Overview of Hypothesis Testing

Engineering Statistics
Section 8.1

Josh Engwer

TTU

08 April 2016

PART I:

MOTIVATING THE NEED FOR HYPOTHESIS TESTING

The Need for Hypothesis Testing

Very often, one does not care about a point estimate nor a confidence interval of a population parameter.

Instead, the interest will be in deciding which of two contradictory claims about a population parameter is correct/reasonable.

Such a decision can be made using a Statistical Inference method called **Hypothesis Testing**.

1^{st} Motivating Example for Hypothesis Testing

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

1^{st} Motivating Example for Hypothesis Testing

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

So how do you decide if this claim is reasonable or not??

1^{st} Motivating Example for Hypothesis Testing

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

So how do you decide if this claim is reasonable or not??
TAKE A SAMPLE!

1^{st} Motivating Example for Hypothesis Testing

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

So how do you decide if this claim is reasonable or not??
TAKE A SAMPLE!

You randomly select a sample of 100 bottles from your warehouse stock.
You measure the amount of soda in each bottle and compute the mean:

$$\bar{x}_{soda} = 16.6 \text{ oz}$$

Would you believe the stores' claim?? Yes, probably so!

1^{st} Motivating Example for Hypothesis Testing

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

So how do you decide if this claim is reasonable or not??
TAKE A SAMPLE!

Suppose instead you randomly select a sample of 100 bottles from your stock.
You measure the amount of soda in each bottle and compute the mean:

$$\bar{x}_{soda} = 19.8 \text{ oz}$$

Would you believe the stores' claim?? No, probably not!

2^{nd} Motivating Example for Hypothesis Testing

Suppose your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.

2^{nd} Motivating Example for Hypothesis Testing

Suppose your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.

Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$

But Joe is claiming that $p_{six} = 0.80$ (!!)

So how do you decide if Joe's claim is reasonable or not??

2^{nd} Motivating Example for Hypothesis Testing

Suppose your friend Joe claims he has a fair six-sided die that's ”magical” in the sense that he rolls a ”6” every four out of five rolls.

Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$

But Joe is claiming that $p_{six} = 0.80$ (!!)

So how do you decide if Joe's claim is reasonable or not??
TAKE A SAMPLE!!

2^{nd} Motivating Example for Hypothesis Testing

Suppose your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.

Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$

But Joe is claiming that $p_{six} = 0.80$ (!!)

So how do you decide if Joe's claim is reasonable or not??
TAKE A SAMPLE!!

So you roll the "magical" die ten times and it showed "6" only two times.
This means the point estimate is: $\widehat{p}_{six} = 0.2$

Would you believe Joe's claim?? No, probably not!

2^{nd} Motivating Example for Hypothesis Testing

Suppose your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.

Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$

But Joe is claiming that $p_{six} = 0.80$ (!!)

So how do you decide if Joe's claim is reasonable or not??
TAKE A SAMPLE!!

Suppose instead when you rolled the die ten times it showed "6" seven times.
This means the point estimate is: $\widehat{p}_{six} = 0.7$

Would you believe Joe's claim?? Yes, probably so!

2^{nd} Motivating Example for Hypothesis Testing

Suppose your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.

Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$

But Joe is claiming that $p_{six} = 0.80 \quad$ (!!)

So how do you decide if Joe's claim is reasonable or not??
TAKE A SAMPLE!!

Suppose instead when you rolled the die ten times it showed "6" seven times.
This means the point estimate is: $\widehat{p}_{six} = 0.7$

Would you believe Joe's claim?? Yes, probably so!

But this still leaves some questions:

- How does one pick the "cutoff"??
  - Some people would believe Joe with a lower point estimate like $\widehat{p}_{six} = 0.6$
  - Some people are more skeptical and would only believe Joe if $\widehat{p}_{six} = 0.8$
- How do we know that this particular point estimate was not a "fluke"??
  - i.e. Did an extreme point estimate like $\widehat{p}_{six} = 0.7$ happen solely by chance... (...rather than happen because the die is "magical"??)

3^{rd} Motivating Example for Hypothesis Testing

Suppose you have a well on your property.
Last year you tested the well water for its pH variability and found: $\sigma_{H_2O} = 0.1$
This means the well water pH hovers closely around a normal pH of 7.0.

But lately you've noticed that the well water sometimes tastes very acidic (high pH) or very alkaline (low pH), both of which can cause health problems!!
So now you suspect that the pH variability is much larger: $\sigma_{H_2O} > 0.1$

So how do you decide if your claim is reasonable or not??
TAKE A SAMPLE!!

So you draw ten buckets of water from the well and measure their pH values.
Then you compute the sample std dev: $s_{H_2O} = 0.14$

Would you believe your own claim?? No, probably not!

3^{rd} Motivating Example for Hypothesis Testing

Suppose you have a well on your property.
Last year you tested the well water for its pH variability and found: $\sigma_{H_2O} = 0.1$
This means the well water pH hovers closely around a normal pH of 7.0.

But lately you've noticed that the well water sometimes tastes very acidic (high pH) or very alkaline (low pH), both of which can cause health problems!!
So now you suspect that the pH variability is much larger: $\sigma_{H_2O} > 0.1$

So how do you decide if your claim is reasonable or not??
TAKE A SAMPLE!!

So you draw ten buckets of water from the well and measure their pH values.
Then you compute the sample std dev: $s_{H_2O} = 1.4$

Would you believe your own claim?? Yes, probably so!

But this still leaves some questions:

- How does one pick the "cutoff"??
  - Some people would believe you with a lower point estimate like $s_{H_2O} = 0.14$
  - Some people are more skeptical and would only believe you if $s_{H_2O} = 2.0$
- How do we know that this particular point estimate was not a "fluke"??
  - i.e. Did an extreme point estimate like $s_{H_2O} = 1.4$ happen solely by chance... (...rather than happen because the well water is tainted??)

PART II:

HYPOTHESIS TESTING: DEFINITION, TERMINOLOGY, PROCEDURE

Null Hypothesis & Alternative Hypothesis

Definition

The **null hypothesis**, denoted H_0, is the claim initially assumed to be true.
The **alternative hypothesis**, denoted H_A, is a claim that contradicts H_0.

A **hypothesis test** is a method that uses sample data & probability to decide whether the null hypothesis is a reasonable statement.

If sample evidence strongly contradicts the null hypothesis H_0,
then the null hypothesis will be rejected in favor of the alternative hypothesis.

Otherwise, if the sample evidence does not strongly contradict H_0,
then it's reasonable to continue to believe that H_0 is still quite plausible.

i.e. THINK: "Innocent until proven guilty."
i.e. THINK: The Scientific Method

The two possible conclusions from a test are: "reject H_0" OR "accept H_0"

NOTE: Sometimes one says "fail to reject H_0" instead of "accept H_0".

Statistical Hypotheses

In Statistical Inference, hypothesis tests involve one or more populations:

Definition

A **statistical hypothesis** is a testable claim about facet(s) of population(s):

- The value of a single parameter (Chapter 8) (i.e. this chapter)
- The values of two parameters (Chapter 9)

- The values of many parameters (Chapters 10-11)
- The shape of fitted models (Chapters 12-13)
- The distribution type (Chapter 14)

Examples of statistical hypotheses:

- The value of a single population parameter:

  $\mu = 3, \quad \mu \geq 7, \quad p = 0.4, \quad p < 0.9, \quad \sigma^2 \leq 2, \quad \sigma > 2, \quad \sigma^2 \neq 1,$ etc...

- The values of two population parameters:

  $\mu_1 = \mu_2, \quad \mu_1 \geq \mu_2, \quad p_1 - p_2 < 0.4, \quad \sigma_1^2/\sigma_2^2 \leq 2, \quad \sigma_1/\sigma_2 \neq 1,$ etc...

The Correct Format for Hypothesis Tests

Hypothesis tests must conform to the following rules:

- Population parameter(s) must be involved.
- Statistics must not be involved.
- The null hypothesis H_0 must only involve equality.
- The alternative hypothesis H_A must not include equality.
- The asserted value in H_0 should also appear in H_A.

COMPLIANT GENERIC 1-SAMPLE HYPOTHESIS TESTS: $(\theta_0 \in \mathbb{R})$

$$\begin{aligned} H_0 &: \ \theta = \theta_0 \\ H_A &: \ \theta > \theta_0 \end{aligned} \qquad \begin{aligned} H_0 &: \ \theta = \theta_0 \\ H_A &: \ \theta < \theta_0 \end{aligned} \qquad \begin{aligned} H_0 &: \ \theta = \theta_0 \\ H_A &: \ \theta \neq \theta_0 \end{aligned}$$

COMPLIANT GENERIC 2-SAMPLE HYPOTHESIS TESTS: $(\delta_0 \in \mathbb{R})$

$$\begin{aligned} H_0 &: \ \theta_1 - \theta_2 = \delta_0 \\ H_A &: \ \theta_1 - \theta_2 > \delta_0 \end{aligned} \qquad \begin{aligned} H_0 &: \ \theta_1 - \theta_2 = \delta_0 \\ H_A &: \ \theta_1 - \theta_2 < \delta_0 \end{aligned} \qquad \begin{aligned} H_0 &: \ \theta_1 - \theta_2 = \delta_0 \\ H_A &: \ \theta_1 - \theta_2 \neq \delta_0 \end{aligned}$$

$$\begin{aligned} H_0 &: \ \theta_1/\theta_2 = \delta_0 \\ H_A &: \ \theta_1/\theta_2 > \delta_0 \end{aligned} \qquad \begin{aligned} H_0 &: \ \theta_1/\theta_2 = \delta_0 \\ H_A &: \ \theta_1/\theta_2 < \delta_0 \end{aligned} \qquad \begin{aligned} H_0 &: \ \theta_1/\theta_2 = \delta_0 \\ H_A &: \ \theta_1/\theta_2 \neq \delta_0 \end{aligned}$$

The Correct Format for Hypothesis Tests

Hypothesis tests must conform to the following rules:

- Population parameter(s) must be involved.
- Statistics must not be involved.
- The null hypothesis H_0 must only involve equality.
- The alternative hypothesis H_A must not include equality.
- The asserted value in H_0 should also appear in H_A.

COMPLIANT PARTICULAR 1-SAMPLE HYPOTHESIS TESTS:

$$\begin{aligned} H_0 &: \quad \mu = 3.1 \\ H_A &: \quad \mu > 3.1 \end{aligned} \qquad \begin{aligned} H_0 &: \quad p = 0.72 \\ H_A &: \quad p < 0.72 \end{aligned} \qquad \begin{aligned} H_0 &: \quad \sigma = 1.9 \\ H_A &: \quad \sigma \neq 1.9 \end{aligned}$$

COMPLIANT PARTICULAR 2-SAMPLE HYPOTHESIS TESTS:

$$\begin{aligned} H_0 &: \quad \mu_1 - \mu_2 = 0 \\ H_A &: \quad \mu_1 - \mu_2 > 0 \end{aligned} \qquad \begin{aligned} H_0 &: \quad p_1 - p_2 = 0.1 \\ H_A &: \quad p_1 - p_2 < 0.1 \end{aligned} \qquad \begin{aligned} H_0 &: \quad p_1 - p_2 = 0 \\ H_A &: \quad p_1 - p_2 \neq 0 \end{aligned}$$

$$\begin{aligned} H_0 &: \quad \sigma_1^2/\sigma_2^2 = 2 \\ H_A &: \quad \sigma_1^2/\sigma_2^2 > 2 \end{aligned} \qquad \begin{aligned} H_0 &: \quad \sigma_1/\sigma_2 = 0.6 \\ H_A &: \quad \sigma_1/\sigma_2 < 0.6 \end{aligned} \qquad \begin{aligned} H_0 &: \quad \sigma_1/\sigma_2 = 1 \\ H_A &: \quad \sigma_1/\sigma_2 \neq 1 \end{aligned}$$

Hypothesis Test for the 1^{st} Motivating Example

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

The corresponding compliant hypothesis test can be written as:

$$H_0: \quad \mu_{soda} = 20$$
$$H_A: \quad \mu_{soda} < 20$$

————OR————

$H_0: \quad \mu_{soda} = 20$ versus $H_A: \quad \mu_{soda} < 20$

————OR————

$H_0: \quad \mu_{soda} = 20$ against $H_A: \quad \mu_{soda} < 20$

Hypothesis Test for the 2^{nd} Motivating Example

Your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.
Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$
But Joe is claiming that $p_{six} = 0.80$ (!!)

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad p_{six} = 0.167 \\ H_A &: \quad p_{six} > 0.167 \end{aligned}$$

——————OR——————

$$H_0 : \quad p_{six} = 0.167 \quad \text{versus} \quad H_A : \quad p_{six} > 0.167$$

——————OR——————

$$H_0 : \quad p_{six} = 0.167 \quad \text{against} \quad H_A : \quad p_{six} > 0.167$$

Hypothesis Test for the 3^{rd} Motivating Example

Suppose you have a well on your property.
Last year you tested the well water for its pH variability and found: $\sigma_{H_2O} = 0.1$
This means the well water pH hovers closely around a normal pH of 7.0.

But lately you've noticed that the well water sometimes tastes very acidic (high pH) or very alkaline (low pH), both of which can cause health problems!!
So now you suspect that the pH variability is much larger: $\sigma_{H_2O} > 0.1$

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad \sigma_{H_2O} = 0.1 \\ H_A &: \quad \sigma_{H_2O} > 0.1 \end{aligned}$$

OR

$H_0 : \quad \sigma_{H_2O} = 0.1$ versus $H_A : \quad \sigma_{H_2O} > 0.1$

OR

$H_0 : \quad \sigma_{H_2O} = 0.1$ against $H_A : \quad \sigma_{H_2O} > 0.1$

Errors in Hypothesis Testing: Type I & Type II Errors

Alas, since only a small portion of the population is sampled rather than the entire population, a hypothesis test may lead to one of two types of errors:

Definition

A **Type I error** consists of rejecting null hypothesis H_0 when it is true.
A **Type II error** consists of not rejecting null hypothesis H_0 when it is false.

The probability of a Type I error is denoted by α.
The probability of a Type II error is denoted by β.

$$\begin{aligned} \alpha &= \mathbb{P}(\text{Type I Error}) &= \mathbb{P}(\text{Rejecting } H_0 \text{ when } H_0 \text{ is true}) \\ \beta &= \mathbb{P}(\text{Type II Error}) &= \mathbb{P}(\text{Accepting } H_0 \text{ when } H_0 \text{ is false}) \end{aligned}$$

	Reject H_0	Accept H_0
H_0 is true	**Type I Error**	**(Correct Decision)**
H_A is true	**(Correct Decision)**	**Type II Error**

In some scenarios, it's easier to think of a Type I Error as a **false positive**.
In some scenarios, it's easier to think of a Type II Error as a **false negative**.

Based on the scenario, one error type may be more serious than the other.

Example of Type I & Type II Errors from US Law

Consider the scenario of a suspect being tried for a petty crime in the US:

$$\begin{aligned} H_0 &: \quad \text{Suspect is innocent} \\ H_A &: \quad \text{Suspect is guilty} \end{aligned}$$

Then:

- Type I Error would be the court deciding the suspect is guilty when in reality the suspect is innocent.
- Type II Error would be the court deciding the suspect is innocent when in reality the suspect is guilty.

	Decide Suspect is Guilty	Decide Suspect is Innocent
Suspect is really Innocent	**Type I Error**	**(Correct Decision)**
Suspect is really Guilty	**(Correct Decision)**	**Type II Error**

Example of Type I & Type II Errors from US Law

Consider the scenario of a suspect being tried for a petty crime in the US:

$$\begin{aligned} H_0 &: \quad \text{Suspect is innocent} \\ H_A &: \quad \text{Suspect is guilty} \end{aligned}$$

Then:

- Type I Error would be the court deciding the suspect is guilty when in reality the suspect is innocent.
- Type II Error would be the court deciding the suspect is innocent when in reality the suspect is guilty.

In this scenario, which error type is more serious??

Example of Type I & Type II Errors from US Law

Consider the scenario of a suspect being tried for a petty crime in the US:

$$\begin{aligned} H_0 &: \quad \text{Suspect is innocent} \\ H_A &: \quad \text{Suspect is guilty} \end{aligned}$$

Then:

- Type I Error would be the court deciding the suspect is guilty when in reality the suspect is innocent.
- Type II Error would be the court deciding the suspect is innocent when in reality the suspect is guilty.

In this scenario, which error type is more serious?? **It depends!**

- Friends/family of the suspect probably view Type I Error as more serious.
- Everyone else probably view Type II Error as more serious.

Type I & Type II Errors from 1^{st} Motivating Example

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

The corresponding compliant hypothesis test can be written as:

$$
\begin{aligned}
H_0 : & \quad \mu_{soda} = 20 \\
H_A : & \quad \mu_{soda} < 20
\end{aligned}
$$

Then:

- Type I Error would be ....
- Type II Error would be ....

Type I & Type II Errors from 1^{st} Motivating Example

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad \mu_{soda} = 20 \\ H_A &: \quad \mu_{soda} < 20 \end{aligned}$$

Then:

- Type I Error would be deciding the soda bottles are being systematically under-filled when in fact they are correctly filled.
- Type II Error would be deciding the soda bottles are being correctly filled when in fact they are being systematically under-filled.

Type I & Type II Errors from 1^{st} Motivating Example

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad \mu_{soda} = 20 \\ H_A &: \quad \mu_{soda} < 20 \end{aligned}$$

Then:

- Type I Error would be deciding the soda bottles are being systematically under-filled when in fact they are correctly filled.
- Type II Error would be deciding the soda bottles are being correctly filled when in fact they are being systematically under-filled.

In this scenario, which error type is more serious??

Type I & Type II Errors from 1^{st} Motivating Example

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

The corresponding compliant hypothesis test can be written as:

$$\begin{array}{rl} H_0: & \mu_{soda} = 20 \\ H_A: & \mu_{soda} < 20 \end{array}$$

Then:

- Type I Error would be deciding the soda bottles are being systematically under-filled when in fact they are correctly filled.
- Type II Error would be deciding the soda bottles are being correctly filled when in fact they are being systematically under-filled.

In this scenario, which error type is more serious??

Type II Error!

(since the result would be a drop in sales/profit and possibly lawsuits!!)

Type I & Type II Errors from 2^{nd} Motivating Example

Your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.
Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$
But Joe is claiming that $p_{six} = 0.80$ (!!)

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad p_{six} = 0.167 \\ H_A &: \quad p_{six} > 0.167 \end{aligned}$$

Then:

- Type I Error would be ....
- Type II Error would be ....

Type I & Type II Errors from 2^{nd} Motivating Example

Your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.
Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$
But Joe is claiming that $p_{six} = 0.80$ (!!)

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad p_{six} = 0.167 \\ H_A &: \quad p_{six} > 0.167 \end{aligned}$$

Then:

- Type I Error would be deciding the die is "magical" when in fact it's not.
- Type II Error would be deciding the die is not "magical" when in fact it is.

Type I & Type II Errors from 2^{nd} Motivating Example

Your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.
Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$
But Joe is claiming that $p_{six} = 0.80 \quad$ (!!)

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad p_{six} = 0.167 \\ H_A &: \quad p_{six} > 0.167 \end{aligned}$$

Then:

- Type I Error would be deciding the die is "magical" when in fact it's not.
- Type II Error would be deciding the die is not "magical" when in fact it is.

In this scenario, which error type is more serious??

Type I & Type II Errors from 2^{nd} Motivating Example

Your friend Joe claims he has a fair six-sided die that's "magical" in the sense that he rolls a "6" every four out of five rolls.
Now, you know from Chapter 2 that $p_{six} = \mathbb{P}(\text{Rolling a "6"}) = 1/6 \approx 0.167$
But Joe is claiming that $p_{six} = 0.80$ (!!)

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad p_{six} = 0.167 \\ H_A &: \quad p_{six} > 0.167 \end{aligned}$$

Then:

- Type I Error would be deciding the die is "magical" when in fact it's not.
- Type II Error would be deciding the die is not "magical" when in fact it is.

In this scenario, which error type is more serious??

Type I Error!

(Since in reality there's no such thing as "magical" dice, Joe deceived you!)

Type I & Type II Errors from 3^{rd} Motivating Example

Suppose you have a well on your property.
Last year you tested the well water for its pH variability and found: $\sigma_{H_2O} = 0.1$
This means the well water pH hovers closely around a normal pH of 7.0.

But lately you've noticed that the well water sometimes tastes very acidic (high pH) or very alkaline (low pH), both of which can cause health problems!!
So now you suspect that the pH variability is much larger: $\sigma_{H_2O} > 0.1$

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad \sigma_{H_2O} = 0.1 \\ H_A &: \quad \sigma_{H_2O} > 0.1 \end{aligned}$$

Then:

- Type I Error would be ...
- Type II Error would be ...

Type I & Type II Errors from 3^{rd} Motivating Example

Suppose you have a well on your property.
Last year you tested the well water for its pH variability and found: $\sigma_{H_2O} = 0.1$
This means the well water pH hovers closely around a normal pH of 7.0.

But lately you've noticed that the well water sometimes tastes very acidic (high pH) or very alkaline (low pH), both of which can cause health problems!!
So now you suspect that the pH variability is much larger: $\sigma_{H_2O} > 0.1$

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad \sigma_{H_2O} = 0.1 \\ H_A &: \quad \sigma_{H_2O} > 0.1 \end{aligned}$$

Then:

- Type I Error would be deciding the water is tainted when in fact it's not.
- Type II Error would be deciding the water is not tainted when in fact it is.

Type I & Type II Errors from 3^{rd} Motivating Example

Suppose you have a well on your property.
Last year you tested the well water for its pH variability and found: $\sigma_{H_2O} = 0.1$
This means the well water pH hovers closely around a normal pH of 7.0.

But lately you've noticed that the well water sometimes tastes very acidic (high pH) or very alkaline (low pH), both of which can cause health problems!!
So now you suspect that the pH variability is much larger: $\sigma_{H_2O} > 0.1$

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad \sigma_{H_2O} = 0.1 \\ H_A &: \quad \sigma_{H_2O} > 0.1 \end{aligned}$$

Then:

- Type I Error would be deciding the water is tainted when in fact it's not.
- Type II Error would be deciding the water is not tainted when in fact it is.

In this scenario, which error type is more serious??

Type I & Type II Errors from 3^{rd} Motivating Example

Suppose you have a well on your property.
Last year you tested the well water for its pH variability and found: $\sigma_{H_2O} = 0.1$
This means the well water pH hovers closely around a normal pH of 7.0.

But lately you've noticed that the well water sometimes tastes very acidic (high pH) or very alkaline (low pH), both of which can cause health problems!!
So now you suspect that the pH variability is much larger: $\sigma_{H_2O} > 0.1$

The corresponding compliant hypothesis test can be written as:

$$\begin{aligned} H_0 &: \quad \sigma_{H_2O} = 0.1 \\ H_A &: \quad \sigma_{H_2O} > 0.1 \end{aligned}$$

Then:

- Type I Error would be deciding the water is tainted when in fact it's not.
- Type II Error would be deciding the water is not tainted when in fact it is.

In this scenario, which error type is more serious??

Type II Error!!

(Since a Type II Error in this scenario can cause health problems!!)

P-Values

So, how to measure how likely an extreme test statistic value is a ”fluke”??

Definition

Given random sample $\mathbf{X} := (X_1, X_2, \ldots, X_n)$ of a population with parameter θ. Suppose a sample $\mathbf{x} := (x_1, x_2, \ldots, x_n)$ is taken from the population. Finally, let $W(\mathbf{X}; \theta_0)$ be the test statistic for null hypothesis $H_0 : \quad \theta = \theta_0$.

The **P-value** is the probability of obtaining a value of the test statistic at least as contradictory to H_0 as the value computed from the sample, all while assuming that the null hypothesis is true:

$$\begin{array}{l} H_0 : \quad \theta = \theta_0 \\ H_A : \quad \theta > \theta_0 \end{array} \implies \text{P-value} := \mathbb{P}\left(W(\mathbf{X};\theta_0) \geq W(\mathbf{x};\theta_0) \text{ assuming } H_0 \text{ is true}\right)$$

$$\begin{array}{l} H_0 : \quad \theta = \theta_0 \\ H_A : \quad \theta < \theta_0 \end{array} \implies \text{P-value} := \mathbb{P}\left(W(\mathbf{X};\theta_0) \leq W(\mathbf{x};\theta_0) \text{ assuming } H_0 \text{ is true}\right)$$

$$\begin{array}{l} H_0 : \quad \theta = \theta_0 \\ H_A : \quad \theta \neq \theta_0 \end{array} \implies \text{P-value} := \left(\begin{array}{c} \text{Requires distribution type of population} \\ \text{Will be encountered in the rest of Ch8} \end{array} \right)$$

Significance Levels

So, how does one set a ”cutoff” in a hypothesis test??

Definition

Given random sample $\mathbf{X} := (X_1, X_2, \ldots, X_n)$ of a population with parameter θ.
Suppose a sample $\mathbf{x} := (x_1, x_2, \ldots, x_n)$ is taken from the population.
Finally, let $W(\mathbf{X}; \theta_0)$ be the test statistic for null hypothesis $\quad H_0 : \quad \theta = \theta_0$.

Then, a conclusion is reached in a hypothesis test for θ by choosing a **significance level** α that is reasonably close to zero.

If P-value $\leq \alpha$, then H_0 will be rejected in favor of H_A
If P-value $> \alpha$, then H_0 will be accepted (still considered plausible)

The significance levels used in practice are: $\alpha = 0.05, \quad \alpha = 0.01, \quad \alpha = 0.001$

Recall from earlier that $\alpha := \mathbb{P}(\text{Type I Error})$

Also, α is the same α used in $100(1-\alpha)\%$ confidence intervals.

i.e. The lower the α-level, the more skeptical the decision maker is.

P-values & Significance Levels (Example)

Suppose the P-value = 0.02.

Then, for a $(\alpha = 0.05)$-level test, P-value $= 0.02 \leq 0.05 = \alpha \implies$ Reject H_0
But, for a $(\alpha = 0.01)$-level test, P-value $= 0.02 > 0.01 = \alpha \implies$ Accept H_0

For one α-level, H_0 was rejected, but for another α-level, H_0 was accepted??!!

This is a problem, right?????

No, because α must be chosen before the sample is even taken!!
This implies that α must be chosen before a P-value is computed!!

Otherwise, choosing/changing the α-level after the sample is taken & P-value is computed is an example of "moving the goal post!"

"Moving the goal post" like this is an ethical violation in Statistics!!

Hypothesis Tests (Procedure)

Proposition

GIVEN: *Random sample* $\mathbf{X} := (X_1, \ldots, X_n)$ *of a population with parameter* θ.
TASK: *Perform a hypothesis test involving parameter* θ.
(1) State the null hypothesis H_0 *and alternative hypothesis* H_A.

$$\begin{array}{ll} H_0: & \theta = \theta_0 \\ H_A: & \theta > \theta_0 \end{array} \quad \text{OR} \quad \begin{array}{ll} H_0: & \theta = \theta_0 \\ H_A: & \theta < \theta_0 \end{array} \quad \text{OR} \quad \begin{array}{ll} H_0: & \theta = \theta_0 \\ H_A: & \theta \neq \theta_0 \end{array}$$

(2) Select significance level α*:* $\alpha = 0.05$ *OR* $\alpha = 0.01$ *OR* $\alpha = 0.001$
(3) Identify the test statistic $W(\mathbf{X}; \theta_0)$
(4) Formulate a decision rule involving P-value & significance level α
(5) Obtain a sample $\mathbf{x} := (x_1, \ldots, x_n)$ *from the population*
(6) Compute the test statistic value for the sample $W(\mathbf{x}; \theta_0)$
(7) Compute the corresponding P-value
(8) Use the decision rule to arrive at decision to either reject H_0 *or accept* H_0

NOTE: Usually steps 1-5 will be done for you by the problem at hand.

How to Minimize Type I & Type II Errors

Recall the definitions of Type I & Type II Errors:

Definition

A **Type I error** consists of rejecting null hypothesis H_0 when it is true.
A **Type II error** consists of not rejecting null hypothesis H_0 when it is false.

The probability of a Type I error is denoted by α.
The probability of a Type II error is denoted by β.

$$\begin{aligned} \alpha &= \mathbb{P}(\text{Type I Error}) &= \mathbb{P}(\text{Rejecting } H_0 \text{ when } H_0 \text{ is true}) \\ \beta &= \mathbb{P}(\text{Type II Error}) &= \mathbb{P}(\text{Accepting } H_0 \text{ when } H_0 \text{ is false}) \end{aligned}$$

It turns out reducing the chances of a Type I Error (α) will increase the chances of a Type II Error (β)!!

So, how to reduce both error type chances simultaneously???

INCREASE THE SAMPLE SIZE!!

Finally, in practice choose the more serious error to be the Type I Error as that can be easily controlled by choosing a low value for α.

With Hypothesis Tests, ignore Critical Values

Critical values from Confidence Intervals are not used with Hypothesis Tests!

Instead, the **cdf's** of the test statistic will be used to compute P-values!

IGNORE CRITICAL VALUE(S):	INSTEAD USE CDF:
$z^*_{\alpha/2}$	$\Phi(z)$
$t^*_{\nu,\alpha/2}$	$\Phi_t(t;\nu)$
$\chi^{2*}_{\nu,\alpha/2}, \quad \chi^{2*}_{\nu,1-\alpha/2}$	$\Phi_{\chi^2}(x;\nu)$

- Recall that the std normal cdf $\Phi(z)$ table was first seen in section 4.3
- The t cdf $\Phi_t(t;\nu)$ table will be encountered in section 8.3
- The χ^2 cdf will never be encountered in this course.
- The **Binomial cdf** Bi$(x;n,p)$ will be needed for section 8.4
- The Binomial cdf table was first seen in section 3.4

Textbook Logistics for Section 8.1

- Difference(s) in Notation:

CONCEPT	TEXTBOOK NOTATION	SLIDES/OUTLINE NOTATION
Probability of Event	$P(A)$	$\mathbb{P}(A)$
Alternative Hypothesis	H_a	H_A

Fin.