Probability and Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

April 30, 2019

Topics

Objectives

Objectives:

- State the null and alternative hypotheses for a given problem.
- State the role of the level of significance in hypothesis testing, and describe how the choice of a significance level can affect the conclusion of a hypothesis test.
- Interpret the p-value of a hypothesis test.
- Carry out a one-sample z test for a normal mean μ when the population standard deviation σ is known.

Introduction to Hypothesis Testing (8.1)

- A statistical ***hypothesis*** is a **claim** about the value of one or more population parameters.

Introduction to Hypothesis Testing (8.1)

- A statistical ***hypothesis*** is a **claim** about the value of one or more population parameters.
- A ***hypothesis test*** is a statistical procedure for deciding between two competing hypotheses.

Introduction to Hypothesis Testing (8.1)

- A statistical ***hypothesis*** is a **claim** about the value of one or more population parameters.
- A ***hypothesis test*** is a statistical procedure for deciding between two competing hypotheses.
- The ***null hypothesis*** (H_0) is the hypothesis we seek to **discredit**, but to which we give the **benefit of the doubt.**

Introduction to Hypothesis Testing (8.1)

- A statistical ***hypothesis*** is a **claim** about the value of one or more population parameters.
- A ***hypothesis test*** is a statistical procedure for deciding between two competing hypotheses.
- The ***null hypothesis*** (H_0) is the hypothesis we seek to **discredit**, but to which we give the **benefit of the doubt.**
- The ***alternative hypothesis*** (H_a) is the hypothesis we seek to **substantiate**.

- The conclusion of any hypothesis test will be to either **reject** or **fail to reject** H_0.

- The conclusion of any hypothesis test will be to either **reject** or **fail to reject** H_0.
- The decision will be based on a ***test statistic***, its associated ***p-value***, and a ***decision rule*** involving a ***level of significance***.

Steps for Carrying Out a Hypothesis Test:

1. Identify the parameter of interest.
2. Determine the null value and state H_0 and H_a.
3. Write the test statistic formula and compute its value.
4. Determine the p-value.
5. Use the decision rule, level of significance, and p-value to decide whether H_0 should be rejected, and state the conclusion.

- The ***p-value*** is a **probability** that answers the question:

  *"**If $\mathbf{H_0}$ was true**, what's the chance we'd get a test statistic value that's as contradictory to H_0 (and consistent with H_a) as the one we got?"*

- The ***p-value*** is a **probability** that answers the question:

  *"**If $\boldsymbol{H_0}$ was true**, what's the chance we'd get a test statistic value that's as contradictory to H_0 (and consistent with H_a) as the one we got?"*

  Smaller p-values indicate **stronger evidence against $\boldsymbol{H_0}$** in favor of H_a.

- The ***p-value*** is a **probability** that answers the question:

  *"**If $\boldsymbol{H_0}$ was true**, what's the chance we'd get a test statistic value that's as contradictory to H_0 (and consistent with H_a) as the one we got?"*

  Smaller p-values indicate **stronger evidence against $\boldsymbol{H_0}$** in favor of H_a.

- The ***level of significance***, denoted α, is a threshold used in the ***decision rule***, which states:

Decision Rule:

Reject H_0 if p-value $< \alpha$.
Fail to reject H_0 if p-value $\geq \alpha$.

- The most commonly used values for α are **0.01**, **0.05**, and **0.10**.

Decision Rule:

Reject H_0 if p-value $< \alpha$.
Fail to reject H_0 if p-value $\geq \alpha$.

- The most commonly used values for α are **0.01**, **0.05**, and **0.10**.
- Using a **smaller** value for α means we require **stronger evidence** against H_0 before we're willing to reject H_0.

- When we reject H_0, we say the result is ***statistically significant***, and conclude that the result is **not just due to chance.**

Test for a Normal Mean μ When σ is Known (8.2)

- Suppose
  1. $X_1, X_2, \ldots, X_n$ are a random sample from a **normal** population.
  2. The population mean μ **is unknown** but the standard deviation σ **is known.**

Test for a Normal Mean μ When σ is Known (8.2)

- Suppose
  1. $X_1, X_2, \ldots, X_n$ are a random sample from a **normal** population.
  2. The population mean μ **is unknown** but the standard deviation σ **is known.**

Test for a Normal Mean μ When σ is Known (8.2)

- Suppose
  1. $X_1, X_2, \ldots, X_n$ are a random sample from a **normal** population.
  2. The population mean μ **is unknown** but the standard deviation σ **is known.**

- We'll see how to use the sample to decide if μ is different from some **hypothesized value** $\boldsymbol{\mu_0}$.

Test for a Normal Mean μ When σ is Known (8.2)

- Suppose
  1. $X_1, X_2, \ldots, X_n$ are a random sample from a **normal** population.
  2. The population mean μ **is unknown** but the standard deviation σ **is known.**

- We'll see how to use the sample to decide if μ is different from some **hypothesized value** μ_0.

  The appropriate hypothesis test (when σ is known) is called the ***one-sample*** z ***test for*** μ.

- Because we're seeking to "disprove" the claim that μ is equal to μ_0, the **null hypothesis** is that it *is* equal to μ_0 .

Null Hypothesis:

$$H_0 : \ \mu \ = \ \mu_0$$

- The **alternative hypothesis** will depend on what we're trying to "prove":

Alternative Hypothesis: The alternative hypothesis will be one of

1. $H_a : \mu > \mu_0$ (**one-sided, upper-tailed**)
2. $H_a : \mu < \mu_0$ (**one-sided, lower-tailed**)
3. $H_a : \mu \neq \mu_0$ (**two-sided, two-tailed**)

depending on what we're trying to verify using the data.

Example

Hemoglobin is a protein in red blood cells that carries oxygen from the lungs to body tissues. People with less than **12** grams of hemoglobin per deciliter of blood (g/dl) are said to be anemic.

Example

Hemoglobin is a protein in red blood cells that carries oxygen from the lungs to body tissues. People with less than **12** grams of hemoglobin per deciliter of blood (g/dl) are said to be anemic.

A health official in Jordan suspects that the mean μ for all children in that country is **less than 12.** To test his claim, he measures the hemoglobin a random sample of $n = 50$ children.

Example

Hemoglobin is a protein in red blood cells that carries oxygen from the lungs to body tissues. People with less than **12** grams of hemoglobin per deciliter of blood (g/dl) are said to be anemic.

A health official in Jordan suspects that the mean μ for all children in that country is **less than 12**. To test his claim, he measures the hemoglobin a random sample of $n = 50$ children.

The **null hypothesis** is

$$H_0 : \ \mu \ = \ 12$$

Which of the following three **alternative hypotheses** would he test?

- $H_a : \mu > 12$
- $H_a : \mu < 12$
- $H_a : \mu \neq 12$

Which of the following three **alternative hypotheses** would he test?

- $H_a : \mu > 12$
- $H_a : \mu < 12$
- $H_a : \mu \neq 12$

Would he be performing a **lower-tailed**, **upper-tailed**, or **two-tailed test**?

Example

The Survey of Study Habits and Attitudes (SSHA) is a psychological test that measures students' study habits and attitudes toward school. The mean SSHA score for all college students is **115**.

Example

The Survey of Study Habits and Attitudes (SSHA) is a psychological test that measures students' study habits and attitudes toward school. The mean SSHA score for all college students is **115**.

A teacher suspects that the true mean μ for *older* students is **higher than 115.** To test her claim, she gives the SSHA to a random sample of $n = 25$ students who are at least 30 years old.

Example

The Survey of Study Habits and Attitudes (SSHA) is a psychological test that measures students' study habits and attitudes toward school. The mean SSHA score for all college students is **115**.

A teacher suspects that the true mean μ for *older* students is **higher than 115.** To test her claim, she gives the SSHA to a random sample of $n = 25$ students who are at least 30 years old.

The **null hypothesis** is

$$H_0 : \ \mu \ = \ 115$$

Which of the following three **alternative hypotheses** would she test?

- $H_a : \mu > 115$
- $H_a : \mu < 115$
- $H_a : \mu \neq 115$

Which of the following three **alternative hypotheses** would she test?

- $H_a : \mu > 115$
- $H_a : \mu < 115$
- $H_a : \mu \neq 115$

Would she be performing a **lower-tailed**, **upper-tailed**, or **two-tailed test**?

Example

The diameter of a spindle in a small motor is supposed to be **5** mm. If the spindle is either too small or too large, the motor will not work properly.

The manufacturer measures the diameters in a random sample of $n = 10$ spindles to determine whether the true mean diameter μ is **any different** from the target value **5** mm.

Example

The diameter of a spindle in a small motor is supposed to be **5** mm. If the spindle is either too small or too large, the motor will not work properly.

The manufacturer measures the diameters in a random sample of $n = 10$ spindles to determine whether the true mean diameter μ is **any different** from the target value **5** mm.

The **null hypothesis** is

$$H_0 : \ \mu \ = \ 5$$

Which of the following three **alternative hypotheses** would they test?

- $H_a: \mu > 5$
- $H_a: \mu < 5$
- $H_a: \mu \neq 5$

Which of the following three **alternative hypotheses** would they test?

- $H_a : \ \mu \ > \ 5$
- $H_a : \ \mu \ < \ 5$
- $H_a : \ \mu \ \neq \ 5$

Would they be performing a **lower-tailed**, **upper-tailed**, or **two-tailed test**?

- The **test statistic** for the ***one-sample z test for μ*** is

One-Sample Z Test Statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} .$$

When H_0 is true, $Z \sim$ N$(0, 1)$.

- The **test statistic** for the ***one-sample z test for μ*** is

One-Sample Z Test Statistic:

$$Z \;=\; \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\,.$$

When H_0 is true, $Z \;\sim\;$ N$(0, 1)$.

- Z measures how many standard errors $\bar{X}$ is away from $\boldsymbol{\mu_0}$.

- The **test statistic** for the ***one-sample z test for μ*** is

One-Sample Z Test Statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

When H_0 is true, $Z \sim \mathrm{N}(0,1)$.

- Z measures how many standard errors $\bar{X}$ is away from μ_0.
- $\bar{X}$ is an estimator of the unknown population mean μ, so ...

- The **test statistic** for the ***one-sample z test for μ*** is

One-Sample Z Test Statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

When H_0 is true, $Z \sim \mathrm{N}(0, 1)$.

- Z measures how many standard errors $\bar{X}$ is away from $\boldsymbol{\mu_0}$.
- $\bar{X}$ is an estimator of the unknown population mean $\boldsymbol{\mu}$, so ...
  1. Z will be approximately **zero** if $\mu = \mu_0$.
  2. It will be **positive** if $\mu > \mu_0$.
  3. It will be **negative** if $\mu < \mu_0$.

1. **Large positive** values of Z provide **evidence against H_0 in favor of**
   $\boldsymbol{H_a : \mu > \mu_0}$.
2. **Large negative** values of Z provide **evidence against H_0 in favor of**
   $\boldsymbol{H_a : \mu < \mu_0}$.
3. **Large positive *and* large negative** values of Z provide **evidence**
   against H_0 in favor of
   $\boldsymbol{H_a : \mu \neq \mu_0}$.

- Recall that

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathsf{N}(0,1).$$

- Recall that

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathsf{N}(0, 1).$$

- It follows that **if H_0 is true** (so $\mu = \mu_0$),

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathsf{N}(0, 1).$$

- Recall that
$$\frac{\bar{X} - \boldsymbol{\mu}}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1).$$
- It follows that **if $\boldsymbol{H_0}$ is true** (so $\mu = \mu_0$),
$$\frac{\bar{X} - \boldsymbol{\mu_0}}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1).$$

Sampling Distribution of the Test Statistic Under H_0:

If Z is the one-sample Z test statistic, then when

$$H_0: \ \mu \ = \ \mu_0$$

is true,

$$Z \sim \mathsf{N}(0, 1).$$

Sampling Distribution of the Test Statistic Under H_0:
If Z is the one-sample Z test statistic, then when

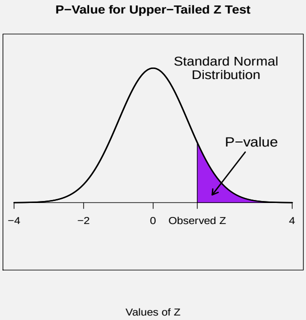
$$H_0: \ \mu \ = \ \mu_0$$

is true,

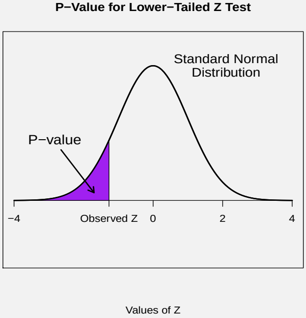
$$Z \sim \mathsf{N}(0,1).$$

- The ***p-value*** is the probability that just by chance (under H_0) we'd get a test statistic value as far from zero, in the direction predicted by H_a, as the observed value.

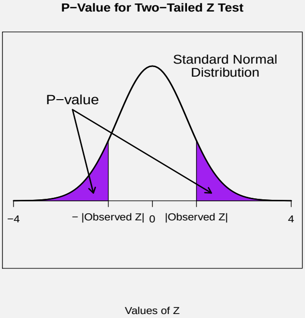
1. **P-value** = Area to the **right** of the observed z if the alternative hypothesis is $H_a : \mu > \mu_0$.



1. **P-value** = Area to the **left** of the observed z if the alternative hypothesis is $H_a : \mu < \mu_0$.



1. **P-value** = Area to the **left** of $-|z|$ **and right** of $|z|$ if the alternative hypothesis is $H_a : \mu \neq \mu_0$.



One-Sample Z Test for μ when σ is Known

Assumptions: The data $x_1, x_2, \ldots, x_n$ are a random sample from a N(μ, σ) distribution where σ is known.

Null hypothesis: $H_0 : \mu = \mu_0$.

Test statistic value: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$.

Decision rule: Reject H_0 if p-value $< \alpha$.

Alternative hypothesis	**P-value** = area under N$(0, 1)$ distribution:
$H_a : \mu > \mu_0$	to the right of z
$H_a : \mu < \mu_0$	to the left of z
$H_a : \mu \neq \mu_0$	to the left of $-\lvert z \rvert$ and right of $\lvert z \rvert$

Example

Recall that a health official in Jordan suspects that the mean hemoglobin level μ for all children in that country is **less than 12.** To test his claim, he measures the hemoglobin a random sample of $n = 50$ children.

Example

Recall that a health official in Jordan suspects that the mean hemoglobin level μ for all children in that country is **less than 12.** To test his claim, he measures the hemoglobin a random sample of $n = 50$ children.

He'll test the **hypotheses**

$$\begin{aligned} H_0 : \mu &= 12 \\ H_a : \mu &< 12 \end{aligned}$$

In the **sample**, the mean hemoglobin level is

$$\bar{x} = 11.7.$$

In the **sample**, the mean hemoglobin level is

$$\bar{x} = 11.7.$$

Suppose the population standard deviation is **known** to be $\sigma = 2$.

In the **sample**, the mean hemoglobin level is

$$\bar{x} = 11.7.$$

Suppose the population standard deviation is **known** to be $\sigma = 2.$

Then the observed **test statistic** is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

In the **sample**, the mean hemoglobin level is

$$\bar{x} = 11.7.$$

Suppose the population standard deviation is **known** to be $\sigma = 2$.

Then the observed **test statistic** is

$$\begin{aligned} z &= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\ &= \frac{11.7 - 12}{2/\sqrt{50}} \end{aligned}$$

In the **sample**, the mean hemoglobin level is

$$\bar{x} = 11.7.$$

Suppose the population standard deviation is **known** to be $\sigma = 2$.

Then the observed **test statistic** is

$$\begin{aligned} z &= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\ &= \frac{11.7 - 12}{2/\sqrt{50}} \\ &= -1.06. \end{aligned}$$

Thus the **sample mean** hemoglobin level, $\bar{x} = 11.7$, is **1.06 standard errors below 12**.

Thus the **sample mean** hemoglobin level, $\bar{x} = 11.7$, is **1.06 standard errors below 12**.

The **p-value** is the **probability** that we'd get this result by chance **if** the **population mean** μ was **12.**

Thus the **sample mean** hemoglobin level, $\bar{x} = 11.7$, is **1.06 standard errors below 12**.

The **p-value** is the **probability** that we'd get this result by chance **if** the **population mean** μ was **12.**

From the **left tail** of the **sampling distribution** that the test statistic would follow under H_0 (the N$(0, 1)$ distribution), the **p-value** is **0.1446**.

Thus we'd get a result like the one we got **14.46%** of the time **even if** the **population mean** μ was **12**.

Thus we'd get a result like the one we got **14.46%** of the time **even if** the **population mean** μ was **12**.

Using a **level of significance** $\alpha = \mathbf{0.05}$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Thus we'd get a result like the one we got **14.46%** of the time **even if** the **population mean** μ was **12**.

Using a **level of significance** $\alpha = 0.05$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Because $0.1446 \geq 0.05$, we **fail to reject** H_0.

Thus we'd get a result like the one we got **14.46%** of the time **even if** the **population mean** μ was **12**.

Using a **level of significance** $\boldsymbol{\alpha = 0.05}$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Because $\mathbf{0.1446 \geq 0.05}$, we **fail to reject** $\boldsymbol{H_0}$.

There's **no statistically significant evidence** that the population mean hemoglobin level μ is less than 12.

Thus we'd get a result like the one we got **14.46%** of the time **even if** the **population mean** μ was **12**.

Using a **level of significance** $\alpha = 0.05$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Because $\mathbf{0.1446 \geq 0.05}$, we **fail to reject** $\boldsymbol{H_0}$.

There's **no statistically significant evidence** that the population mean hemoglobin level μ is less than 12.

The result he got (by taking a random sample) can be explained by chance variation (sampling error).

Example

A concern of employers is time spent surfing the Internet and emailing friends during work hours.

Example

A concern of employers is time spent surfing the Internet and emailing friends during work hours.

An article in the *San Luis Obispo Tribune* (Aug. 3, 2016) ran under the headline "Who Goofs Off 2 Hours a Day? Most Workers, Survey Says".

Example

A concern of employers is time spent surfing the Internet and emailing friends during work hours.

An article in the *San Luis Obispo Tribune* (Aug. 3, 2016) ran under the headline ”Who Goofs Off 2 Hours a Day? Most Workers, Survey Says”.

The CEO of a company wants to decide whether the average amount of wasted time per work day for her employees is **less than** the reported **120** minutes.

Example

A concern of employers is time spent surfing the Internet and emailing friends during work hours.

An article in the *San Luis Obispo Tribune* (Aug. 3, 2016) ran under the headline "Who Goofs Off 2 Hours a Day? Most Workers, Survey Says".

The CEO of a company wants to decide whether the average amount of wasted time per work day for her employees is **less than** the reported **120** minutes.

A random sample of $n = 10$ employees was asked about daily wasted time at work.

Example

A concern of employers is time spent surfing the Internet and emailing friends during work hours.

An article in the *San Luis Obispo Tribune* (Aug. 3, 2016) ran under the headline "Who Goofs Off 2 Hours a Day? Most Workers, Survey Says".

The CEO of a company wants to decide whether the average amount of wasted time per work day for her employees is **less than** the reported **120** minutes.

A random sample of $n = 10$ employees was asked about daily wasted time at work.

(They were guaranteed anonymity to obtain truthful answers!)

She'll test the **hypotheses**

$$\begin{aligned} H_0 : \ \mu &= 120 \\ H_a : \ \mu &< 120 \end{aligned}$$

She'll test the **hypotheses**

$$\begin{aligned} H_0 : \mu &= 120 \\ H_a : \mu &< 120 \end{aligned}$$

Here are the data:

108	112	117	122	111	131	113	113	105	128

The sample mean is

$$\bar{x} = \mathbf{116.0}.$$

Suppose we know that in the company's employee population, wasted time follows a **normal** distribution with $\sigma = 9.5$.

Suppose we know that in the company's employee population, wasted time follows a **normal** distribution with $\boldsymbol{\sigma = 9.5}$.

Do the data provide **statistically significant** evidence, at the $\boldsymbol{\alpha = 0.05}$ **level**, that the population mean wasted time μ for this company is **less than 120** minutes?

The observed **test statistic** is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

The observed **test statistic** is

$$\begin{aligned} z &= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \\ &= \frac{116 - 120}{9.5/\sqrt{10}} \end{aligned}$$

The observed **test statistic** is

$$\begin{aligned} z &= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \\ &= \frac{116-120}{9.5/\sqrt{10}} \\ &= \mathbf{-1.33}. \end{aligned}$$

Thus the **sample mean** wasted time, $\bar{x} = 116$, is **1.33 standard errors below 120**.

The observed **test statistic** is

$$\begin{aligned} z &= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \\ &= \frac{116-120}{9.5/\sqrt{10}} \\ &= \mathbf{-1.33}. \end{aligned}$$

Thus the **sample mean** wasted time, $\bar{x} = 116$, is **1.33 standard errors below 120**.

The **p-value** is the **probability** that we'd get this result by chance **if** the **population mean** μ was **120.**

The **p-value** is the **probability** that we'd get this result by chance **if** the **population mean** μ was **120**.

From the **left tail** of the **sampling distribution** that the test statistic would follow under H_0 (the N$(0, 1)$ distribution), the **p-value** is **0.0918**.

Thus we'd get a result like the one we got **9.18%** of the time **even if** the **population mean** μ was **120**.

Thus we'd get a result like the one we got **9.18%** of the time **even if** the **population mean** μ was **120**.

Using a **level of significance** $\alpha = 0.05$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Thus we'd get a result like the one we got **9.18%** of the time **even if** the **population mean** μ was **120**.

Using a **level of significance** $\boldsymbol{\alpha = 0.05}$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Because $\mathbf{0.0918 \geq 0.05}$, we **fail to reject** $\boldsymbol{H_0}$.

Thus we'd get a result like the one we got **9.18%** of the time **even if** the **population mean** μ was **120**.

Using a **level of significance** $\boldsymbol{\alpha = 0.05}$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Because $\mathbf{0.0918 \geq 0.05}$, we **fail to reject** $\boldsymbol{H_0}$.

There's **no statistically significant evidence** that the population mean wasted time μ is less than 120 minutes.

Thus we'd get a result like the one we got **9.18%** of the time **even if** the **population mean** μ was **120**.

Using a **level of significance** $\alpha = 0.05$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Because $0.0918 \geq 0.05$, we **fail to reject** H_0.

There's **no statistically significant evidence** that the population mean wasted time μ is less than 120 minutes.

The result she got (by taking a random sample) can be explained by chance variation (sampling error).

If instead she had used a **level of significance** $\alpha = 0.10$, would the conclusion have been different?