# Recitation, Week 9

Ye Wang

New York University

POL-850

Spring 2018

1. Hypothesis
2. Null and alternative hypotheses
3. Test Statistic
4. Hypothesis Testing

# Hypothesis



## Definition

A testable statement about the empirical relationship between an independent variable and a dependent variable

## A Template

In a comparison of **[units of analysis]** those having **[one value on the independent variable]** will be more likely to have **[one value on the dependent variable]** than will those having **[a different value on the independent variable]**

# Hypothesis



## A Clinical Trial Example

In a comparison of **patients** those **receiving a new treatment** will be more likely to have **longer life expectancy** than will those **receiving no treatment**

# Hypothesis



## A Clinical Trial Example

In a comparison of **patients** those **receiving a new treatment** will be more likely to have **longer life expectancy** than will those **receiving no treatment**

Suppose we run a clinical trial and collect data.
How do we know that the new treatment is more effective?

# Hypothesis



## A Clinical Trial Example

In a comparison of **patients** those **receiving a new treatment** will be more likely to have **longer life expectancy** than will those **receiving no treatment**

Suppose we run a clinical trial and collect data.
How do we know that the new treatment is more effective?

We compare those two groups!

# Hypothesis

**Methods of comparison** depends on **the level of measurement of variable**

<table>
<tr><td></td><td></td><td colspan="3">Y</td></tr>
<tr><td></td><td></td><td>Nominal</td><td>Ordinal</td><td>Interval</td></tr>
<tr><td rowspan="3">X</td><td>Nominal</td><td colspan="2" rowspan="2">Cross-tab</td><td rowspan="2">Comparison of means</td></tr>
<tr><td>Ordinal</td></tr>
<tr><td>Interval</td><td></td><td></td><td></td></tr>
</table>

# Hypothesis



This trial is going to be an ideal experiment:

- Random sampling of the patients

# Hypothesis



This trial is going to be an ideal experiment:

- Random sampling of the patients
- Random assignment of the new treatment

# Hypothesis



This trial is going to be an ideal experiment:

- Random sampling of the patients
- Random assignment of the new treatment
- Measure the life expectancy

# Hypothesis

This trial is going to be an ideal experiment:

- Random sampling of the patients
- Random assignment of the new treatment
- Measure the life expectancy
- We collect data and compare them

# Hypothesis



This trial is going to be an ideal experiment:

- Random sampling of the patients
- Random assignment of the new treatment
- Measure the life expectancy
- We collect data and compare them

But wait.. what do we want to know **ultimately**?

# Hypothesis

But wait.. what do we want to know **ultimately**?

# Hypothesis



But wait.. what do we want to know **ultimately**?

- We want to know whether our hypothesis is correct

# Hypothesis



## But wait.. what do we want to know **ultimately**?

- We want to know whether our hypothesis is correct
- More precisely, the new treatment is effective at the population level

# Hypothesis



## But wait.. what do we want to know **ultimately**?

- We want to know whether our hypothesis is correct
- More precisely, the new treatment is effective at the population level
- Again this problem is about making inference about the population parameter $\mu$ with a sample statistic (e.g. $\overline{X}$)
- Note that $\mu$ represents population life expectancy
- Because basically what we want to know: $\mu_T > \mu_C$

# Hypothesis



Suppose we run the trial many times with a different sample each time.

| (1) | Treatment | Control |
|---|---|---|
| Mean | 5 years | 2 years |
| N | 100 | 100 |

# Hypothesis



Suppose we run the trial many times with a different sample each time.

| (1) | Treatment | Control |
|---|---|---|
| Mean | 5 years | 2 years |
| N | 100 | 100 |

| (2) | Treatment | Control |
|---|---|---|
| Mean | 2.5 years | 2 years |
| N | 100 | 100 |

# Hypothesis

| (3) | Treatment | Control |
|---|---|---|
| Mean | 2.1 years | 2 years |
| N | 100 | 100 |

# Hypothesis

| (3) | Treatment | Control |
|---|---|---|
| Mean | 2.1 years | 2 years |
| N | 100 | 100 |

| (4) | Treatment | Control |
|---|---|---|
| Mean | 5 years | 2 years |
| N | 10 | 10 |

# Group exercise 1



Q1. When would you be more confident that the new treatment is effective?

1. Result (1) vs. Result (2)
2. Result (1) vs. Result (3)
3. Result (1) vs. Result (4)
4. Result (2) vs. Result (4)

Q2. Can you quantify your confidence?

# Hypothesis



- They all show a positive change when receiving the new treatment

# Hypothesis

- They all show a positive change when receiving the new treatment
- Sometimes we feel intuitively that the change is not “real”

# Hypothesis



- They all show a positive change when receiving the new treatment
- Sometimes we feel intuitively that the change is not “real”
- This is because things can happen just by chance!

# Hypothesis



- They all show a positive change when receiving the new treatment
- Sometimes we feel intuitively that the change is not “real”
- This is because things can happen just by chance!
- In other words, we happen to have a specific sample but that sample can be very different from the population, why?

# Hypothesis



- They all show a positive change when receiving the new treatment
- Sometimes we feel intuitively that the change is not "real"
- This is because things can happen just by chance!
- In other words, we happen to have a specific sample but that sample can be very different from the population, why?
- (Because there is random sampling error)

# Hypothesis



- They all show a positive change when receiving the new treatment
- Sometimes we feel intuitively that the change is not “real”
- This is because things can happen just by chance!
- In other words, we happen to have a specific sample but that sample can be very different from the population, why?
- (Because there is random sampling error)
- As a result, you can’t tell if the hypothesis is correct or not in a systematic way

We need a statistical theory of **hypothesis testing**!

# Null Hypothesis

## Definition

$H_0$: There is **no** relationship between dependent and independent variables in **population** based on our data at hand (sample)

| | Treatment | Control |
|---|---|---|
| Mean | 2.1 years | 2 years |
| N | 100 | 100 |

$H_0$: receiving the new treatment does not prolong life expectancy (of population)

$$H_0 \colon \mu_T = \mu_C$$

# Alternative Hypothesis



## Definition

$H_A$: There **is** a relationship between dependent and independent variables in **population** based on our data at hand (sample)

For example, suppose we have the result like (3)

| | Treatment | Control |
|---|---|---|
| Mean | 2.1 years | 2 years |
| N | 100 | 100 |

$H_A$: receiving the new treatment prolongs or shortens life expectancy (of population)

$$H_A\colon \mu_T \neq \mu_C$$

# Language

We do hypothesis testing (statistical tests) and then

1. If our evidence is consistent with the alternative hypothesis $H_A$:

   we reject $H_0$

2. Otherwise (not consistent with the alternative hypothesis $H_A$):

   we failed to reject $H_0$

# Language



Suppose we have the result (1)

| | Treatment | Control |
|---|---|---|
| Mean | 5 years | 2 years |
| N | 100 | 100 |

Then after a statistical test, we find evidence supportive of our hypothesis

Then we say we **reject** our null hypothesis $H_0$: $\mu_T = \mu_C$ that the new treatment has no effect

# Language

Suppose we have the result (3)

| | Treatment | Control |
|---|---|---|
| Mean | 2.1 years | 2 years |
| N | 100 | 100 |

Then after a statistical test, we find evidence not supportive of our hypothesis

Then we say we **failed to reject** our null hypothesis $H_0$: $\mu_T = \mu_C$ that the new treatment has no effect

# Group exercise 2

But why can't we just say our alternative hypothesis is correct (or true)?

# Group exercise 2



## But why can't we just say our alternative hypothesis is correct (or true)?

This is because our hypothesis testing always involves **uncertainty**!
We'll never know whether our hypothesis about population is true

Also be careful not to say "accept the alternative hypothesis" (because there's always some level of uncertainty)

# Test Statistic



## Definition of (Sample) Statistic

A single measure of some attribute of a sample. It is calculated by applying a function to the values of the items of the sample.

For example, we've learned some sample statistics

- $\overline{X}$: the sample mean
- $s$: the sample standard deviation

We use a **test statistic** to test a hypothesis!

# Test Statistic



## Definition of **Test Statistic**

A statistic (a quantity derived from the sample) used in statistical hypothesis testing. This number compares our data at hand with what is expected under the null hypothesis.

# Test Statistic



## Definition of **Test Statistic**

A statistic (a quantity derived from the sample) used in statistical hypothesis testing. This number compares our data at hand with what is expected under the null hypothesis.

It is noteworthy that

- A test statistic has its sampling distribution
- Why?

# Test Statistic



## Definition of **Test Statistic**

A statistic (a quantity derived from the sample) used in statistical hypothesis testing. This number compares our data at hand with what is expected under the null hypothesis.

It is noteworthy that

- A test statistic has its sampling distribution
- Why? Test statistics are *random variables* (random numbers)!
- There are many test statistics we can use
- t-statistic, chi-square statistic, z-statistic, and so on...

# Test Statistic

## Definition of **Test Statistic**

A statistic (a quantity derived from the sample) used in statistical hypothesis testing. This number compares our data at hand with what is expected under the null hypothesis.

It is noteworthy that

- A test statistic has its sampling distribution
- Why? Test statistics are *random variables* (random numbers)!
- There are many test statistics we can use
- t-statistic, chi-square statistic, z-statistic, and so on...
- The choice of a test statistic depends on the level of measurement, statistical assumptions we make, and the problem at our hand.

# Test Statistic

For example, **two-sample tests** are appropriate for comparing two samples, typically treatment and control samples from a scientifically controlled experiment.

We use a z-statistic as below:

$$z = \frac{\overline{x}_T - \overline{x}_C}{\sqrt{\frac{\sigma_T^2}{n_T} + \frac{\sigma_C^2}{n_C}}}$$

Assume that we know **population standard deviation**

cf. If $\sigma$ is unknown, then we use t-statistic and use $s$ (sample standard deviation)

# Test Statistic



$$z = \frac{\overline{x}_T - \overline{x}_C}{\sqrt{\frac{\sigma_T^2}{n_T} + \frac{\sigma_C^2}{n_C}}}$$

- $\overline{X}_T$: the sample mean of the treatment group
- $\overline{X}_C$: the sample mean of the control group
- $\sigma_T$: the population standard deviation of the treatment group
- $\sigma_C$: the population standard deviation of the control group
- $n_T$: the sample size of the treatment group
- $n_C$: the sample size of the control group

# Hypothesis Testing

Suppose we have the result (3)

| | Treatment | Control |
|---|---|---|
| Mean | 2.1 years | 2 years |
| $\sigma$ | 0.4 years | 0.6 years |
| N | 100 | 100 |

# Hypothesis Testing



Suppose we have the result (3)

| | Treatment | Control |
|---|---|---|
| Mean | 2.1 years | 2 years |
| $\sigma$ | 0.4 years | 0.6 years |
| N | 100 | 100 |

Then we can calculate the test statistic:

$$z = \frac{\overline{x}_T - \overline{x}_C}{\sqrt{\frac{\sigma_T^2}{n_T} + \frac{\sigma_C^2}{n_C}}} = \frac{2.1 - 2}{\sqrt{\frac{0.4^2}{100} + \frac{0.6^2}{100}}} = 1.39$$

# Hypothesis Testing



Suppose we have the result (3)

| | Treatment | Control |
|---|---|---|
| Mean | 2.1 years | 2 years |
| $\sigma$ | 0.4 years | 0.6 years |
| N | 100 | 100 |

Then we can calculate the test statistic:

$$z = \frac{\overline{x}_T - \overline{x}_C}{\sqrt{\frac{\sigma_T^2}{n_T} + \frac{\sigma_C^2}{n_C}}} = \frac{2.1 - 2}{\sqrt{\frac{0.4^2}{100} + \frac{0.6^2}{100}}} = 1.39$$

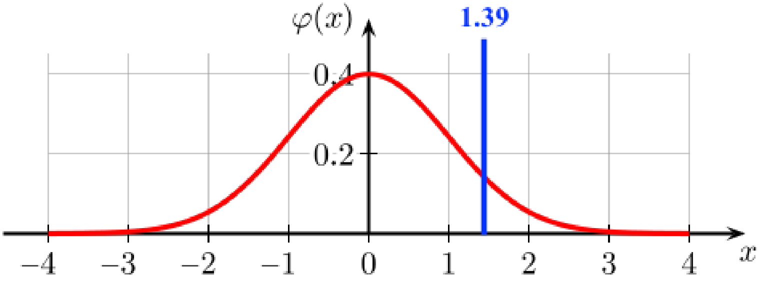
**But what does this statistic (number) mean?**

# Hypothesis Testing

- We know the test statistic ($z = z_{\text{observed}} = 1.39$) and know that this test statistic (Z) has the standard normal distribution.

# Hypothesis Testing

- We know the test statistic ($z = z_{\text{observed}} = 1.39$) and know that this test statistic (Z) has the standard normal distribution.




- It means that we're more likely to see some test statistics than others given the null hypothesis (how about 1.39? later..)
- Large test statistics (either positive or negative) are rarely seen

# Hypothesis Testing

| | Treatment | Control |
|---|---|---|
| Mean | 5 years | 2 years |
| $\sigma$ | 0.4 years | 0.6 years |
| N | 100 | 100 |

- What happens to $z$ if we have large difference $\overline{X}_T - \overline{X}_C$?

# Hypothesis Testing



| | Treatment | Control |
|---|---|---|
| Mean | 5 years | 2 years |
| $\sigma$ | 0.4 years | 0.6 years |
| N | 100 | 100 |

- What happens to $z$ if we have large difference $\overline{X}_T - \overline{X}_C$?

$$z = \frac{\overline{x}_T - \overline{x}_C}{\sqrt{\frac{\sigma_T^2}{n_T} + \frac{\sigma_C^2}{n_C}}} = \frac{5 - 2}{\sqrt{\frac{0.4^2}{100} + \frac{0.6^2}{100}}} = 41.6$$

- In this case, it is very very unlikely to see this test statistic given the null hypothesis

# Hypothesis Testing



| | Treatment | Control |
|---|---|---|
| Mean | 5 years | 2 years |
| $\sigma$ | 0.4 years | 0.6 years |
| N | 100 | 100 |

- What happens to $z$ if we have large difference $\overline{X}_T - \overline{X}_C$?

$$z = \frac{\overline{x}_T - \overline{x}_C}{\sqrt{\frac{\sigma_T^2}{n_T} + \frac{\sigma_C^2}{n_C}}} = \frac{5-2}{\sqrt{\frac{0.4^2}{100} + \frac{0.6^2}{100}}} = 41.6$$

- In this case, it is very very unlikely to see this test statistic given the null hypothesis
- $Pr[Z < -41.6 \text{ or } Z > 41.6] \approx 0$ (p-value is close to zero)

# Hypothesis Testing

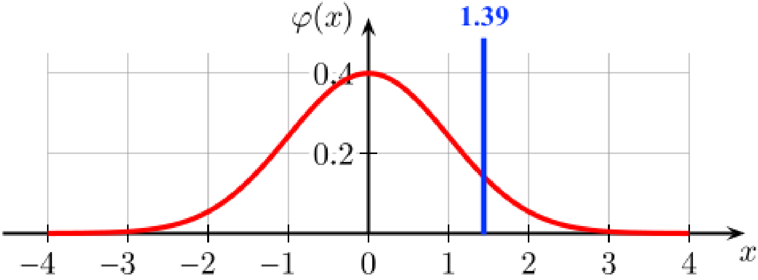
| | Treatment | Control |
|---|---|---|
| Mean | 5 years | 2 years |
| $\sigma$ | 0.4 years | 0.6 years |
| N | 100 | 100 |

- What happens to $z$ if we have large difference $\overline{X}_T - \overline{X}_C$?

$$z = \frac{\overline{x}_T - \overline{x}_C}{\sqrt{\frac{\sigma_T^2}{n_T} + \frac{\sigma_C^2}{n_C}}} = \frac{5-2}{\sqrt{\frac{0.4^2}{100} + \frac{0.6^2}{100}}} = 41.6$$

- In this case, it is very very unlikely to see this test statistic given the null hypothesis
- $Pr[Z < -41.6 \text{ or } Z > 41.6] \approx 0$ (p-value is close to zero)
- Then we can convincingly reject our null hypothesis

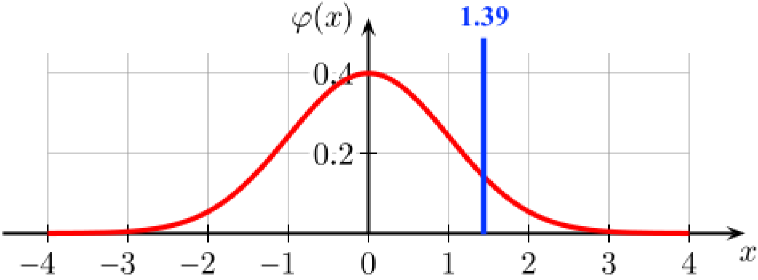
# Hypothesis Testing




- But how about the previous case? $z = 1.39$
- Should we reject our hypothesis or not?

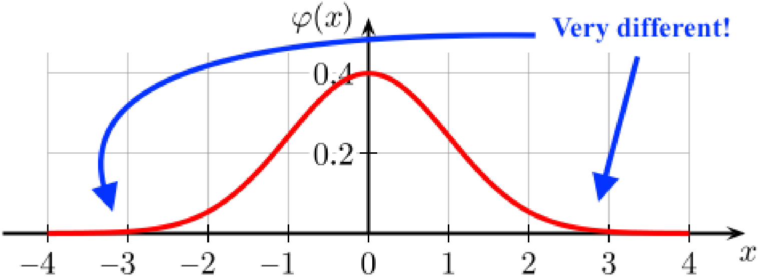
# Hypothesis Testing






- But how about the previous case? $z = 1.39$
- Should we reject our hypothesis or not?
- This is why we need **a criteria (a rejection rule)** to reject our null hypothesis: **a significance level**

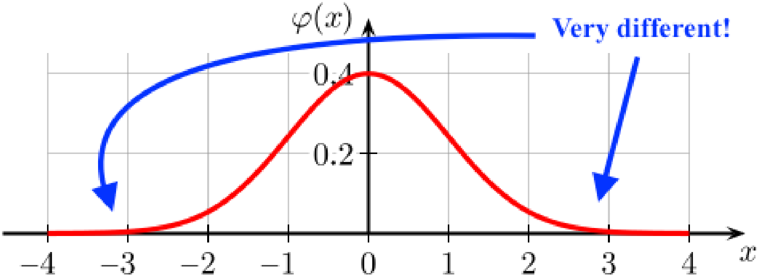
# Hypothesis Testing




- We reject the null hypothesis when two means are **significantly different**

# Hypothesis Testing






- We reject the null hypothesis when two means are **significantly different**
- For those cases, the test statistic would be a large number (either positive or negative)

# Significance level

- But **how “significant**” do we want?

## Significance level

A **threshold** probability beyond which we **reject** the null hypothesis
More precisely we want to pin down $p$ such that (two-tailed test)

$$Pr\big[|Z| > c\big] = p$$

# Significance level



- But **how "significant**'' do we want?

## Significance level

A **threshold** probability beyond which we **reject** the null hypothesis
More precisely we want to pin down $p$ such that (two-tailed test)

$$Pr\left[|Z| > c\right] = p$$

## Rejection rule

We reject $H_0$ in favor of $H_A$ when $|z_{\text{obs}}| > c$ or $|p_{\text{obs}}| < p$

# Individual exercise 1



Suppose we want the 0.05 significance level for our hypothesis testing. Then what are $c$ and $p$?

## Significance level

A **threshold** probability beyond which we **reject** the null hypothesis

More precisely we want to pin down $p$ such that (two-tailed test)

$$Pr\big[|Z| > c\big] = p$$

## Rejection rule

We reject $H_0$ in favor of $H_A$ when $|z_{\text{obs}}| > c$ or $|p_{\text{obs}}| < p$

# Significance level

- Can choose different significance levels (0.001, 0.05, 0.01...)

# Significance level

- Can choose different significance levels (0.001, 0.05, 0.01...)

Example 1

- Suppose our p-value is 0.04 (very unlikely to see this given $H_0$)
- And we choose the 0.05 significance level,

# Significance level

- Can choose different significance levels (0.001, 0.05, 0.01...)

Example 1

- Suppose our p-value is 0.04 (very unlikely to see this given $H_0$)
- And we choose the 0.05 significance level, then we reject $H_0$ since 0.04¡0.05

Example 2

- Suppose our z-value is 1.74
- Again the 0.05 significance level

# Significance level

- Can choose different significance levels (0.001, 0.05, 0.01...)

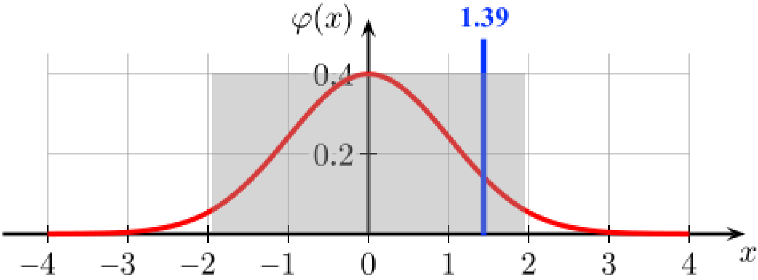
<u>Example 1</u>

- Suppose our p-value is 0.04 (very unlikely to see this given $H_0$)
- And we choose the 0.05 significance level, then we reject $H_0$ since 0.04¡0.05

<u>Example 2</u>

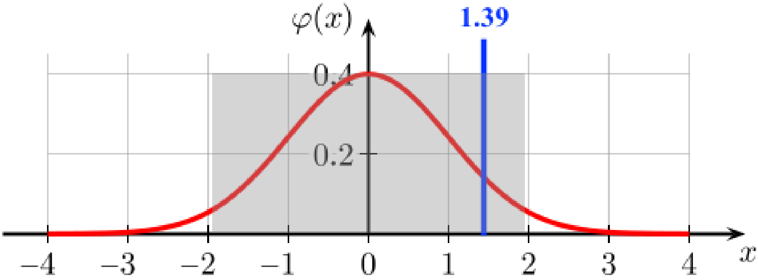
- Suppose our z-value is 1.74
- Again the 0.05 significance level
- We fail to reject $H_0$ since 1.74¡1.96 (NB: $c = 1.96$)

# Hypothesis Testing




- Let's choose the 0.05 level (p-value is 0.05)

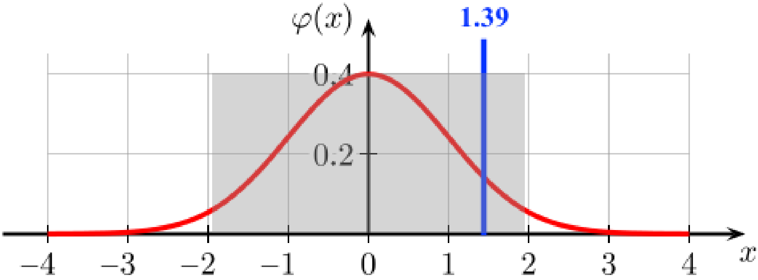
# Hypothesis Testing




- Let's choose the 0.05 level (p-value is 0.05)
- $Pr[-z_{0.025} < Z < z_{0.025}] = .95$ and p-value is 0.05
- And $z_{0.025} = 1.96$

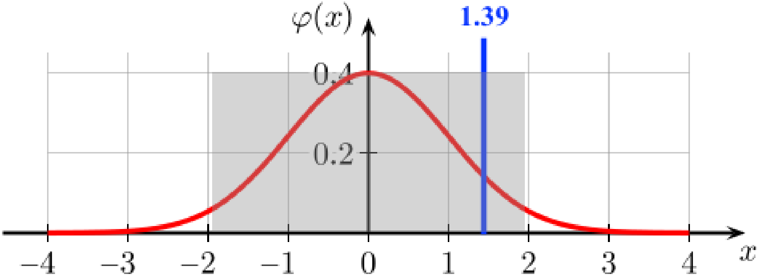
# Hypothesis Testing






- Let's choose the 0.05 level (p-value is 0.05)
- $Pr[-z_{0.025} < Z < z_{0.025}] = .95$ and p-value is 0.05
- And $z_{0.025} = 1.96$
- This means that we reject our null hypothesis when $z > 1.96$

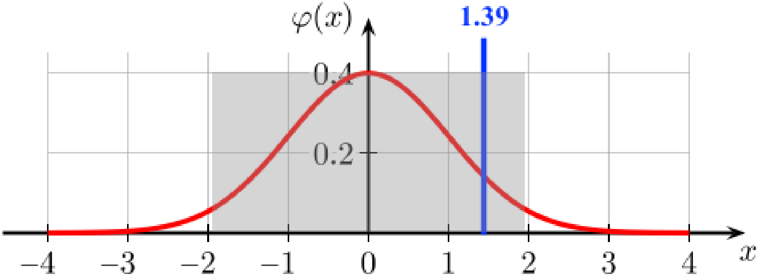
# Hypothesis Testing




- Let's choose the 0.05 level (p-value is 0.05)
- $Pr[-z_{0.025} < Z < z_{0.025}] = .95$ and p-value is 0.05
- And $z_{0.025} = 1.96$
- This means that we reject our null hypothesis when $z > 1.96$
- Here we fail to reject our null hypothesis

# Hypothesis Testing






- Alternatively, we can use our $z$ to obtain the corresponding p-value
- $Pr[-1.39 < z < 1.39] = 0.84$ and p-value is 0.16
- This p-value is not smaller than 0.05
- Therefore, we fail to reject our null hypothesis

# Individual exercise 2



| | Treatment | Control |
|---|---|---|
| Mean | 2.5 years | 2 years |
| $\sigma$ | 1 year | 1 year |
| N | 18 | 18 |

- Would you reject the null hypothesis $H_0$: $\mu_T = \mu_C$ given this data? (suppose you choose the 0.05 significance level)

$$z = \frac{\overline{x}_T - \overline{x}_C}{\sqrt{\frac{\sigma_T^2}{n_T} + \frac{\sigma_C^2}{n_C}}}$$

- Hint: to get the p-value, you should refer to the z-score table.

# To sum up

Suppose we have the result (1)

| | Treatment | Control |
|---|---|---|
| Mean | 5 years | 2 years |
| $\sigma$ | 0.4 years | 0.6 years |
| N | 100 | 100 |

- Let's choose 0.05 as our significance level for the test
- Its p-value is close to zero (smaller than 0.05)
- This means that it's very unlikely to see this given $H_0$
- (And the z-statistic 41.6 is greater than 1.96)
- Therefore we reject the null hypothesis
- In other words, our result is *statistically significant* at the 0.05 significance level.

# To sum up

Suppose we have the result (3)

| | Treatment | Control |
|---|---|---|
| Mean | 2.1 years | 2 years |
| $\sigma$ | 0.4 years | 0.6 years |
| N | 100 | 100 |

- Let's choose 0.05 as our significance level for the test
- Its p-value is 0.16 (greater than 0.05)
- This means that it's not unlikely to see this given $H_0$
- (And the z-statistic 1.39 is smaller than 1.96)
- Therefore we fail to reject the null hypothesis
- In other words, our result is *not statistically significant* at the 0.05 significance level.

# Key takeaways

You should be able to answer these six questions:

1. Hypothesis?

# Key takeaways

You should be able to answer these six questions:

1. Hypothesis? A **testable** statement about the **empirical** relationship between an **independent** variable and a **dependent** variable
2. Null hypothesis?

# Key takeaways



You should be able to answer these six questions:

1. Hypothesis? A **testable** statement about the **empirical** relationship between an **independent** variable and a **dependent** variable
2. Null hypothesis? There is **no** relationship between those two
3. Alternative hypothesis?

# Key takeaways



You should be able to answer these six questions:

1. Hypothesis? A **testable** statement about the **empirical** relationship between an **independent** variable and a **dependent** variable
2. Null hypothesis? There is **no** relationship between those two
3. Alternative hypothesis? There **is** a relationship between those two
4. Hypothesis testing?

# Key takeaways



You should be able to answer these six questions:

1. Hypothesis? A **testable** statement about the **empirical** relationship between an **independent** variable and a **dependent** variable
2. Null hypothesis? There is **no** relationship between those two
3. Alternative hypothesis? There **is** a relationship between those two
4. Hypothesis testing? **Statistical** test on whether to **reject** the null hypothesis
5. Test statistic?

# Key takeaways

You should be able to answer these six questions:

1. Hypothesis? A **testable** statement about the **empirical** relationship between an **independent** variable and a **dependent** variable
2. Null hypothesis? There is **no** relationship between those two
3. Alternative hypothesis? There **is** a relationship between those two
4. Hypothesis testing? **Statistical** test on whether to **reject** the null hypothesis
5. Test statistic? A **statistic** (a number from a sample) used in hypothesis testing
6. Significance level?

# Key takeaways



You should be able to answer these six questions:

1. Hypothesis? A **testable** statement about the **empirical** relationship between an **independent** variable and a **dependent** variable
2. Null hypothesis? There is **no** relationship between those two
3. Alternative hypothesis? There **is** a relationship between those two
4. Hypothesis testing? **Statistical** test on whether to **reject** the null hypothesis
5. Test statistic? A **statistic** (a number from a sample) used in hypothesis testing
6. Significance level? A **threshold** probability beyond which we **reject** the null hypothesis