard error
- The two numerical values are referred to as confidence interval

Confidence Interval (C.I)

-
- $\text{statistic} \pm \text{reliability coefficient} \cdot S.E(statistic)$
- $\text{reliability coefficient} \cdot \text{S.E(statistic)} = \text{precision of estim}$

Confidence Interval (C.I)



Source. MathBlog

Reliability coefficient

- Depends on:
  - Desired degree of confidence **(95% confidence level)**
  - Normality of the data distribution
  - Knowledge of population variance
  - Sample size

Interpreting Confidence Intervals

Comparing mean values or proportions

- Confidence intervals containing 0 implies data is consistent with Null Hypothesis.

For Relative Risks or Odds Ratio

- Confidence intervals containing 1 implies risk factor not statistically significant.

Random sampling error

Advantages of CI

- Gives range of plausible values for the population.
- Conveys information on the magnitude of differences, let's say systolic blood pressure values between male and female patients.
- Explains the level of exposure a subject is affected by.