Statistics I – Chapter 9 Hypothesis Testing for One Population (Part 1)

Ling-Chieh Kung

Department of Information Management
National Taiwan University

December 12, 2012

Introduction

- How do **scientists** (physicists, chemists, etc.) do research?
  - Observe phenomena.
  - Make hypotheses.
  - Test the hypotheses through experiments (or other methods).
  - Make conclusions about the hypotheses.
- In the business world, business researchers do the same thing with **hypothesis testing**.
  - One of the most important technique of inferential Statistics.
  - A technique for (statistically) **proving** things.
  - Again relies on **sampling distributions**.

Road map

- **Basic ideas of hypothesis testing**.
- The first example.
- The *p*-value.
- Type I and Type II errors.

People ask questions

- In the business (or social science) world, people ask questions:
  - Are order workers more loyal to a company?
  - Does the newly hired CEO enhance our profitability?
  - Is one candidate preferred by more than 50% voters?
  - Do teenagers eat fast food more often than adults?
  - Is the quality of our products stable enough?
- How should we answer these questions?
- Statisticians suggest:
  - First make a **hypothesis**.
  - Then **test** it with samples and statistical methods.

Hypotheses

- We make hypotheses also because we want to find **explanations** for business phenomena.
  - E.g., suppose we observe one product creates a larger sales volume than another product.
  - We need to know why so that in the future we can make and market popular products.
  - We first guess based on intuitions: "It is because product 1 is cheaper than product 2." Such a guess is a hypothesis.
  - Then we put relevant questions in questionnaires, collect data, analyze data, and then decide whether the hypothesis is true.
- **Guess** by observations or intuitions. **Test** by facts.

Hypotheses

- According to Merriam Webster's Collegiate Dictionary (tenth edition):
  - A **hypothesis** is a tentative explanation of a principle operating in nature.
- So we try to prove hypotheses to find reasons that explain phenomena and enhance decision making.
- There are three types of hypotheses:
  - Research hypotheses.
  - Statistical hypotheses.
  - Substantive hypotheses.

Research hypotheses

- In a **research hypothesis**, the researcher predicts **the outcome** of an experiment of a study.
- It is presented in words with **no specific format**:
  - Older workers are more loyal to a company.
  - The newly hired CEO is useless.
  - This candidate is supported by more than 50% voters.
  - Teenagers eat fast food more often than adults.
  - The quality of our products is not stable.
- To test research hypotheses, we typically state them into statistical hypotheses.

Statistical hypotheses

- A **statistical hypothesis** is a formal way of stating a research hypothesis.
  - Typically with parameters and numbers.
- It contains two parts:
  - The **null hypothesis** (denoted as H_0).
  - The **alternative hypothesis** (denoted as H_a or H_1).
- The alternative hypothesis is:
  - The thing that we want (need) to **prove**.
  - The conclusion that can be made only if we have **a strong evidence**.
- The null hypothesis corresponds to a **default** position.

Statistical hypotheses: example 1

- In our factory, we produce packs of candy whose average weight should be 1 kg.
- One day, a consumer told us that his pack only weighs 900 g.
- We need to know whether this is just a rare event or our production system is out of control.
- If (we believe) the system is out of control, we need to shutdown the machine and spend two days for inspection and maintenance. This will cost us at least $100,000.
- So we should not to believe that our system is out of control just because of one complaint. What should we do?

Statistical hypotheses: example 1

- We may state a research hypothesis "Our production system in under control."
- Then we ask: Is there a strong enough evidence showing that the hypothesis is **wrong**, i.e., the system is out of control?
  - Initially, we assume our system is under control.
  - Then we do a survey for a "strong enough evidence".
  - We should shutdown machines **only if** we prove that the system is out of control.
- Let μ be the average weight, the **statistical hypothesis** is

$$H_0 : \mu = 1$$
$$H_a : \mu \neq 1.$$

Statistical hypotheses: example 1

- Why don't we use

$$H_0\colon \mu \neq 1$$
$$H_a\colon \mu = 1.$$

as the statistical hypothesis?
  - We need a **default position** before we start a survey. $\mu \neq 1$ cannot be a position: We do not know where to stand on.
  - We should shutdown machines only if we have a strong evidence showing that $\mu \neq 1$.
  - The conclusion that requires **a strong evidence** is put in H_a.
- We will have more discussions on how to set up a hypothesis.

Statistical hypotheses

- In the previous example, it does not matter whether the research hypothesis is "our production system in under control" or "our production system in out of control".
- The statistical hypothesis will be the same. We always start by assuming $\mu = 1$, the null hypothesis.
- For beginners in Statistics, one of the most confusing thing is to determine the statements of a statistical hypothesis.
- Let's see some more examples.

Statistical hypotheses: example 2

- In our society, we adopt the presumption of innocence.
  - One is considered **innocent** until proven **guilty**.
- So when there is a person who probably stole some money:

$$H_0: \text{The person is innocent}$$
$$H_a: \text{The person is guilty.}$$

  - It is unacceptable that an innocent person is considered guilty.
  - We will say one is guilty only if there is a strong evidence.

Statistical hypotheses: example 3

- Consider the research hypothesis "The candidate is preferred by more than 50% voters."
- As we need a default position and the percentage that we care about is 50%, we will choose our null hypothesis as

$$H_0 : p = 0.5.$$

- How about the alternative hypothesis? Should it be

$$H_a : p > 0.5 \quad \text{or} \quad H_a : p < 0.5?$$

Statistical hypotheses: example 3

- The choice of the alternative hypothesis depends on the related **decisions** or **actions** to make.
- Suppose one will go for the election only if she thinks she will win (i.e., $p > 0.5$), the alternative hypothesis will be

$$H_a\colon p > 0.5.$$

- Suppose one tends to participate in the election and will give up only if the chance is slim, the alternative hypothesis will be

$$H_a\colon p < 0.5.$$

Remarks

- For setting up a statistical hypothesis:
  - Our default position will be put in the null hypothesis.
  - The thing we want to prove (i.e., the thing that needs a strong evidence) will be put in the alternative hypothesis.
- For writing the mathematical statement:
  - The **equal sign** (=) will always be put in the null hypothesis.
  - The alternative hypothesis contains an **unequal sign** or **strict inequality**: $\neq$, $>$, or $<$.
- The statement of the alternative hypothesis depends on the business context.
- Some studies have H_0, H_1, H_2, ...

One-tailed tests and two-tailed tests

- If the alternative hypothesis contains an unequal sign ($\neq$), the test is a **two-tailed** test.
- If it contains a strict inequality ($>$ or $<$), the test is a **one-tailed** test.
- Suppose we want to test the value of the population mean.
  - In a two-tailed test, we test whether the population mean significantly deviates from a value. We do not care whether it is larger than or smaller than.
  - In a one-tailed test, we test whether the population mean significantly deviates from a value **in a specific direction**.

Substantive hypotheses

- Once we establish a statistical hypothesis, we will do survey and analysis to get conclusions.
- If a strong evidence is found to support the alternative hypothesis, we say the result is (**statistically**) **significant**.
- The concluding statements may be:
  - Old workers are significantly more loyal than young workers.
  - The proportion of voters supporting the candidate is not significantly higher than 50%.
  - Teenagers significantly eat fast food more often than adults.

Substantive hypotheses

- But that one result is statistically significant does not imply it is also **substantively significant**.
  - Suppose the candidate did a survey and get a sample proportion $\hat{p} = 0.505$.
  - If the sample size is large enough, it is possible to conclude that "the proportion of voters supporting him is (statistically) significantly higher than 0.5."
  - But for him, probably 0.505 is still not **high enough**. The statistically significant result is not substantively significant.
- A result is substantive only if it will really affect a decision maker's decision.

Summary

- A research hypothesis states a claim in words.
- A statistical hypothesis states a claim formally.
  - The null hypothesis is our default position.
  - The alternative hypothesis is the thing we want to prove.
- A statistically significant result is substantive only if the decision maker will take actions based on it.

Road map

- Basic ideas of hypothesis testing.
- **The first example**.
- The p-value.
- Type I and Type II errors.

The first example

- Now we will demonstrate the process of hypothesis testing.
- Suppose we test the average weight (in g) of our products.

$$H_0\colon \mu = 1000$$
$$H_a\colon \mu \neq 1000.$$

- Once we have a strong evidence supporting H_a, we will claim that $\mu \neq 1000$.
- Suppose we know the variance of the weights of the products produced: $\sigma^2 = 40000$ g^2.

Controlling the error probability

- Certainly the evidence comes from a **random** sample.
- It is natural that we may be **wrong** when we claim $\mu \neq 1$.
  - E.g., it is possible that $\mu = 1000$ but we unluckily get a sample mean $\bar{x} = 912$.
- We want to **control the error probability**.
  - Let α be the maximum probability for us to make this error.
  - α is called the **significance interval**.
  - So when $\mu = 1$, we will claim that $\mu \neq 1$ for at most probability α.
  - Recall confidence intervals!

Rejection rule

- Now let's test with the significance level $\alpha = 0.05$.
- Intuitively, if $\overline{X}$ **deviates** from 1000 **a lot**, we should **reject** the null hypothesis and believe that $\mu \neq 1000$.
  - If $\mu = 1000$, it is so unlikely to observe such a large deviation.
  - So such a large deviation provides a **strong evidence**.
- So we start by sampling and calculating the **sample mean**.
  - Suppose the sample size $n = 100$.
  - Suppose the sample mean $\bar{x} = 963$.
- We want to construct a **rejection rule**: If $|\overline{X} - 1000| > d$, we reject H_0. We need to calculate d.

Rejection rule

- We want a distance d such that

  if H_0 is true, the probability of rejecting H_0 is 5%.

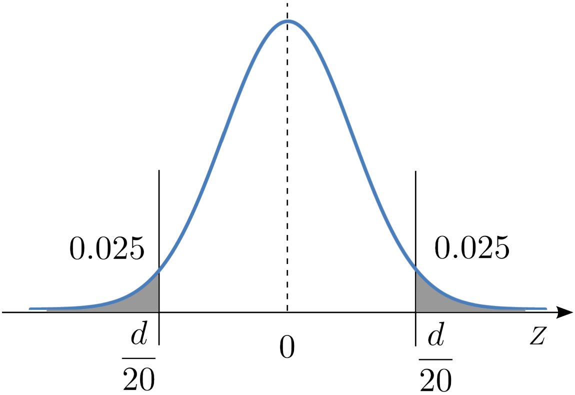
  - If H_0 is true, $\mu = 1000$. We reject H_0 if $|\overline{X} - 1000| > d$.
- Therefore, we need

$$\Pr\left(|\overline{X} - 1000| > d \middle| \mu = 1000\right) = 0.05.$$

  - People typically hide the condition $\mu = 1000$.
- The statistic sample mean $\overline{X}$ has its sampling distribution.
  - Due to the central limit theorem, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \text{ND}(0,1)$. The standard error is $200/\sqrt{100} = 20$.

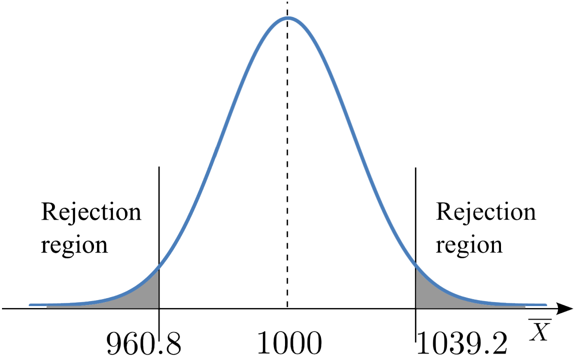
Rejection rule: the critical value

- $0.95 = \Pr(|\overline{X} - 1000| < d) = \Pr(1000 - d < \overline{X} < 1000 + d)$, which is $\Pr(-\frac{d}{20} < Z < \frac{d}{20})$.



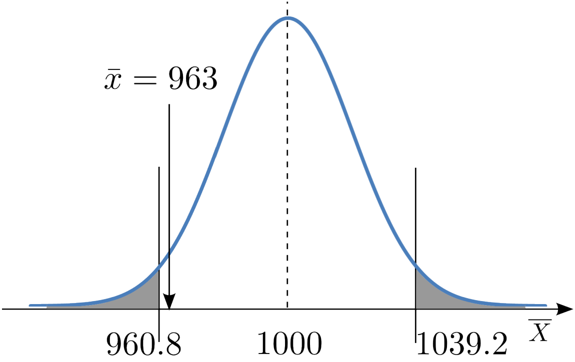
Rejection rule: the critical value

- As $z_{0.025} = 1.96 = \frac{d}{20}$, we have $d = 39.2$.
- The **rejection region** is $R = (-\infty, 960.8) \cup (1039.2, \infty)$.
- If $\overline{X}$ falls in the rejection region, we reject H_0.



Rejection rule: the critical value

- we **cannot reject** H_0 because $\bar{x} = 963 \notin R$.
  - The deviation from 1000 is not large enough.
  - The evidence is not strong enough.



Rejection rule: the critical value

- In this example, the two values 960.8 and 1039.2 are the critical values for rejection.
  - If the sample mean is more extreme than one of the critical values, we reject H_0.
  - Otherwise, we do not reject H_0.
- $\bar{x} = 963$ is not strong enough to support H_a: $\mu \neq 1000$.
- Concluding statement:
  - Because the sample mean does not lie in the rejection region, we **cannot reject H_0**. With a 5% significance level, there **is no** strong evidence showing that the average weight **is not** 1000 g. Based on this result, we **should not** shutdown machines and do an inspection.

Summary

- We want to know whether H_0 is false, i.e., $\mu \neq 1000$.
- We control the probability of making a wrong conclusion.
  - If the machine is actually good, we do not want to reach a conclusion that requires an inspection and maintenance.
  - If H_0 ($\mu = 1000$) is true, we do not want to reject H_0.
  - We limit the probability at the significance level $\alpha = 5\%$.
- We conclude that H_0 is false because the sample mean falls in the rejection region.
  - The calculation of the rejection region (i.e., the critical values) is based on **the z distribution**.
  - We conducted a **z test**.

Not rejecting vs. accepting

- We should be careful in writing our conclusions:
  - **Right:** Because the sample mean does not lie in the rejection region, we **cannot reject** H_0. With a 5% significance level, there **is no** strong evidence showing that the average weight **is not** 1000 g.
  - **Wrong:** Because the sample mean does not lie in the rejection region, we accept H_0. With a 5% significance level, there **is** a strong evidence showing that the average weight **is** 1000 g.
  - Unable to prove one thing is false does not mean it is true!

What probability are we controlling?

- What we have controlled is:
  - If the null hypothesis is true, the probability of rejecting it is no greater than the significance level (α).
- We **did not** ensure that:
  - If we reject the null hypothesis, the probability that the null hypothesis is true is no greater than the significance level (α).
- The key is:
  - **Only if** we know (actually, assume) the null hypothesis is true, we may calculate the probability of rejecting it.
  - The probability cannot be controlled in the opposite way.

What probability are we controlling?

- The significance level α is a **conditional probability**:
  - $\Pr(\text{rejecting } H_0 | H_0 \text{ is true}) = \alpha$.
  - $\Pr(H_0 \text{ is true} | \text{rejecting } H_0)$ cannot be calculated.
- Is the following a correct joint probability table?

	H_0 is true	H_0 is false	Total
Do not reject H_0			
Rejecting H_0	α		
Total			1

The first example (part 2)

- Suppose we modify the hypothesis into a directional one:

$$H_0\colon \mu = 1000.$$
$$H_a\colon \mu < 1000.$$

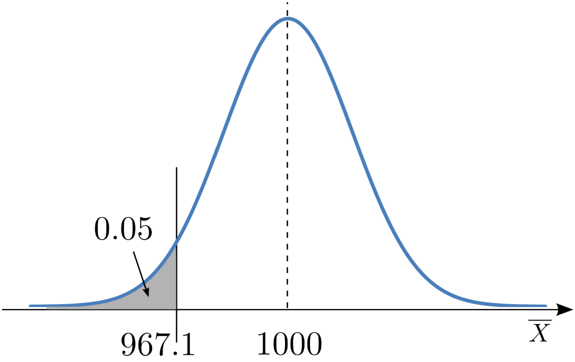
$\sigma^2 = 40000$, $n = 100$, $\alpha = 0.05$.

  - This is a one-tailed test.
  - Once we have a strong evidence supporting H_a, we will claim that $\mu < 1000$.
  - We need to find a distance d such that

$$\Pr\left(1000 - \overline{X} > d \middle| \mu = 1000\right) = 0.05.$$

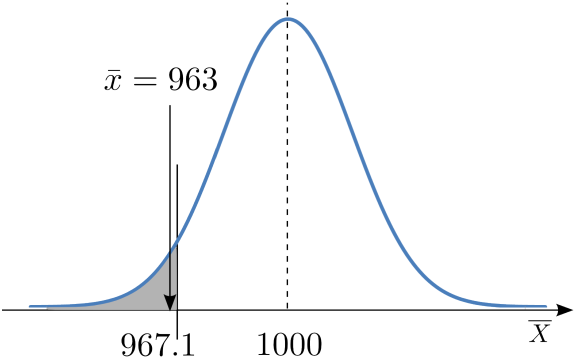
Rejection rule: the critical value

- We have $0.05 = \Pr(1000 - \overline{X} > d) = \Pr(Z < -\frac{d}{20})$.
  - The critical value $z_{0.05} = 1.645$. $d = 1.645 \times 20 = 32.9$.
  - The rejection region is $(-\infty, 967.1)$.



Rejection rule: the critical value

- Because the observed sample mean $\bar{x} = 963 \in (-\infty, 967.1)$, we **reject** H_0.
  - The deviation from 1000 is large enough.
  - The evidence is strong enough.



Rejection rule: the critical value

- In this example, 967.1 is the critical values for rejection.
  - If the sample mean is more extreme than (in this case, below) the critical value, we reject H_0.
  - Otherwise, we do not reject H_0.
- There is a strong evidence supporting H_a: $\mu < 1000$.
- Concluding statement:
  - Because the sample mean lies in the rejection region, we **reject H_0**. With a 5% significance level, there **is a** strong evidence showing that the average weight **is less than** 1000 g.

The other form of the null hypothesis

- Some statisticians write the one-tailed hypothesis as

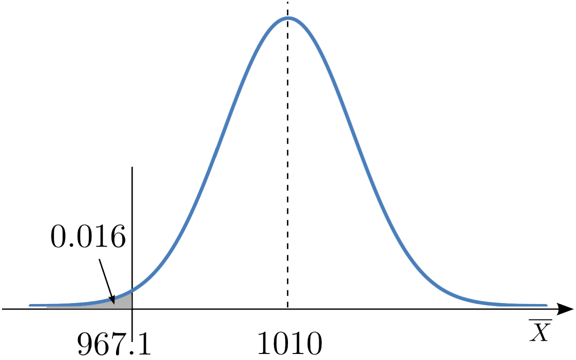
$$H_0\colon \mu \geq 1000$$
$$H_a\colon \mu < 1000.$$

  - When H_0 is true, μ is not fixed to a single value.
- With the rejection region $(-\infty, 967.1)$, what is the error probability Pr(rejecting $H_0|H_0$ is true)?
  - If $\mu = 1000$, Pr(rejecting $H_0|H_0$ is true) $= 0.05$.
  - If $\mu > 1000$,

$$\begin{aligned} &\Pr(\text{rejecting } H_0|H_0 \text{ is true}) \\ =\ &\Pr(\overline{X} < 967.1|H_0 \text{ is true}) < 0.05. \end{aligned}$$

The other form of the null hypothesis

- E.g., suppose $\mu = 1010$.



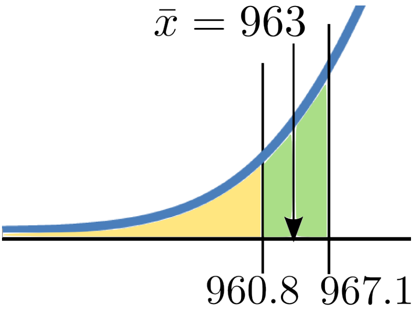
- In general, we control the probability of rejecting H_0 when it is true to be **at most** α.

One-tailed tests vs. two-tailed tests

- When should we use a two-tailed test?
  - We should use a two-tailed test to be **conservative**.
  - E.g., we suspect that the parameter **has changed**, but we are **unsure** whether it becomes larger or smaller.
- If we know or believe that the change is possible **only in one direction**, we may use a one-tailed test.
- If we do not know it, using one-tailed test is **dangerous**.
  - In the previous example with $H_a : \mu < 1000$.
  - If $\bar{x} = 2000$, all we can say is "there is no strong evidence that $\mu < 1000$."
  - We are unable to conclude that $\mu \neq 1000$.

One-tailed tests vs. two-tailed tests

- Having more information (i.e., knowing the direction of change) makes rejection "easier".
- Easier to find a strong enough evidence.



Summary

- Distinguish the following pairs:
  - One- and two-tailed tests.
  - No evidence showing H_0 is false and having evidence showing H_0 is true.
  - Not rejecting H_0 and accepting H_0.
  - Using $=$ and using $\geq$ or $\leq$ in the null hypothesis.

Road map

- Basic ideas of hypothesis testing.
- The first example.
- **The p-value**.
- Type I and Type II errors.

The p-value

- The **p-value** is an important, meaningful, and widely-adopted tool for hypothesis testing.

Definition 1

In a hypothesis testing, for an observed value of the statistic, the p-value is the probability of observing a value that is at least as extreme as the observed value under the assumption the null hypothesis is true.

  - Based on an **observed** value of the statistic.
  - Is the **tail probability** of the observed value.
  - Assuming that the null hypothesis is true.

The p-value

- Mathematically:
  - Suppose we test a population mean μ with a one-tailed test

$$H_0 : \mu = 1000$$
$$H_a : \mu < 1000.$$

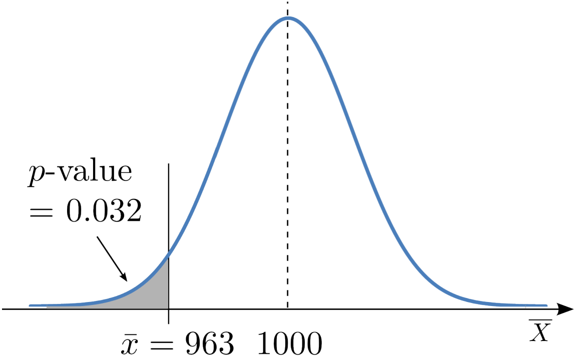
  - Given an observed $\bar{x}$, the p-value is defined as

$$\Pr(\overline{X} < \bar{x}).$$

- In the previous example:
  - $\sigma^2 = 40000$, $n = 100$, $\alpha = 0.05$, $\bar{x} = 963$.
  - How to calculate the p-value of $\bar{x}$?

The p-value

- If H_0 is true, i.e., $\mu = 1000$, we have:
  - $\Pr(\overline{X} \leq 963) = \Pr(Z \leq -1.85) = 0.032.$



What factors affect the p-value?

- Which of the following factors affect the p-value

$$\Pr(\overline{X} < \bar{x})?$$

  - The observed value of the statistic.
  - The population mean assumed in the null hypothesis.
  - The population variance.
  - The sample size.
  - The significance level α.
  - Whether the test is one-tailed or two-tailed.

How to use the p-value?

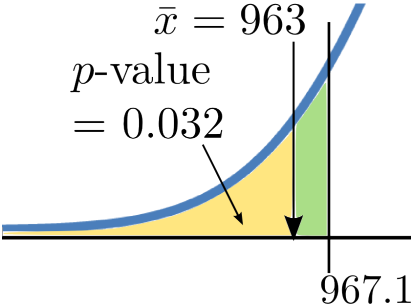
- The p-value can be used for constructing a **rejection rule**.
- For a one-tailed test:
  - If the p-value is **smaller** than α, we **reject** H_0.
  - If the p-value is greater than α, we do not reject H_0.
- Consider the one-tailed test

$$H_0 : \mu = 1000$$
$$H_a : \mu < 1000.$$

  - Suppose we still adopt $\alpha = 0.05$.
  - Because the p-value $0.032 < 0.05$, we reject H_0.

p-values vs. critical values

- Using the p-value is **equivalent** to using the critical values.
  - The rejection-or-not decision we make will be the same based on the two methods.



The benefit of using the p-value

- In calculating the p-value, we do not need α.
- After the p-value is calculated, we compare it with α.
- The p-value, which needs to be calculated **only once**, allows us to know whether the evidence is strong enough under various significance levels.

α	0.1	0.05	0.01
Rejecting H_0?	Yes ($0.032 < 0.1$)	Yes ($0.032 < 0.05$)	No ($0.032 > 0.01$)

  - If we use the critical-value method, we need to calculate the critical value for three times, one for each value of α.

The benefit of using the p-value

- In many studies, the researchers do not determine the significance level α before a test is conducted.
- They calculate the p-value and then mark **how significant** the result is with **stars**.

p-value	< 0.01	< 0.05	< 0.1	> 0.1
Significant?	Highly significant	Moderately significant	Slightly significant	Insignificant
Mark	***	**	*	(Empty)

The benefit of using the p-value

- As an example, suppose one is testing whether people sleep at least eight hours per day in average.
  - Age groups: $[10, 15)$, $[15, 20)$, $[20, 35)$, etc.
  - For group i, a one-tailed test is conducted. $H_a : \mu_i > 8$.
  - The result may be presented in a table:

Group	Age group	p-value
1	[10,15)	0.002***
2	[15,20)	0.2
3	[20,25)	0.06*
4	[25,30)	0.04**
5	[30,35)	0.03**

Interpreting the p-value

- A smaller p-value does NOT mean **a larger deviation**!
  - We cannot conclude that $\mu_5 > \mu_4$, $\mu_1 > \mu_3$, etc.
- A smaller p-value means **a higher probability** to reject the null hypothesis.
  - If $\alpha = 0.01$, we will conclude that only μ_1 is statistically significantly larger than 8.
  - We do not believe that μ_1 is larger than 8 **by a huge amount**!
  - It is **more probable** (i.e., with a larger range of α) for us to conclude that μ_1 "significantly" deviate from 8.

The p-value for two-tailed tests

- How to construct the rejection rule for a **two-tailed** test?
  - If the p-value is **smaller** than $\frac{\alpha}{2}$, we **reject** H_0.
  - If the p-value is greater than $\frac{\alpha}{2}$, we do not reject H_0.
- Consider the two-tailed test

$$H_0 : \mu = 1000.$$
$$H_a : \mu \neq 1000.$$

  - Suppose we still adopt $\alpha = 0.05$.
  - Because the p-value $0.032 > \frac{\alpha}{2} = 0.025$, we do not reject H_0.

The p-value for two-tailed tests

- In most commercial statistical software, there are functions that help one calculate p-values.
- Some functions return the p-value for a one-tailed test but **twice of the p-value** for a two-tailed test.
  - E.g., the function TTEST() in MS Excel.
- With these functions, we will always compare the returned value with α directly.
- Read the instructions before using those functions!

Summary

- The p-value is the tail probability of the realization of a statistics assuming the null hypothesis is true.
- The p-value method is an alternative way of making the rejection decision.
  - It is equivalent to the critical-value method.
- The p-value measure how probable to reject H_0.
- It does not measure how larger the deviation is.

Road map

- Basic ideas of hypothesis testing.
- The first example.
- The p-value.
- **Type I and Type II errors**.

Type I error

- We discussed a lot in controlling a probability:
  - If the null hypothesis is true, we want to avoid rejecting it.
  - Typically we set $\Pr(\text{rejecting } H_0 | H_0 \text{ is true}) = \alpha$.
  - In general, it is $\Pr(\text{rejecting } H_0 | H_0 \text{ is true}) \leq \alpha$.
  - What we have controlled is not $\Pr(H_0 \text{ is true} | \text{rejecting } H_0)$.
- If we reject a true null hypothesis, we make a **Type I error**.
- What if the null hypothesis is false?

Type II error

- What if the null hypothesis is false? How to avoid not rejecting a false null hypothesis?
- Not rejecting a false null hypothesis is a **Type II error**.
- The probability of making a type II error is denoted as β:

$$\Pr(\text{rejecting } H_0 | H_0 \text{ is true}) = \alpha.$$
$$\Pr(\text{not rejecting } H_0 | H_0 \text{ is false}) = \beta.$$

- We controlled the probability of making a Type I error. We know it is at most α.
- Do we know the probability of making a Type II error?

Type II error

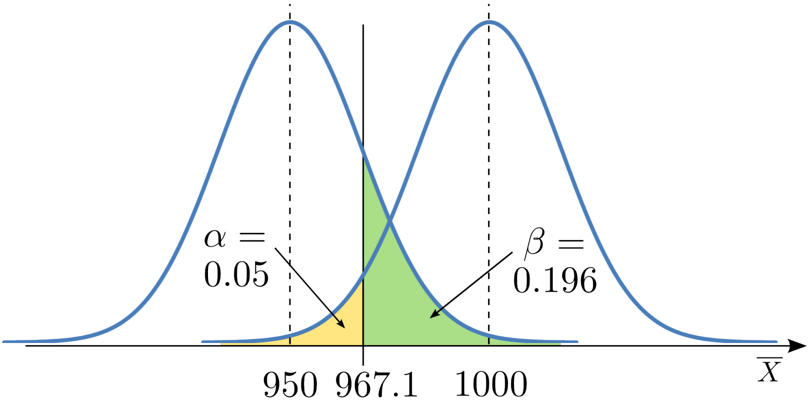
- Recall our one-tailed test with $\alpha = 0.05$ again:

$$H_0 : \mu = 1000.$$
$$H_a : \mu < 1000.$$

- If H_0 is false and μ is actually 950, we know how to calculate β:
  - The **rejection rule** (which is constructed by assuming H_0 is true) will be the same: Reject H_0 if $\overline{X} < 967.1$.
  - The probability of **not rejecting H_0** is

$$\Pr(\overline{X} > 967.1) = \Pr(Z > 0.855) = 0.196 = \beta.$$

α and β



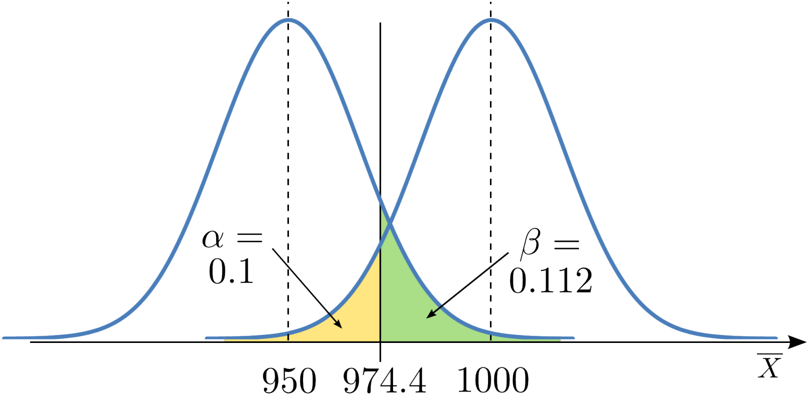
Type II error

- For every different value of μ, we have a different β:

μ	950	960	970	980	990
β	0.196	0.361	0.558	0.74	0.874

- As the true value of μ is never known, we **never** know β.
- To lower β, one way is to increase α.

Increasing α to decrease β



Type I errors vs. Type II errors

- If we control α, we **cannot** control β.
- As α is controlled, β (as a function of the parameter) determines **how good a test is**.
  - $1 - \beta$ is called the **power** of a test. Smaller β means a better test.
- Summary:

Action	State on nature	
	H_0 is true	H_0 is false
Do not reject H_0	Correct decision ($1 - \alpha$)	Type II error (β)
Reject H_0	Type I error (significance level: α)	Correct decision (power: $1 - \beta$)

Why controlling α only?

- We cannot control α and β at the same time.
- Why do we control α only?
- Recall what we did in setting up a hypothesis:
  - We put the claim that requires a strong evidence in H_a.
  - We will conclude that H_a is true only with a strong evidence.
- We did so because it is more important to:
  - Avoid rejecting H_0 when it is true.
  - Avoid a type I error.
- That is, a type I error is **more costly** than a type II error.
  - This is why controlling α is our first priority.

Setting up a hypothesis

- As a judge, which one will you choose?
  - H_0: Innocent. H_a: Guilty.
  - H_0: Guilty. H_a: Innocent.
- As a manufacturer, which one will you choose?
  - μ is the weight of a bag of candy. Ideally it should be 1000.
  - H_0: $\mu = 1000$. H_a: $\mu < 1000$.
  - H_0: $\mu = 1000$. H_a: $\mu > 1000$.
- What if we conduct a two-tailed test?
  - H_0: $\mu = 1000$. H_a: $\mu \neq 1000$.
  - H_0: $\mu \neq 1000$. H_a: $\mu = 1000$. (Can we?)
  - But we may adjust α.

Summary

- Type I errors and Type II errors.
  - Type I: Rejecting a true H_0.
  - Type II: Not rejecting a false H_0.
- We control α, the probability of making a Type I error.
- We do not (cannot) control β directly.
- To reduce both α and β, increase the sample size.