Hypothesis Testing I

Nate Wells

Math 141, 3/16/22

Outline

In this lecture, we will...

Outline

In this lecture, we will...

- Use P-values to quantify the strength of evidence against the null hypothesis
- Investigate significance level as means of making decisions
- Discuss decision errors and statistical power

Section 1

Hypothesis Testing Framework

Framework for Hypothesis Testing

Hypothesis Testing represents a type of scientific experiment, and so should follow the general scientific method.

1. Present research question
2. Identify hypotheses
3. Obtain data
4. Calculate relevant statistics
5. Compute likelihood of observing statistic under original hypothesis
6. Determine statistical significance and make conclusion on research question

Review: Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective or that there is no relationship among several variables.

Review: Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.

Review: Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.

Review: Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.

Review: Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.
- The alternate hypothesis in a **two-sided hypothesis test** proposes that the population parameter is not equal null value. (i.e. $p \neq .5$)

Review: Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.
- The alternate hypothesis in a **two-sided hypothesis test** proposes that the population parameter is not equal null value. (i.e. $p \neq .5$)
- The alternate hypothesis in a **one-sided hypothesis test** proposes that the population parameter is less than (or greater than) the null value (i.e. one of $p > .5$ or $p < .5$)

Review: Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.
- The alternate hypothesis in a **two-sided hypothesis test** proposes that the population parameter is not equal null value. (i.e. $p \neq .5$)
- The alternate hypothesis in a **one-sided hypothesis test** proposes that the population parameter is less than (or greater than) the null value (i.e. one of $p > .5$ or $p < .5$)

Approximating the Null Distribution

- The distribution of the statistic of interest, *if the null hypothesis were true*, is called the **Null Distribution**

Approximating the Null Distribution

- The distribution of the statistic of interest, *if the null hypothesis were true*, is called the **Null Distribution**
- We can use R to approximate the null distribution by running 2000 experiments of 8 coin flips:

Approximating the Null Distribution

- The distribution of the statistic of interest, *if the null hypothesis were true*, is called the **Null Distribution**

- We can use R to approximate the null distribution by running 2000 experiments of 8 coin flips:

```
coin %>% rep_sample_n(size = 8, replace = T, reps = 2000) %>%
  summarize(n_heads = sum(face == "Heads")) %>% mutate(p_hat = n_heads/8)
```

```
## # A tibble: 2,000 x 3
##    replicate n_heads p_hat
##        <int>   <int> <dbl>
##  1         1       5 0.625
##  2         2       5 0.625
##  3         3       4 0.5
##  4         4       4 0.5
##  5         5       3 0.375
##  6         6       3 0.375
##  7         7       3 0.375
##  8         8       2 0.25
##  9         9       3 0.375
## 10        10       2 0.25
## # ... with 1,990 more rows
```

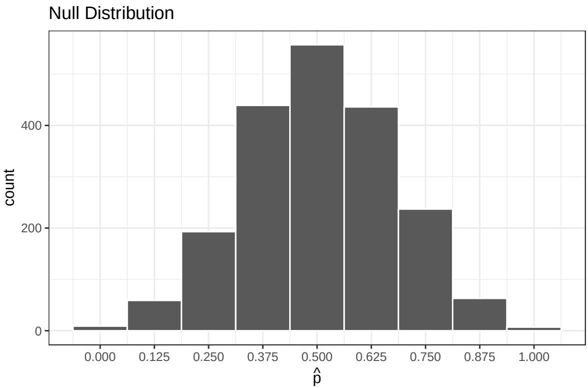
Visualizing the Null Distribution

- We can use a histogram to visualize the Null Distribution of the sample proportion $\hat{p}$

Visualizing the Null Distribution

- We can use a histogram to visualize the Null Distribution of the sample proportion $\hat{p}$

```
null_stats %>% ggplot(aes(x = p_hat))+geom_histogram(bins = 9, color = "white")
```



P-Values

- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.

P-Values

- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.
- To distinguish between sample statistics generally and the particular one obtained from the sample, we call the latter the **test statistic**

P-Values

- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.
- To distinguish between sample statistics generally and the particular one obtained from the sample, we call the latter the **test statistic**
    - In the prior experiment, we flipped a coin 8 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.

P-Values

- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.
- To distinguish between sample statistics generally and the particular one obtained from the sample, we call the latter the **test statistic**
  - In the prior experiment, we flipped a coin 8 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.
  - The p-value for this test statistic is

$$\text{Probability of at least 8 heads in 8 flips} = 0.5^8 = 0.0039$$

P-Values

- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.
- To distinguish between sample statistics generally and the particular one obtained from the sample, we call the latter the **test statistic**
  - In the prior experiment, we flipped a coin 8 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.
  - The p-value for this test statistic is

$$\text{Probability of at least 8 heads in 8 flips} = 0.5^8 = 0.0039$$

- The p-value quantifies the strength of evidence against the Null Hypothesis. Smaller p-values represent stronger evidence to reject H_0.

P-Values

- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.
- To distinguish between sample statistics generally and the particular one obtained from the sample, we call the latter the **test statistic**
  - In the prior experiment, we flipped a coin 8 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.
  - The p-value for this test statistic is

$$\text{Probability of at least 8 heads in 8 flips} = 0.5^8 = 0.0039$$

- The p-value quantifies the strength of evidence against the Null Hypothesis. Smaller p-values represent stronger evidence to reject H_0.
  - P-values very close to 0 represent statistics that were very unlikely to arise by chance, if the null hypothesis were true.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
  - Then calculate the proportion of simulated statistics as extreme as the test statistic.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
  - Then calculate the proportion of simulated statistics as extreme as the test statistic.

```
null_stats %>% mutate(extreme = ifelse(p_hat >=1.0, "yes", "no")) %>%
  group_by(extreme) %>% summarize(n = n()) %>%
  mutate(proportion = n/sum(n))
```

```
## # A tibble: 2 x 3
##   extreme     n proportion
##   <chr>   <int>      <dbl>
## 1 no       1993     0.996
## 2 yes         7     0.0035
```

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
    - Then calculate the proportion of simulated statistics as extreme as the test statistic.

```
null_stats %>% mutate(extreme = ifelse(p_hat >=1.0, "yes", "no")) %>%
  group_by(extreme) %>% summarize(n = n()) %>%
  mutate(proportion = n/sum(n))
```

```
## # A tibble: 2 x 3
##   extreme      n proportion
##   <chr>    <int>      <dbl>
## 1 no        1993     0.996
## 2 yes          7     0.0035
```

- Method 2: We use theory-based tools to create the theoretical null distribution.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
  - Then calculate the proportion of simulated statistics as extreme as the test statistic.

```
null_stats %>% mutate(extreme = ifelse(p_hat >=1.0, "yes", "no")) %>%
  group_by(extreme) %>% summarize(n = n()) %>%
  mutate(proportion = n/sum(n))
```

```
## # A tibble: 2 x 3
##   extreme     n proportion
##   <chr>   <int>      <dbl>
## 1 no       1993     0.996
## 2 yes         7     0.0035
```

- Method 2: We use theory-based tools to create the theoretical null distribution.
  - Then use the model to calculate the theoretical probability of observing a sample statistic as extreme as the test statistic.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
  - Then calculate the proportion of simulated statistics as extreme as the test statistic.

```
null_stats %>% mutate(extreme = ifelse(p_hat >=1.0, "yes", "no")) %>%
  group_by(extreme) %>% summarize(n = n()) %>%
  mutate(proportion = n/sum(n))
```

```
## # A tibble: 2 x 3
##   extreme     n proportion
##   <chr>   <int>      <dbl>
## 1 no       1993     0.996
## 2 yes         7     0.0035
```

- Method 2: We use theory-based tools to create the theoretical null distribution.
  - Then use the model to calculate the theoretical probability of observing a sample statistic as extreme as the test statistic.
  - Assuming that coin flips heads with probability 0.5 and that each flip is independent of the others, then the probability of 8 consecutive heads is

```
0.5^8
```

```
## [1] 0.00390625
```

P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?

P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
    - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.

P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?

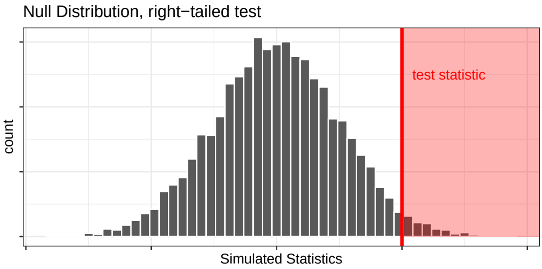
P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?
  - Yes! The **direction** of the alternative hypotheses determines which "tail(s)" of the null distribution correspond to *extreme* values.

P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?
  - Yes! The **direction** of the alternative hypotheses determines which "tail(s)" of the null distribution correspond to *extreme* values.

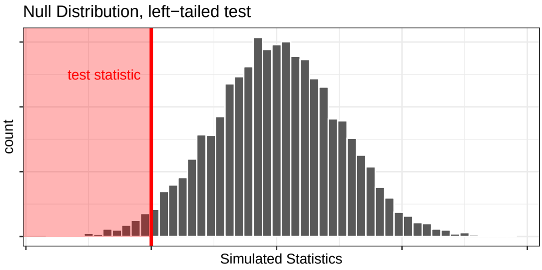
1. If H_a is of the form $\text{parameter} > \text{null value}$, then the p-value is the proportion of simulated statistics greater than or equal to the test statistic (i.e. the right tail)



P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?
  - Yes! The **direction** of the alternative hypotheses determines which "tail(s)" of the null distribution correspond to *extreme* values.

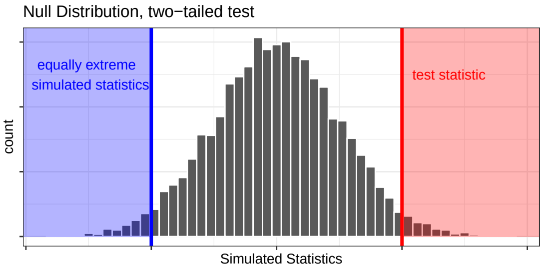
2. If H_a is of the form $\text{parameter} < \text{null value}$, then the p-value is the proportion of simulated statistics less than or equal to the test statistic (i.e. the left tail)



P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?
  - Yes! The **direction** of the alternative hypotheses determines which "tail(s)" of the null distribution correspond to *extreme* values.

3. If H_a is of the form $\text{parameter} \neq \text{null value}$, then the p-value is twice the proportion of simulated statistics more extreme than the test statistic (i.e. both tails)



A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.

A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:

$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$

A Two-Tailed Example

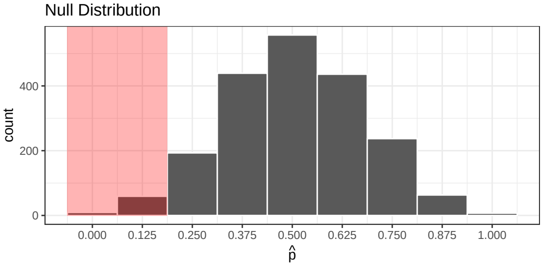
- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:
$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$
- We flip the coin 8 times and obtain 1 heads, for a proportion $\hat{p} = 0.125$.

A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:

$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$

- We flip the coin 8 times and obtain 1 heads, for a proportion $\hat{p} = 0.125$.
- Using the previous null-distribution, we shade values that are as extreme as our statistic:

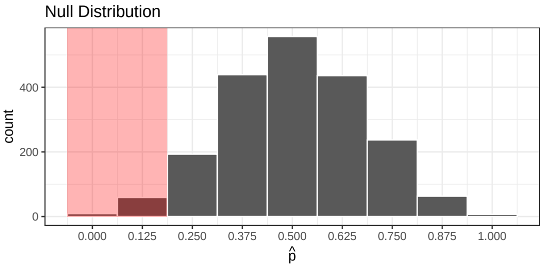


A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:

$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$

- We flip the coin 8 times and obtain 1 heads, for a proportion $\hat{p} = 0.125$.
- Using the previous null-distribution, we shade values that are as extreme as our statistic:



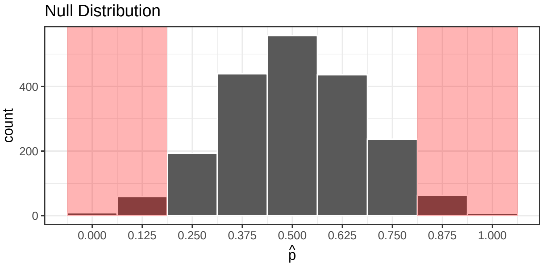
- We find the proportion of simulated statistics in the left tail is 0.034

A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:

$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$

- We flip the coin 8 times and obtain 1 heads, for a proportion $\hat{p} = 0.125$.
- Using the previous null-distribution, we shade values that are as extreme as our statistic:



- We double this to include the right-tail as well, and get a p-value of 0.068.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.
  - In the coin flip experiment, $\text{P-value} < \alpha$ since $0.03125 < 0.05$

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.
  - In the coin flip experiment, $\text{P-value} < \alpha$ since $0.03125 < 0.05$
  - Our test was statistically significant and we reject the hypothesis that the coin is fair in favor of the hypothesis that the coin is more likely to flip heads.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.
  - In the coin flip experiment, $\text{P-value} < \alpha$ since $0.03125 < 0.05$
  - Our test was statistically significant and we reject the hypothesis that the coin is fair in favor of the hypothesis that the coin is more likely to flip heads.
- We should **always** choose the value of α prior to conducting an experiment and observing data. *Usually* the choice is made for us depending on conventions in our field of study.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.
  - In the coin flip experiment, $\text{P-value} < \alpha$ since $0.03125 < 0.05$
  - Our test was statistically significant and we reject the hypothesis that the coin is fair in favor of the hypothesis that the coin is more likely to flip heads.
- We should **always** choose the value of α prior to conducting an experiment and observing data. *Usually* the choice is made for us depending on conventions in our field of study.
- Choosing a significance level of $\alpha = 0.05$ means that we treat any result that would have occurred by chance alone less than 5% of the time as good evidence that the null hypothesis is false.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
    - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

- A **Type 1 Error** occurs when we reject H_0 when it is actually true.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

- A **Type 1 Error** occurs when we reject H_0 when it is actually true.
  - The coin is actually fair. But we saw an unlikely event and claimed the coin was biased.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

- A **Type 1 Error** occurs when we reject H_0 when it is actually true.
  - The coin is actually fair. But we saw an unlikely event and claimed the coin was biased.
- A **Type 2 Error** occurs when we fail to reject H_0 when it is in fact false.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

- A **Type 1 Error** occurs when we reject H_0 when it is actually true.
  - The coin is actually fair. But we saw an unlikely event and claimed the coin was biased.
- A **Type 2 Error** occurs when we fail to reject H_0 when it is in fact false.
  - The coin was indeed biased. But we withheld judgment since unlikely events do happen from time to time.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
  - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
    - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.
- The **power** of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. That is

$$\text{Power} = 1 - \text{Probability of Type II Error}$$

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
  - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.
- The **power** of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. That is

$$\text{Power} = 1 - \text{Probability of Type II Error}$$

  - In general, computing power can be difficult, and requires we investigate the distribution of a sample statistic under the alternative hypothesis.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
    - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.
- The **power** of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. That is

$$\text{Power} = 1 - \text{Probability of Type II Error}$$

    - In general, computing power can be difficult, and requires we investigate the distribution of a sample statistic under the alternative hypothesis.

With great power comes...

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
  - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.
- The **power** of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. That is

$$\text{Power} = 1 - \text{Probability of Type II Error}$$

  - In general, computing power can be difficult, and requires we investigate the distribution of a sample statistic under the alternative hypothesis.

With great power comes...greater chance of Type I error.

Rapid COVID test

A quick and accessible (but unreliable) test for COVID-19 is to match a patient's symptoms to the 10 most common symptoms exhibited by victims of COVID.

Suppose a person walks into a medical clinic with 6 of the 10 symptoms of COVID, and medical personnel are concerned the person may have COVID.

1. What are the Null and Alternate Hypotheses in this case?

Rapid COVID test

A quick and accessible (but unreliable) test for COVID-19 is to match a patient's symptoms to the 10 most common symptoms exhibited by victims of COVID.

Suppose a person walks into a medical clinic with 6 of the 10 symptoms of COVID, and medical personnel are concerned the person may have COVID.

1. What are the Null and Alternate Hypotheses in this case?
2. What 'statistic' is being used to determine whether the person has COVID

Rapid COVID test

A quick and accessible (but unreliable) test for COVID-19 is to match a patient's symptoms to the 10 most common symptoms exhibited by victims of COVID.

Suppose a person walks into a medical clinic with 6 of the 10 symptoms of COVID, and medical personnel are concerned the person may have COVID.

1. What are the Null and Alternate Hypotheses in this case?
2. What 'statistic' is being used to determine whether the person has COVID
3. In the context of this problem, what does a Type I error represent? What are some possible consequences of a Type I error?

Rapid COVID test

A quick and accessible (but unreliable) test for COVID-19 is to match a patient's symptoms to the 10 most common symptoms exhibited by victims of COVID.

Suppose a person walks into a medical clinic with 6 of the 10 symptoms of COVID, and medical personnel are concerned the person may have COVID.

1. What are the Null and Alternate Hypotheses in this case?
2. What 'statistic' is being used to determine whether the person has COVID
3. In the context of this problem, what does a Type I error represent? What are some possible consequences of a Type I error?
4. Similarly, what does a Type II error represent? What are some possible consequences of a Type II error?

Rapid COVID test

A quick and accessible (but unreliable) test for COVID-19 is to match a patient's symptoms to the 10 most common symptoms exhibited by victims of COVID.

Suppose a person walks into a medical clinic with 6 of the 10 symptoms of COVID, and medical personnel are concerned the person may have COVID.

1. What are the Null and Alternate Hypotheses in this case?
2. What 'statistic' is being used to determine whether the person has COVID
3. In the context of this problem, what does a Type I error represent? What are some possible consequences of a Type I error?
4. Similarly, what does a Type II error represent? What are some possible consequences of a Type II error?
5. What significance level are you willing to use for this COVID test? *Remember, decreasing significance level also decreases the power of the test.*

DNA Tests

DNA testing allows researchers to compare markers in a person's DNA to those found at crime scene. Suppose the DNA found at a crime scene will **always** match the perpetrator of the crime. However, there is a small chance that the crime scene DNA will also match the markers for another innocent person.

Suppose a person is on trial for a crime. Forensic scientists attest that the person's DNA matches that found at the crime scene.

1. What are the Null and Alternate Hypotheses in this case?

DNA Tests

DNA testing allows researchers to compare markers in a person's DNA to those found at crime scene. Suppose the DNA found at a crime scene will **always** match the perpetrator of the crime. However, there is a small chance that the crime scene DNA will also match the markers for another innocent person.

Suppose a person is on trial for a crime. Forensic scientists attest that the person's DNA matches that found at the crime scene.

1. What are the Null and Alternate Hypotheses in this case?
2. What ‘statistic‘ is being used to determine whether the person has committed the crime.

DNA Tests

DNA testing allows researchers to compare markers in a person's DNA to those found at crime scene. Suppose the DNA found at a crime scene will **always** match the perpetrator of the crime. However, there is a small chance that the crime scene DNA will also match the markers for another innocent person.

Suppose a person is on trial for a crime. Forensic scientists attest that the person's DNA matches that found at the crime scene.

1. What are the Null and Alternate Hypotheses in this case?
2. What 'statistic' is being used to determine whether the person has committed the crime.
3. In the context of this problem, what does a Type I error represent? What are some possible consequences of a Type I error?

DNA Tests

DNA testing allows researchers to compare markers in a person's DNA to those found at crime scene. Suppose the DNA found at a crime scene will **always** match the perpetrator of the crime. However, there is a small chance that the crime scene DNA will also match the markers for another innocent person.

Suppose a person is on trial for a crime. Forensic scientists attest that the person's DNA matches that found at the crime scene.

1. What are the Null and Alternate Hypotheses in this case?
2. What 'statistic' is being used to determine whether the person has committed the crime.
3. In the context of this problem, what does a Type I error represent? What are some possible consequences of a Type I error?
4. Similarly, what does a Type II error represent? What are some possible consequences of a Type II error?

DNA Tests

DNA testing allows researchers to compare markers in a person's DNA to those found at crime scene. Suppose the DNA found at a crime scene will **always** match the perpetrator of the crime. However, there is a small chance that the crime scene DNA will also match the markers for another innocent person.

Suppose a person is on trial for a crime. Forensic scientists attest that the person's DNA matches that found at the crime scene.

1. What are the Null and Alternate Hypotheses in this case?
2. What 'statistic' is being used to determine whether the person has committed the crime.
3. In the context of this problem, what does a Type I error represent? What are some possible consequences of a Type I error?
4. Similarly, what does a Type II error represent? What are some possible consequences of a Type II error?
5. What significance level are you willing to use for this DNA test? *Remember, decreasing significance level also decreases the power of the test.*