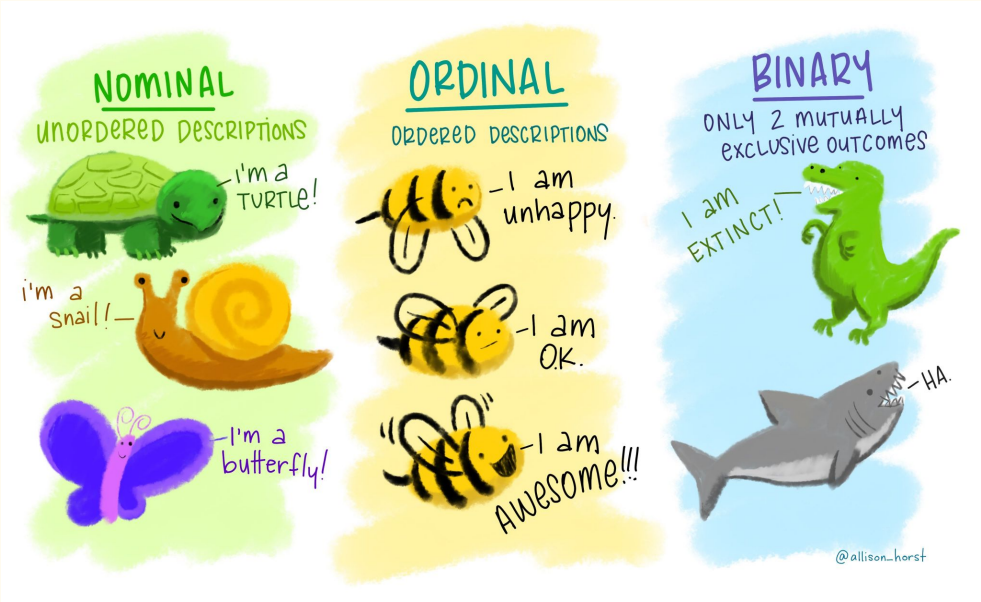
MIDTERM Review

STA130F24

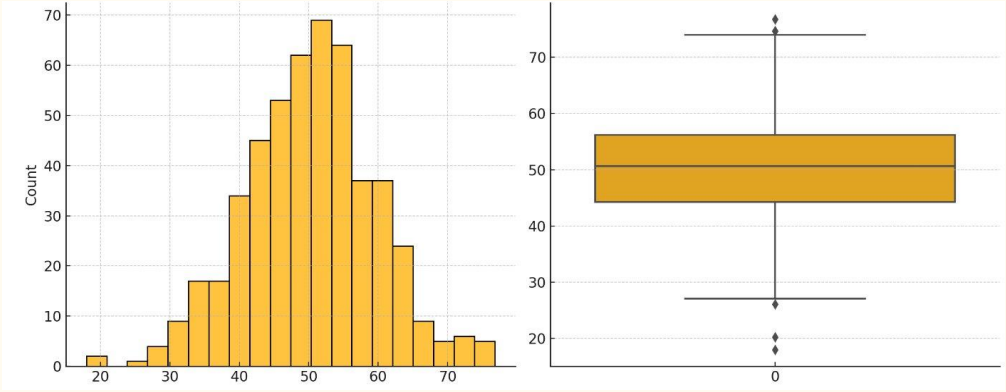
Data Types

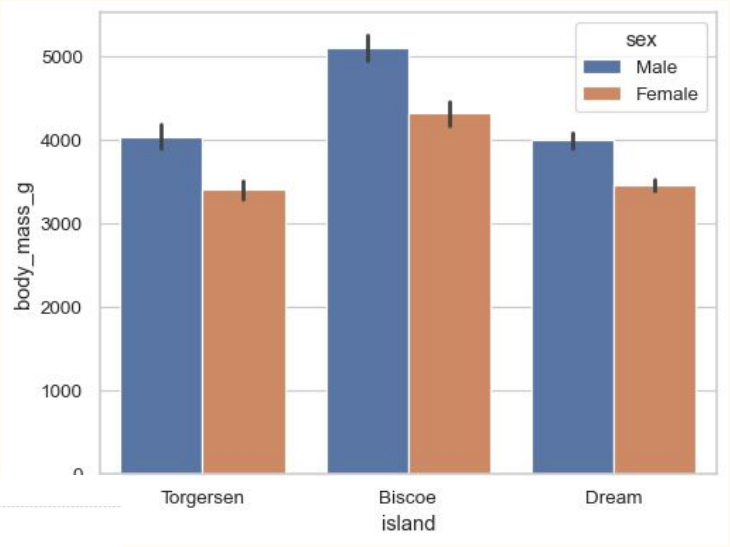
Quantitative variables

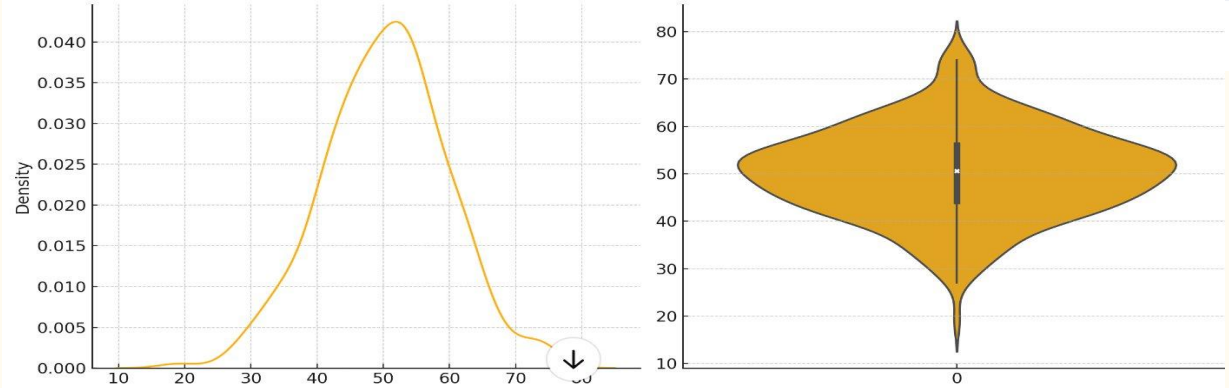
Qualitative variables



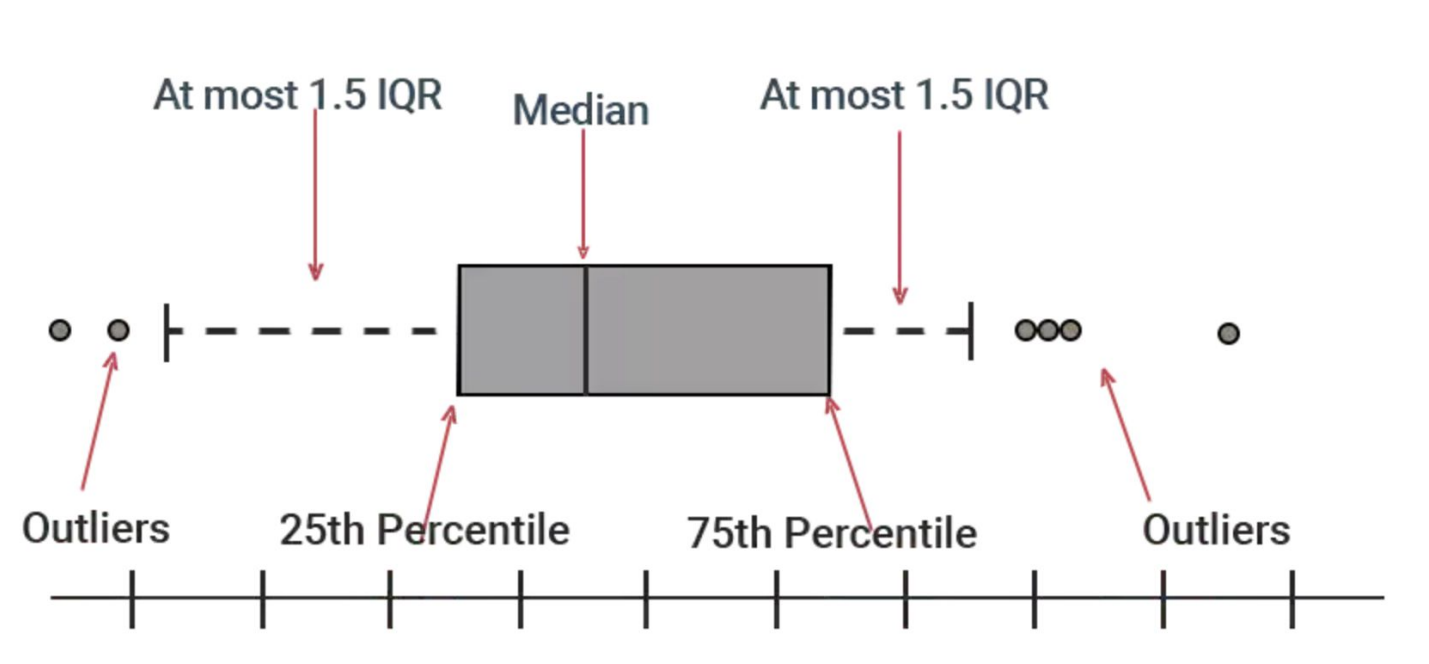
How to Visualize your Data?



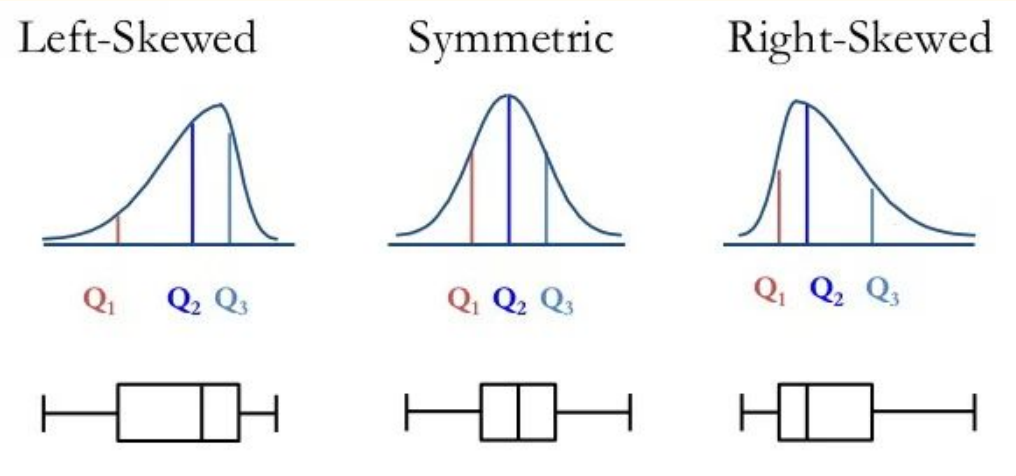


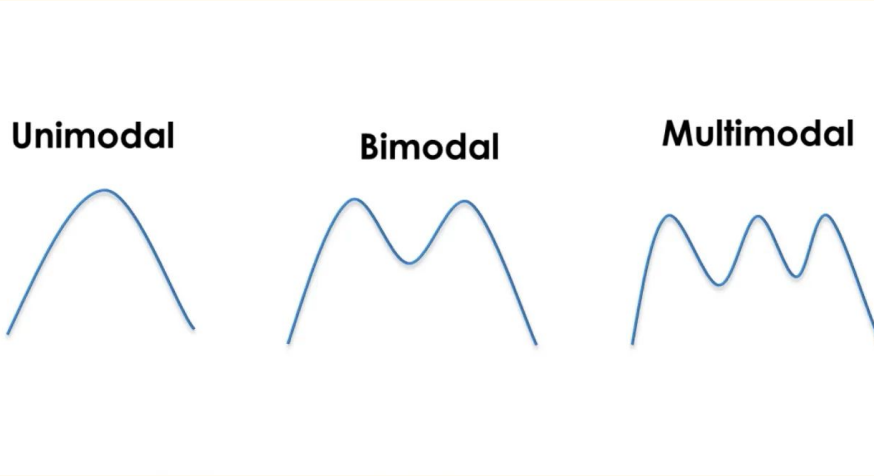


Median and Potential Outliers



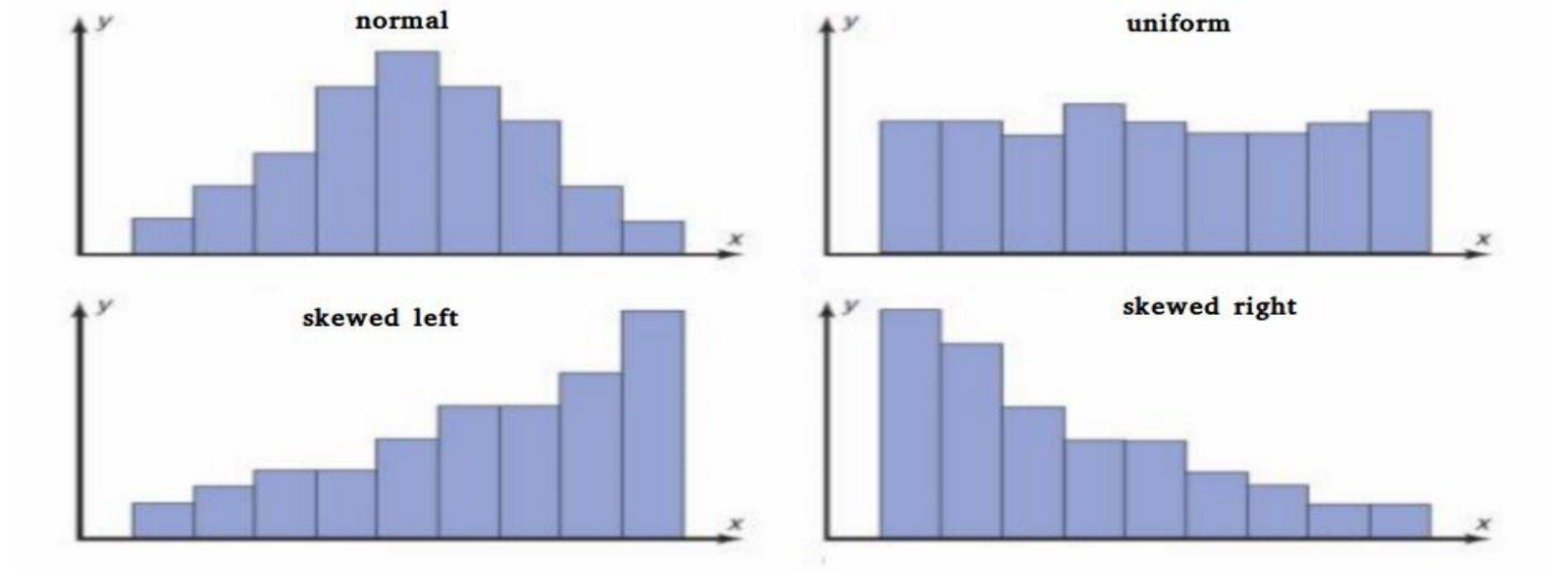
Skewness and Modality



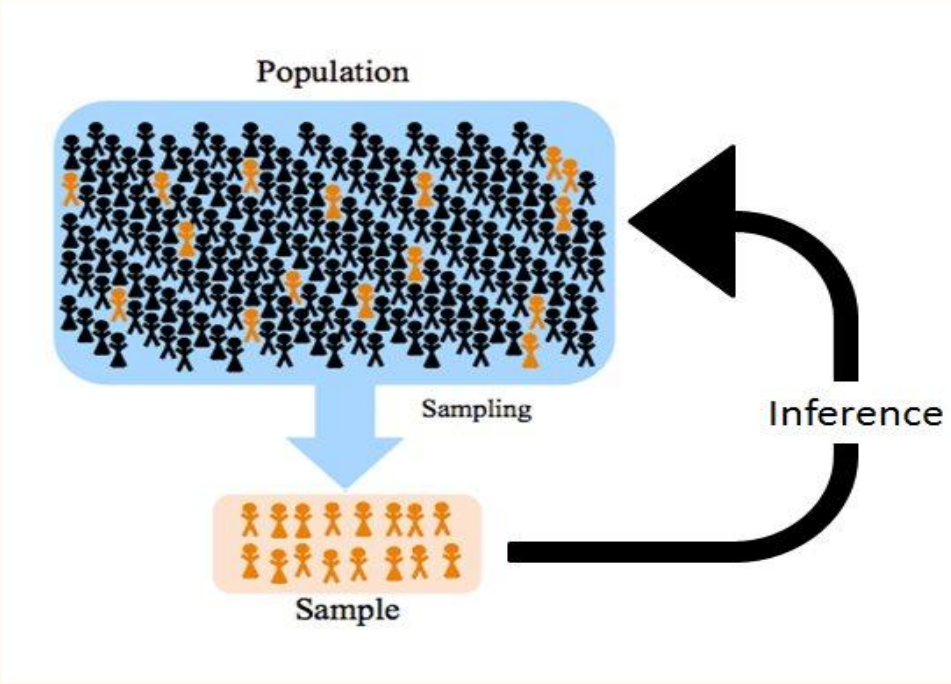


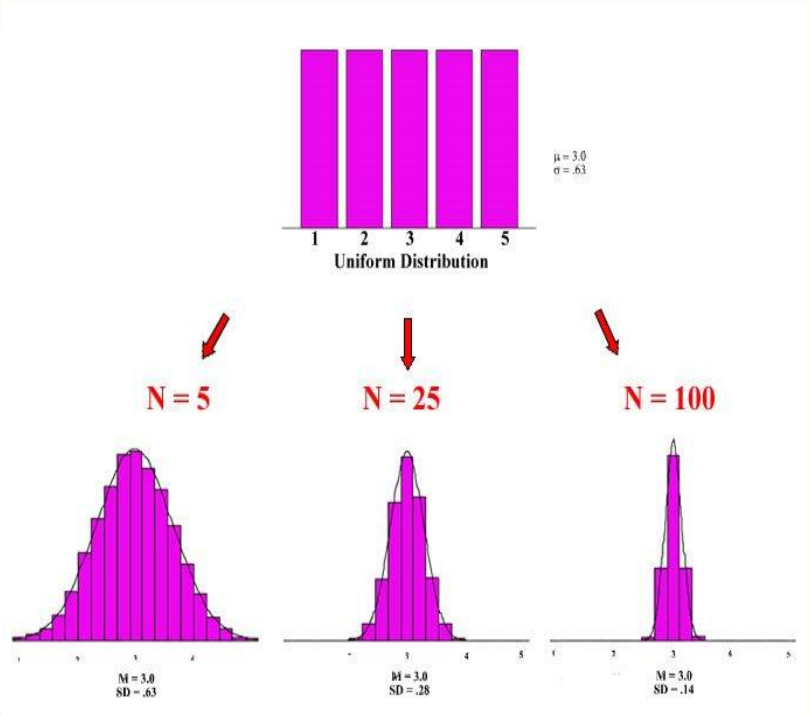
What is Distribution?

The way in which the values of a random variable are spread or allocated.



Sampling Distribution





Standard Deviation vs Standard Error

Standard deviation: the amount of variation or dispersion of individual data points within a single sample. It shows how spread out the data points are around the sample mean.

Standard error: how much variability there is in the sample mean as an estimate of the population mean

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

- x_i = each individual data point.
- $\bar{x}$ = sample mean.
- n = sample size.

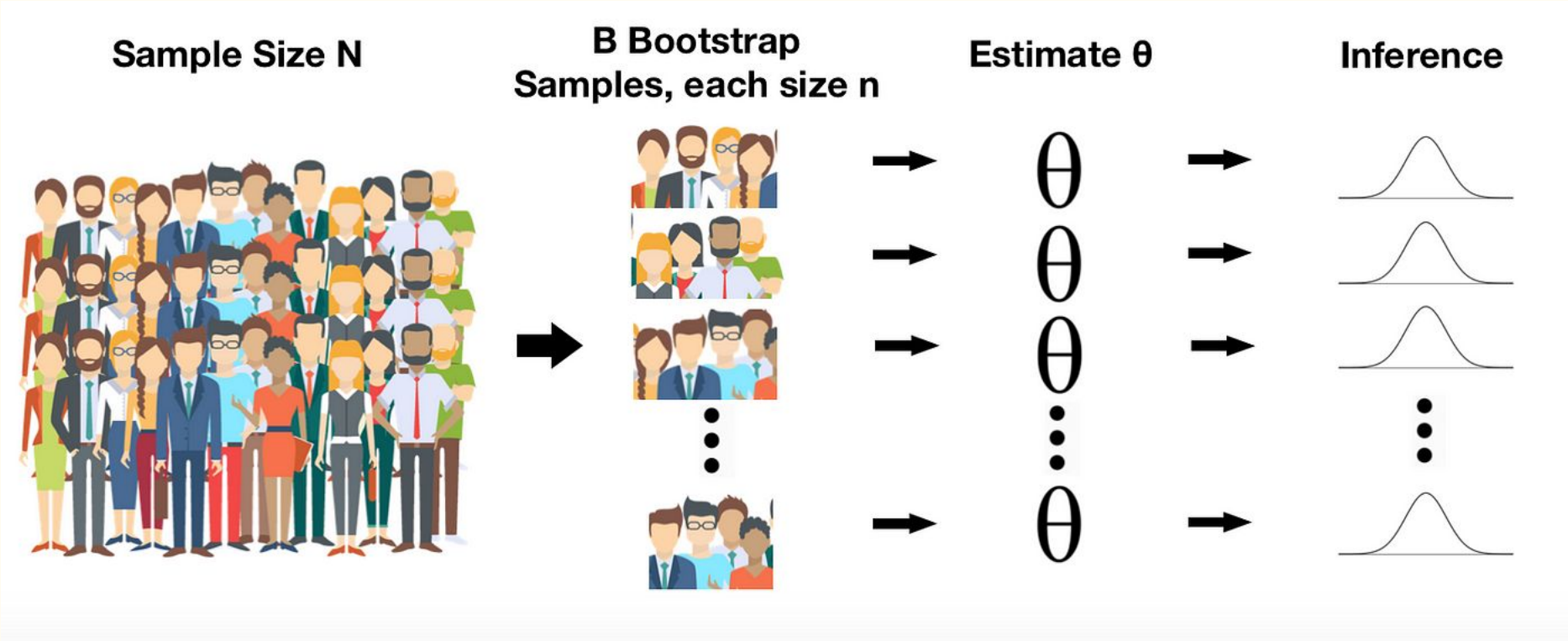
$$SE = \frac{s}{\sqrt{n}}$$

How number of samples (n) drives SE?

The standard error of the mean *decreases as the sample size increases*, indicating more "precise" estimates with larger samples.

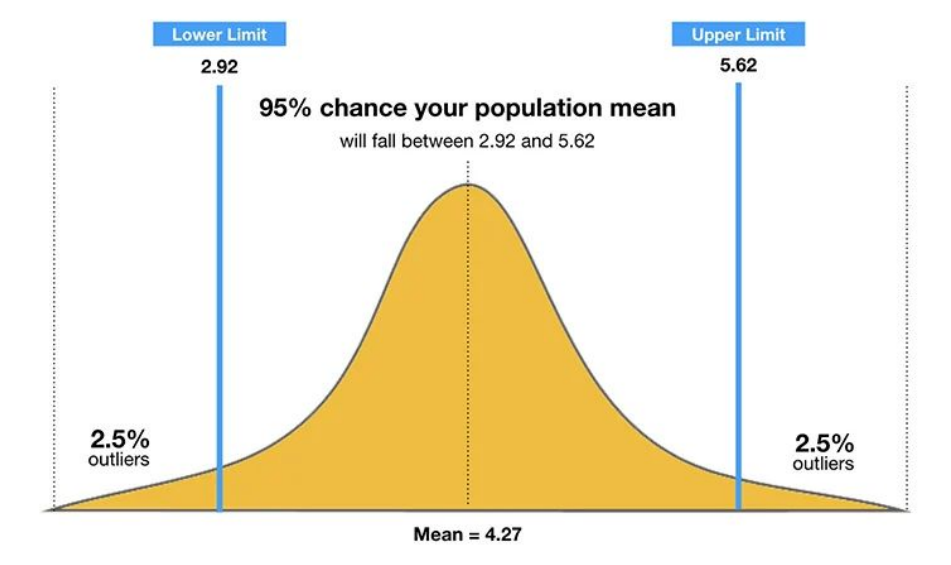
$$SE = \frac{s}{\sqrt{n}}$$

Bootstrapping



Confidence Interval

- A plausible range of values for a population parameter.
- The most common choice for the confidence level is **95%**, meaning that the chance that an independent sample will result in the construction of a confidence interval which "captures" the actual true value of the population parameter is 95%.



Hypothesis Testing

- Hypothesis test give a decision about whether a hypothesis is consistent with the observed data.

  – **Null hypothesis** (H_0)
    - The working theory until proven otherwise
  – **Alternative hypothesis** (H_a)
    - The opposing theory of the null
  – **Test statistic**
    - Evidence from the data to help make the decision about H_0
  – **P-value**
    - Probability the test statistic supports H_0 assuming its true
  – **Significance level** (α)
    - Threshold to determine if we reject H_0 based on the p-value

Null Hypothesis vs Alternative Hypothesis

- **Null hypothesis** (H_0)
  - The working theory until proven otherwise
- **Alternative hypothesis** (H_a)
  - The opposing theory of the null

The **null hypothesis** usually simply states the "no effect" (on average) assumption

H_0 : The vaccine has no effect **(on average)** on patient health
H_1 : H_0 is false

To empasize that "**(on average)**" refers to the pupulation parameter μ (the average effect), it is helpful to more formally (and concisely) express this equivalently as

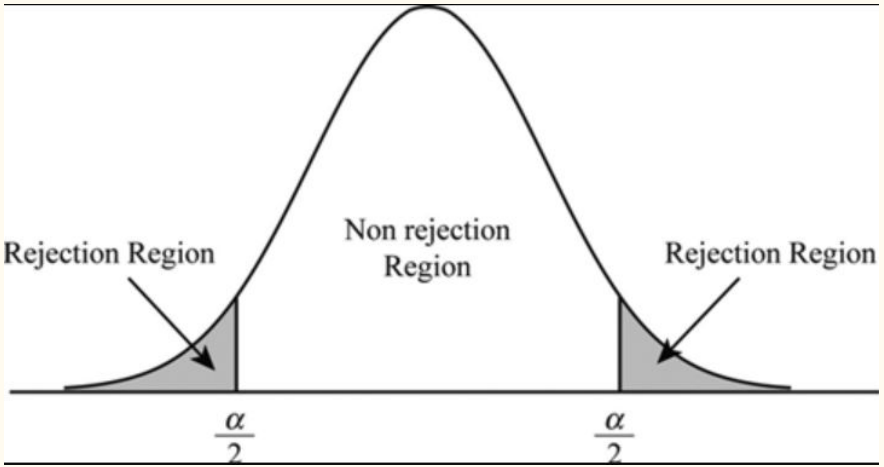
$$H_0 : \mu = 0 \quad \text{and} \quad H_A : H_0 \text{ is false}$$

Test Statistics (Observed vs Simulated)

- **Test statistic**: mean, median, mode, variance, etc.
- **Observed test statistic**: calculated based on sample data.
- **Simulated test statistic**: calculated based on the simulation under the assumption that null hypothesis is true.

P-value & Significance level

- **P-value**: The probability that a test statistic (simulated test statistic) is as or more extreme than the observed test statistic under the **assumption that the null hypothesis is true**.
- **Significance level / Rejection level**: threshold to determine if we reject the null hypothesis based on the p-value.



P-value and the strength of the evidence

p-value	Evidence
$p > 0.1$	No evidence against the null hypothesis
$0.1 \geq p > 0.05$	Weak evidence against the null hypothesis
$0.05 \geq p > 0.01$	Moderate evidence against the null hypothesis
$0.01 \geq p > 0.001$	Strong evidence against the null hypothesis
$0.001 \geq p$	Very strong evidence against the null hypothesis

Type I and Type II Errors

Decision	Null Hypothesis is True	Null Hypothesis is False
Reject Null	**Type I Error (with chance α)**	Correct Decision
Fail to Reject	Correct Decision	**Type II Error (with chance β)**

- **Type I.** If a medical test incorrectly shows that a patient has a disease when they actually don't.
- **Type II.** If a medical test incorrectly shows that a patient does not have a disease when they actually do.

One-sided vs Two-sided Hypothesis Test

One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$ Rejection = (1- a)	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$ Rejection = (1- a) / 2	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$