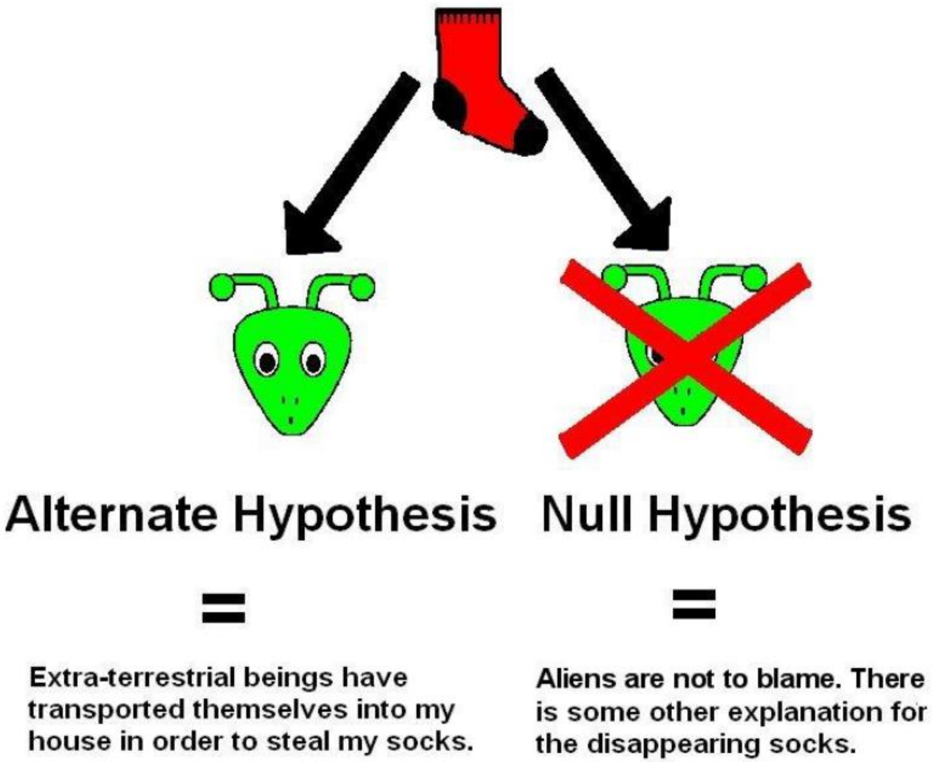
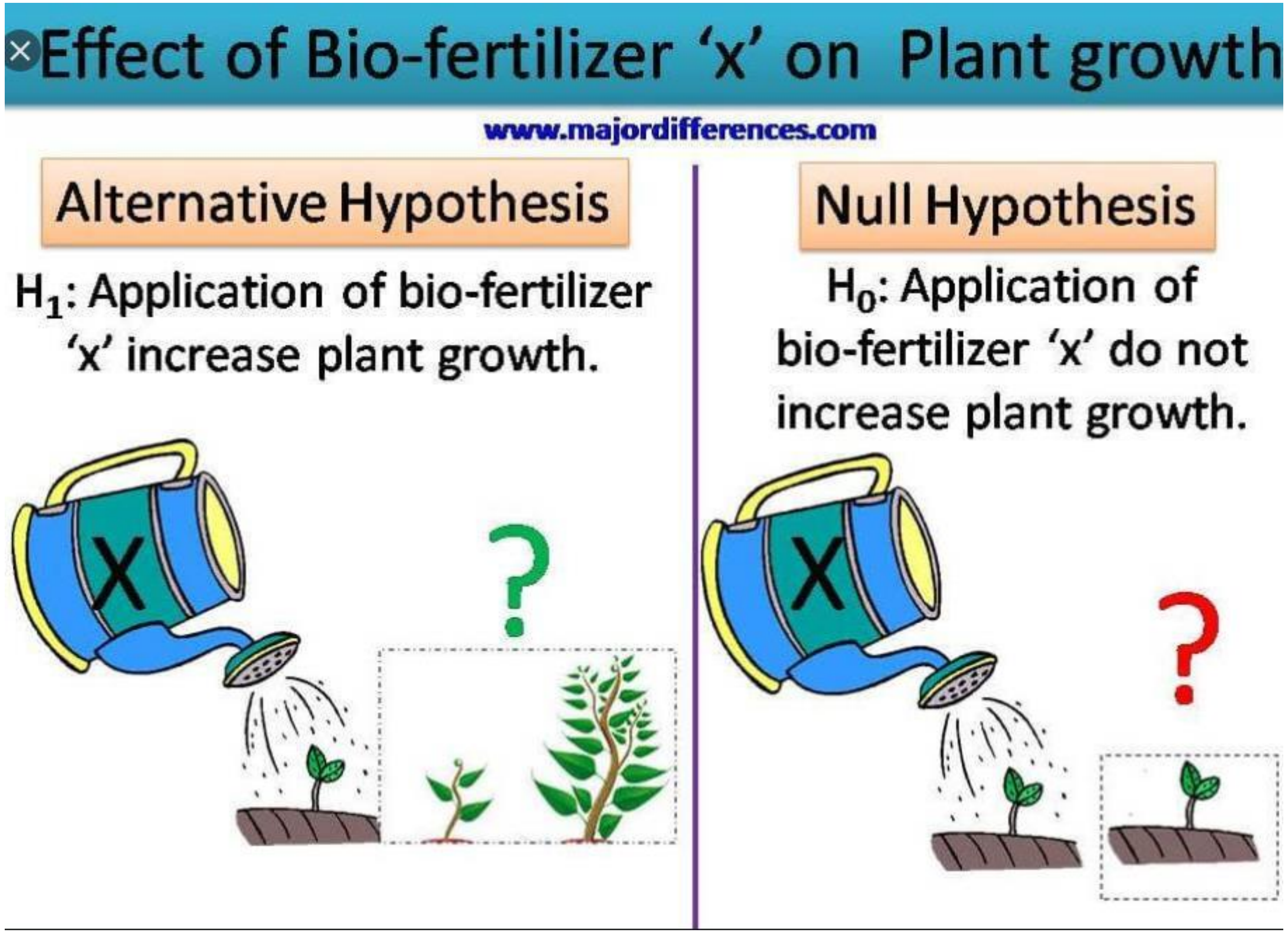
Hypothesis Testing

- Often we are called on to *make decisions* or *draw conclusions* on the new design or improvement in performance of a given process *based on sampled data*.
- In making a decision, we typically *form a hypothesis* concerning what we believe is true and then *collect data to prove or disapprove* the hypothesis.
  - “Eating chocolate everyday can lower the risk of heart disease”
  - “Indian population has better immunity against COVID-19”
  - “Music can enhance plant growth” …
- In statistical hypothesis testing, we generally formulate two hypotheses.
  - **The null hypothesis (H_0):** will be rejected or nullified if the sample data do not support it.
  - **Alterative hypothesis (H_1):** any hypothesis that is different from H0 is called an alternative hypothesis, denoted by H1. ***Any time, H_0 is rejected, H_1 will be considered accepted.***

Q. Where have all my socks gone?



https://www.jcu.edu.au/__data/assets/pdf_file/0018/115344/Basic-Statistics-8_Hypothesis-Testing.pdf



Hypothesis Testing of Population Means

- There are three ways to set up the alternative hypothesis (H_1)

$$\text{Method 1}: \; H_0: \mu_x \leq \mu_0 \qquad H_1: \; \mu > \mu_0$$

$$\text{Method 2}; \; H_0: \mu_x \geq \mu_0 \qquad H_1: \; \mu < \mu_0$$

$$\text{Method 3}: \; H_0: \mu_x = \mu_0 \qquad H_1: \; \mu \neq \mu_0$$

A statistical hypothesis test consists of the following six steps:

1. State the null and alternative hypotheses. Define the test statistic used to analyze the situation.
2. Determine the significance level, 'α', at which the test will be made.
3. Collect the data and calculate the test statistic result.
4. Define the reference distribution for the test statistic.
5. Compare the test statistic and its reference distribution under H0. Carry out the necessary analysis of data.
6. Assess the risk.

Errors and Significance Level

Two kinds of errors may be committed when testing hypotheses. If the null hypothesis is rejected when it is true, a type I error has occurred. If the null hypothesis is *not* rejected when it is false, a type II error has been made. The probabilities of these two errors are given special symbols

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ is true})$$
$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 | H_0 \text{ is false})$$

Sometimes it is more convenient to work with the **power** of the test, where

$$\text{Power} = 1 - \beta = P(\text{reject } H_0 | H_0 \text{ is false})$$

The general procedure in hypothesis testing is to specify a value of the probability of type I error α, often called the **significance level** of the test, and then design the test procedure so that the probability of type II error β has a suitably small value.

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β

Scribbr

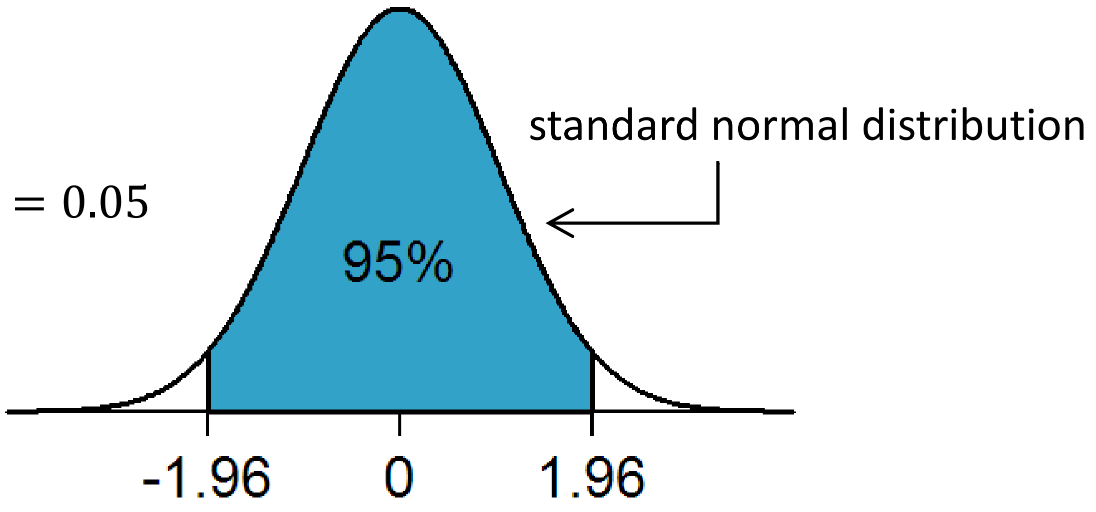
The power of a binary hypothesis test is the probability that the test correctly rejects the null hypothesis when a specific alternative hypothesis is true. It represents the chances of a true positive detection conditional on the actual existence of an effect to detect.

Significance Level

- In statistical hypothesis testing, a result has *statistical significance* when it is very unlikely to have occurred given the null hypothesis.
- More precisely, a study's defined significance level, α, is the *probability of the study rejecting the null hypothesis*, given that the null hypothesis was assumed to be true
- **p-value** of a result, is the probability of obtaining a result at least as extreme, *given that the null hypothesis is true*

The result is statistically significant when $p \leq \alpha$

Example, when $\alpha = 5\% = 0.05$



TRY THIS:

https://www.mathsisfun.com/data/standard-normal-distribution-table.html

Chapter 2.1

Classical Design of Experiments

Comparative Experiments

Comparative Experiments

Three cases of Hypothesis Testing

1. $H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0$

    Example

    H_0: Sachin Tendulkar's ODI batting average is 50 runs

    H_1: Sachin Tendulkar's ODI batting average is NOT 50 runs

2. $H_0: \mu \geq \mu_0, \quad H_1: \mu < \mu_0$

    Example

    H_0: Sachin Tendulkar's ODI batting average is at least 50 runs

    H_1: Sachin Tendulkar's ODI batting average is less than 50 runs

3. $H_0: \mu \leq \mu_0, \quad H_1: \mu > \mu_0$

    Example

    H_0: Sachin Tendulkar's ODI batting average is at max 50 runs

    H_1: Sachin Tendulkar's ODI batting average is greater than 50 runs

How do you draw statistical conclusions?

Comparative Experiments

- **Comparing two processes/products/datasets**
- **Comparing a process/product/dataset with a reference**

Case 1: When we have ALL the population data

Examples:

1. Who is **a better ODI batsman**, based on runs scored in an inning, **Sachin or MS Dhoni**?
2. Is Sachin's ODI average greater than 'X'?

Comparative Experiments

Case 2: What happens when we have PARTIAL population data?

Example

To test the newspaper claim that the mean wage rate of local foundry workers is \$16 an hour, 25 foundry workers were randomly surveyed. It was found that the average wage rate for the sample of workers was \$14.50. Historical data suggest that the wage rates follow the normal distribution and the standard deviation of wage rates is \$3. Can the Union claim that the average wage is not \$16 an hour? Assume α =0.05.

newspaper claim : $\mu = 16$

sample survey $n = 25$, $\bar{y} = 14.50$

we know $\sigma = 3$ (somehow)

Hypothesis testing

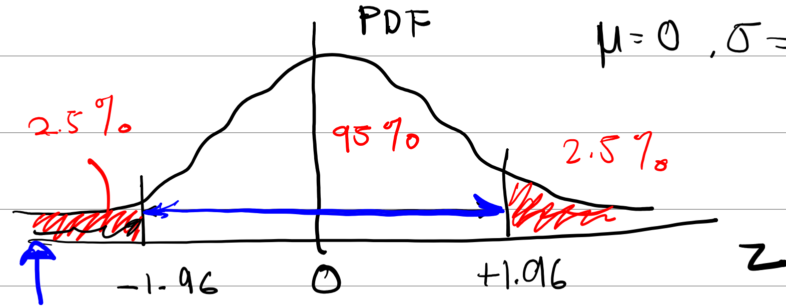
$H_0 : \mu = 16$ ✗

$H_1 : \mu \neq 16$ ✓

$$z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{14.5 - 16}{3/\sqrt{25}}$$

$$z = \frac{-1.5}{0.6} = -2.5$$

Ho



std normal dist.

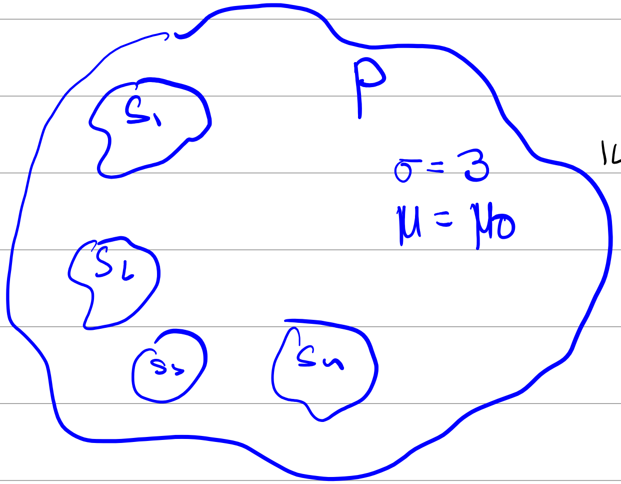
NOTE: You do NOT have permission to share this file or any of its contents with anyone else, and/or upload it on internet or any of the platforms where it can be accessed by others.

Comparative Experiments

$\mu = 16$ (?) $\quad \sigma = 3$

14.5, 14.8, 16.8, 17.5 ...

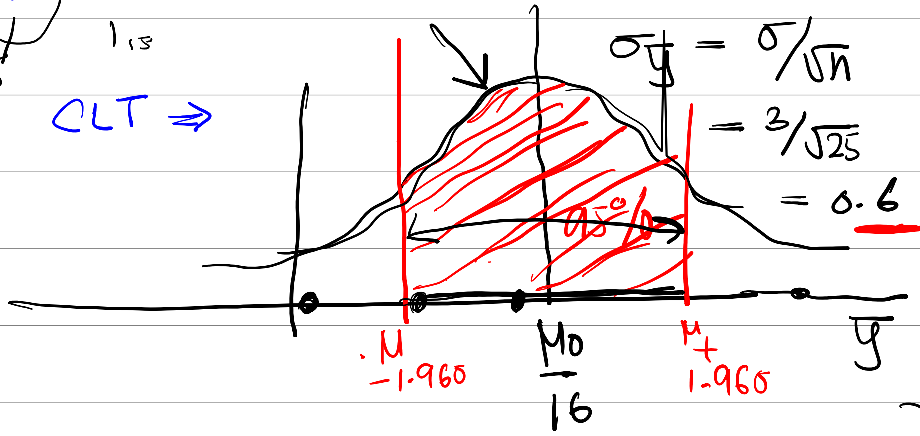
Example



$\bar{y}_1, \bar{y}_2, \dots \bar{y}_n$

14.5

CLT ⇒



$\sigma_{\bar{y}} = \sigma/\sqrt{n}$

$= 3/\sqrt{25}$

$= 0.6$

95% $\quad \alpha = 5\%$

99% $\quad \alpha = 1\%$

99.73% $\quad \alpha = 0.27\%$

$\mu \pm 1.96\sigma$

$16 \pm 1.96 \times 0.6$

[14.824, 17.176]

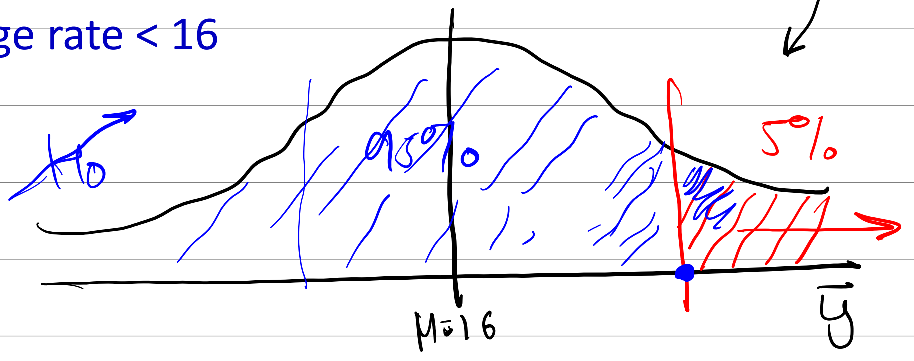
14.5

How do you test for wage rate > 16 or wage rate < 16

$H_0 : \mu \leq 16$ $\quad \bar{y} = 14.5$

$H_1 : \mu > 16$

[−∞, 1.64]



$16 \pm 1.96\sigma$

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}}$$

Case 3: What happens when we have NO population data?

Example 1

An engineer is studying **the formulation of a Portland cement mortar**. He has **added a polymer latex emulsion** during mixing to determine if this impacts the curing time and tension bond strength of the mortar.

The experimenter prepared 10 samples of the original formulation and 10 samples of the modified formulation.

TABLE 2.1

Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Question: Does adding polymer latex emulsion change the strength?

Comparative Experiments

Case 3: What happens when we have NO population data?

Example 2

Who is a better ODI batsman, Virat or Babar? (Based on the runs scored in an inning)

Batsman	One sample each of 10 ODI innings	Sample Mean	Sample Std. Dev
Virat	00, 53, 34, 31, 00, 54, 96, 20, 10, 19	**31.7**	29.6
Babar	12, 09, 91, 79, 51, 45, 41, 46, 29, 33	**43.6**	26.0

What is the hypothesis test?

What is the statistical (mathematical) model based on the hypothesis?