BUDT 730
Data, Models and Decisions

Lecture 7

Hypothesis Testing (I)

Prof. Sujin Kim

Learning Objective

- Learn the principles of hypothesis testing
- Construct the null and alternative hypotheses for business cases
- Distinguish between two different types of errors: Type I and Type II
- Learn how to interpret the two errors
- Perform a hypothesis test for the population mean

- Data Files:
  - 2003salary.xlsx
  - AluminumSheet.xlsx
  - Beverage Bottling.xlsx

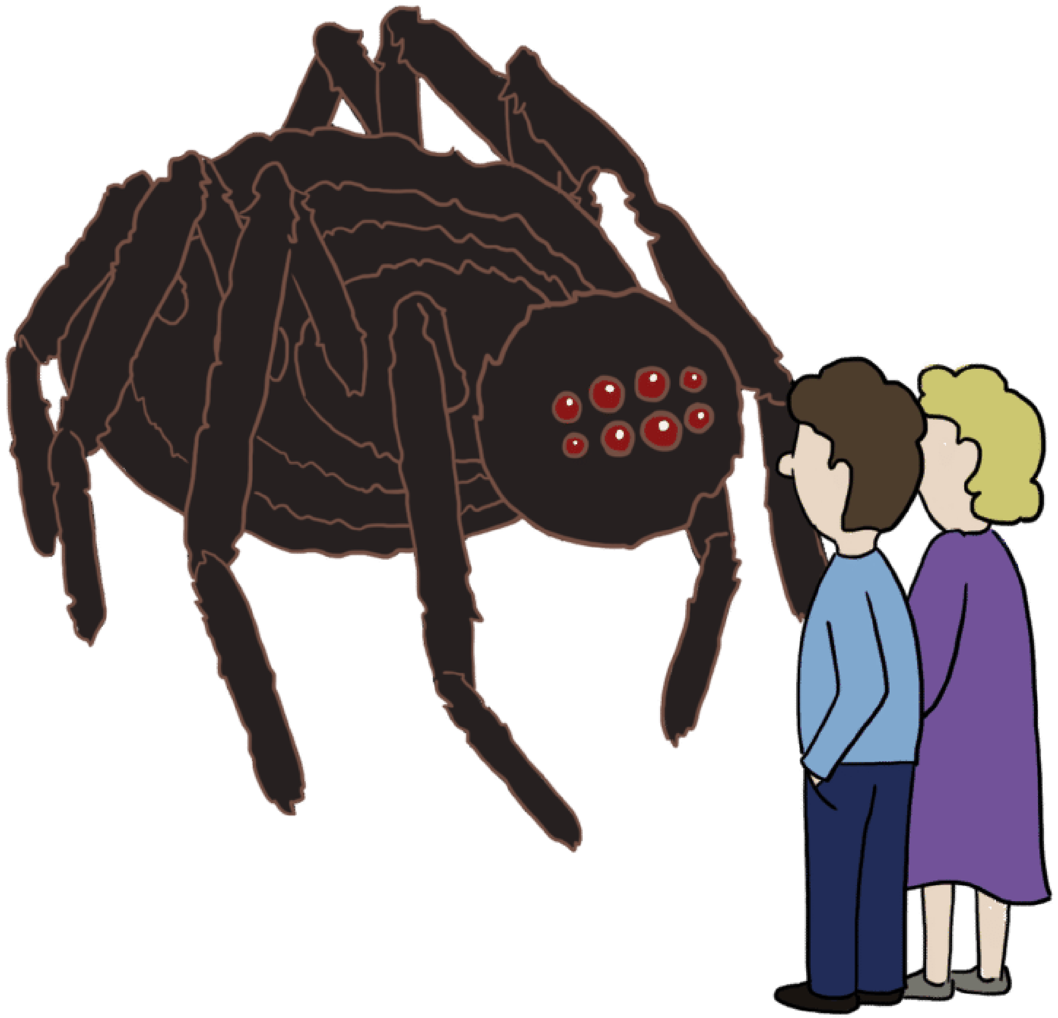
Examples and R Functions

- Four examples:
  - Example 1 - CEO Salary, Data set: 2003Salary.xlsx
  - Example 2 - Aluminum Sheet, Data set: AluminumSheet.xlsx
  - Example 3 – Meal Service problem, No data
  - Example4 – Bottling filling problem, Data set: Beverage Bottling.xlsx

- R libraries and functions for the examples:

Package Name	Function Name
built-in package	pnorm(), pt()
built-in package	qnorm(), qt()
built-in package	t.test()

Concepts of Hypothesis Testing



"I've narrowed it down to two hypothesis: it grew, or we shrunk."

Credit: Dr. Pieter Tans, NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/) and Dr. Ralph Keeling, Scripps Institution of Oceanography (scrippsco2.ucsd.edu/)

Recall Meal Service Problem

- A government contractor provided services to the military in a troubled region.
  - Average of 10,000 daily meals provided.
  - Operations lasted 300 days
  - Cost: $10/meal
  - Total charged: $30 million

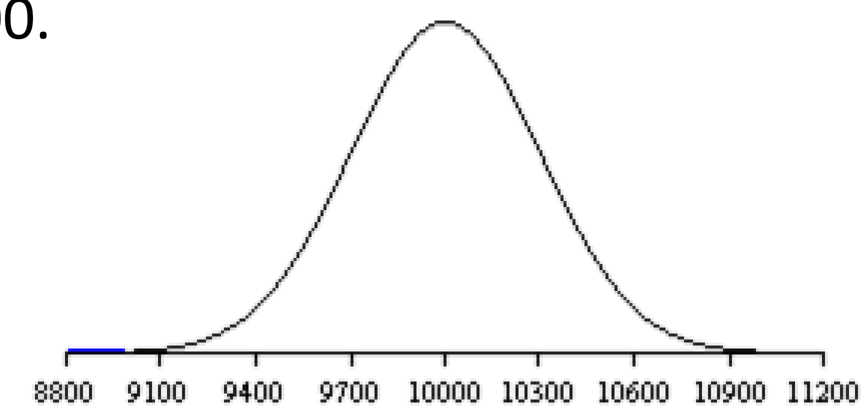
- The government believes that the charges of the contractor are too high.
- The government obtains a random sample of 30 days
  - Average number of meals for 30 days: 8,983 Meals Served
  - Suppose that **population standard deviation** is 1643.17 meals per day

What is the government trying to determine?

Cost of Service and the CLT

How likely is it to obtain a number as small as 8,983 if the contractor's claim was true?

Answer: Invoking the CLT, $\bar{X}$ has a normal distribution, with mean 10,000 and a standard deviation of 300.



$P(\bar{X} \leq 8{,}983) =$ NORM.DIST(*8983*,10000,300,1)=0.00035

$$P\left(Z \leq \frac{8{,}983-10{,}000}{300}\right) = NORM.S.DIST(-3.39) = 0.00035$$

Hypothesis Testing

- Hypothesis testing can be applied to a wide class of problems.
- It enables us to arrive at statistical conclusions about an uncertain outcome
  - Is the production under control?
  - Is a company's tax filing fraudulent?
  - Did the marketing campaign result in higher brand awareness?

- The goal:
  - We want to make a decision in favor of one out of two possible scenarios using only limited amounts of information

Concepts in Hypothesis Testing

- Construct a set of two hypotheses:
  - **H_0** (Null Hypothesis): Baseline case, conservative, status quo
  - **H_a or H_1** (Alternative Hypothesis): The thing we are trying to prove, innovative, new idea, headline of the newspaper article/study title

- Note that the two hypotheses divide all possible outcomes into two non-overlapping sets.
- That is, H_0 and H_a are complements
  - H_0 and H_a must be mutually exclusive (i.e., non-overlapping)
  - H_0 and H_a must be collectively exhaustive (i.e., cover all possibilities)

Constructing Hypothesis

- Example
  - **H_0** (Null Hypothesis): the marketing campaign did not improve brand awareness
  - **H_a** (Alternative Hypothesis): the marketing campaign improved brand awareness

Statistically Speaking

μ_{old} = previous brand awareness

μ_{new} = new brand awareness

- **H_0** (Null Hypothesis): $\mu_{new} \leq \mu_{old}$
- **H_a** (Alternative Hypothesis): $\mu_{new} > \mu_{old}$

Concepts in Hypothesis Testing

- In this course, we will focus on **hypothesis testing on population parameters** $(\boldsymbol{\mu}\ \boldsymbol{or}\ \boldsymbol{p})$**.**

- **Decision is always given in terms of the null hypothesis**:
  - **Reject H_0** or **fail to reject H_0**; we never conclude "reject H_a", or "accept H_a".
  - We reject H_0 because our sample provided evidence against it
    - Interpretation in practice: H_a MAY be true.
  - We fail to reject H_0 because our sample didn't provide enough evidence against it
    - Technically, this does not necessarily mean that H_0 is true
    - Interpretation in practice: H_0 MAY be true.

Confidence Intervals vs. Hypothesis Testing

Confidence interval

- Goal: Estimate an unknown population parameter with sampling distribution

- Procedure: Collect a random sample and compute confidence interval

- Decision: "we are 95% confident that the population mean, μ is in [0.1, 0.9] "

Hypothesis test

- Goal: Test a hypothesis about a specific population parameter.
  - $H_0: \mu = 0.5, H_a: \mu \neq 0.5$

- Procedure: Collect random sample and determine whether to reject the null hypothesis

- Decision: " we reject (or fail to reject) the null hypothesis at 5% significance level "

Example: CEO Salary

- Suppose that average annual percentage salary increase for CEOs of mid-size corporations was 7% from 1999 to 2002
- For the 2003, due to a worsening economic situation, we hypothesize that the average salary increase was lower than in the previous years
- Let μ be the average salary increase in 2003, what is $\boldsymbol{H_a}$?
  - $\mu = 7\%$
  - $\mu > 7\%$
  - $\boxed{\mu < 7\%}$
  - $\mu \neq 7\%$

What is the difference between the two statements?

- … we hypothesize that the average salary increase was lower (greater) than in the previous years (previous years average was 7%)

H_a: μ < (>) 7%

- … we hypothesize that the average salary increase was different than in the previous years (previous years average was 7%)

H_a: μ ≠ 7%

One-Tailed vs. Two-Tailed Tests

- The hypothesis test are either one tailed or two tailed
  - If we are only interested in changes in one direction, we use a **one-tailed test**, which is supported only by evidence in a single direction, framed as < or >
    - H_0: $\mu \geq$ (or $\leq$) 7%
    - H_a: $\mu <$ (or $>$) 7%
  - If we are interested in changes in any direction, we use a **two-tailed test**, which is supported by evidence in either direction, framed as ≠
    - H_0: $\mu = 7\%$
    - H_a: $\mu \neq 7\%$
- **The null must include '='**
  - The hypothesized population mean $= 7\%$ for both tests.
- In general, we first set up the alternative hypothesis and the null is the complement of it.

Example: Aluminum Sheets

- An aircraft manufacturer needs to buy aluminum sheets with an average thickness of **0.05 inches**
- The manufacturer knows that significantly thinner sheets would be unsafe and thicker sheets would be too heavy
- A random sample of **100 sheets** from a potential supplier is collected and the thickness of each sheet in the sample is measured and recorded.
- The manufacturer needs to decide whether to hire the supplier or not based on the sample.
- Write a research hypothesis test for this problem: Identify the null hypothesis and the alternative hypothesis. What test should be run for the problem?

Example: Aluminum Sheets

- Identify the null hypothesis and the alternative hypothesis
  - Let μ be the mean thickness of aluminum sheets.
  - H_0: $\mu = 0.05$
  - H_a: $\mu \neq 0.05$

    The hypothesis test is two-tailed.

General Procedure for Hypothesis Tests

1. Construct H_0 and H_1
    - Identify the parameter of interest: **mean, proportion**, …
    - **H_0** (Null Hypothesis): Baseline case, conservative, status quo
    - **H_a** (Alternative Hypothesis): The thing we are trying to prove, innovative, new idea
    - Determine whether the test is **one-tailed** or **two-tailed**
    - Choose a **significance level** $\boldsymbol{\alpha}$ (0.01=1%, 0.05=5%, 0.1=10%)
2. Compute the value of test statistics
3. Choose an appropriate test statistic and perform the test using sample data:
    - **Critical value** method
    - **p-value** method
4. Interpret the results: **Reject $\boldsymbol{H_0}$** or **fail to reject $\boldsymbol{H_0}$**

Type I and Type II Errors

Type I & II Errors

- Type I Error
  - Reject the null hypothesis, even though it's true
  - CEO salary: "conclude that the average salary increase is less than 7%, while in reality it is the usual 7%"
- Type II Error
  - Fail to reject the null, even though it's false
  - CEO salary: "conclude that the average salary increase is the usual 7%, while in reality it is lower"

	Null hypothesis is TRUE	Null hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct outcome! (True positive)
Fail to reject null hypothesis	Correct outcome! (True negative)	Type II Error (False negative)

Type I & II Errors

- Identify the Type I and Type II errors for the aluminum sheet problem
  - Type I error:
  - Type II error

Type I & II Errors

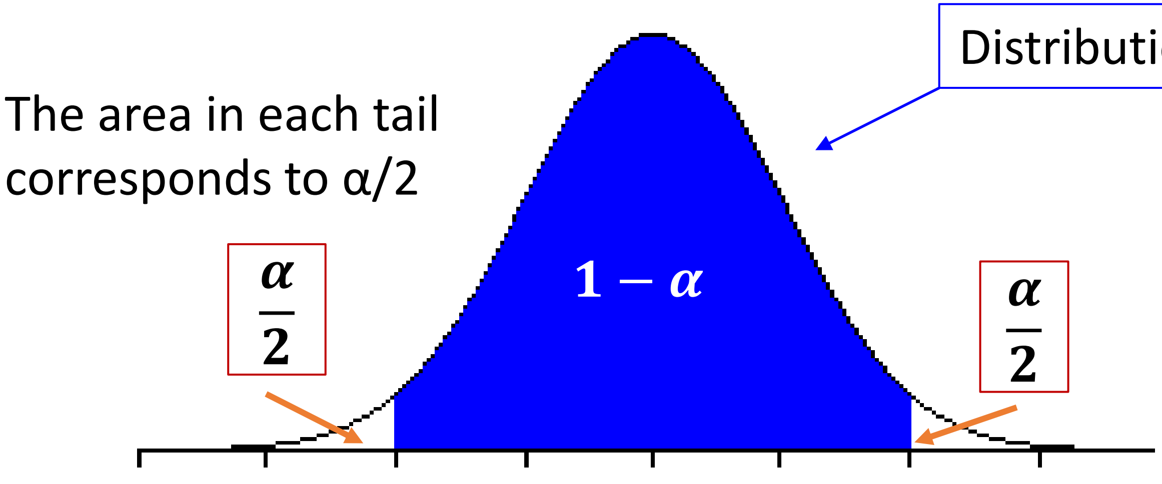
- Identify the Type I and Type II errors for the aluminum sheet problem
  - Type I error: We conclude that the average thickness is not equal to 0.05 inches and do not hire the supplier, while in fact it is 0.05 inches.
  - Type II error: We conclude that the average thickness is equal to 0.05 inches and hire the supplier, while in fact it is NOT equal to 0.05 inches.

Setting the Significance Level

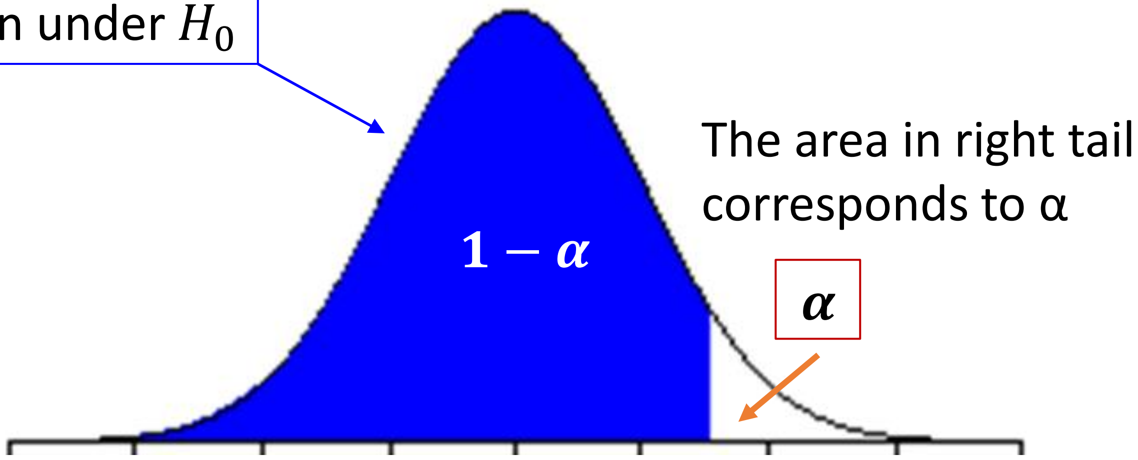
- **Level of significance α**: The probability of Type I error that the researcher is willing to tolerate.
  - Given that H_0 is true, what is your tolerance for rejecting it? 1% of the time? 5% of the time?
- Predetermined by the analyst -must be set before data collection and analysis
- In hypothesis testing we are primarily concerned with "controlling" the Type I Error
  - We often set α =5% (or smaller)

Significance Level α

Recall that the null must include '=': ex) $\mu = 7\%$ or $\mu = 0.05$
We assume that the null is true and perform hypothesis testing under H_0.
Therefore, the decision is always given in terms of the null hypothesis: Reject H_0 or fail to reject H_0



Two tailed test



Right one tailed test

Type I vs. II Errors

- We are "less concerned" with the Type II Error
- The evaluation of the Type II error is not straightforward.
  - The burden of proving H_a is true is on the researcher

- Ideally probability of Type I error should be low. However, **if α is set very low, then the probability of Type II error is high**
- Similarly, if the probability of Type II error is set very low, then the probability of Type I error is high
- Need to strike the right balance between Type I and Type II errors

Type I & II Errors (10/6(W))

- In making the tradeoff between likelihood of Type I and Type II errors and setting α, **what should be considered?**
  - We must consider the <u>costs</u> of making a Type I error relative to the cost of making a Type II error.
  - **If the cost of making a Type I error is high** (relative to the cost of Type II error), then **the level of significance should be low.**

Hypothesis Test

Population Mean

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

Test Statistic for Test of Mean

- μ_0 is the hypothesized population mean

  $H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$

  $H_0: \mu >= \mu_0$ $H_a: \mu < \mu_0$

  $H_0: \mu <= \mu_0$ $H_a: \mu > \mu_0$

- For a population mean with **unknown standard deviation**, we run the t-test. The formula of the test statistic is

$$\boxed{T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}}$$

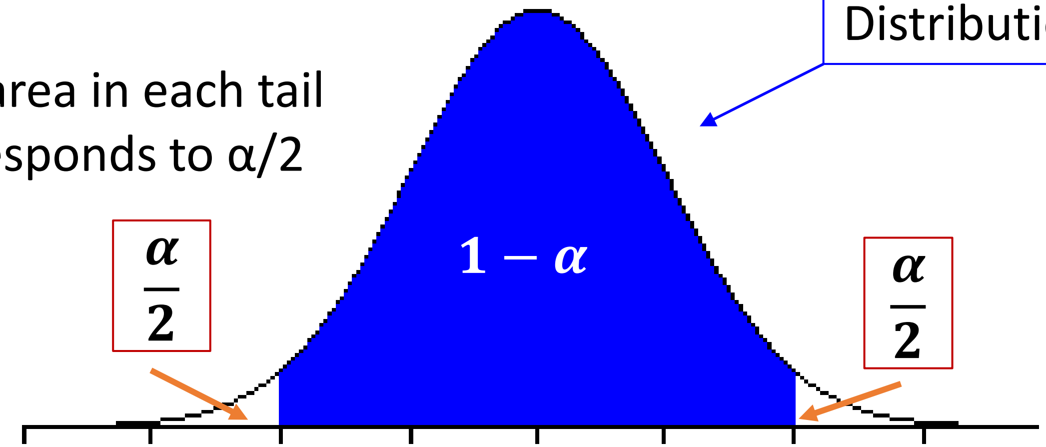
- The test statistic has a t-distribution with $(n - 1)$ degrees of freedom.
- The closer the sample statistic is to zero, the more **unlikely** it is to reject the null hypothesis.
- **It the population standard deviation is known, use Z-test**

Method 1: Critical Value

- A critical value is the standard score such that the area in the tail on the opposite side of the critical value (or values) from zero equals the corresponding significance level, α
- The value depends on whether the hypothesis test is one-tailed or two-tailed

Distribution under H_0

The area in each tail corresponds to α/2

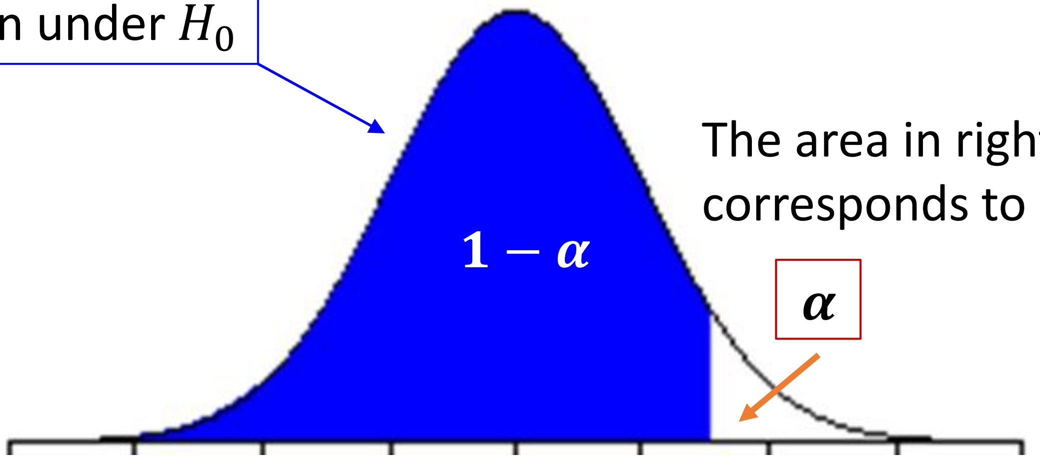


Critical Value (-)

Critical Value (+)

Two tailed test

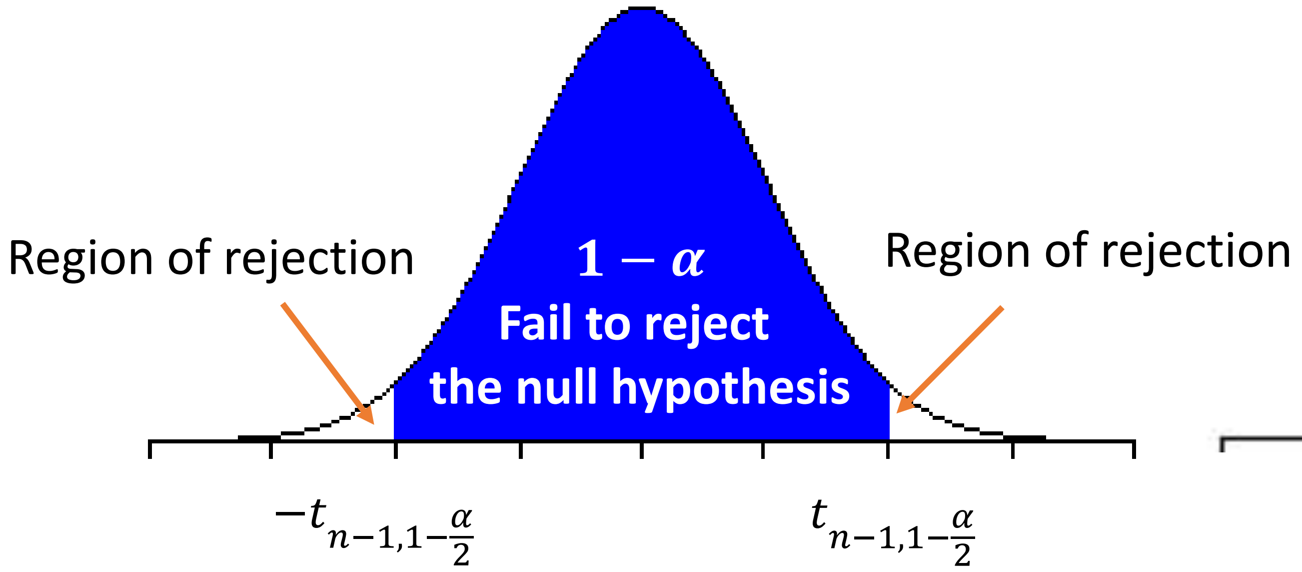
The area in right tail corresponds to α

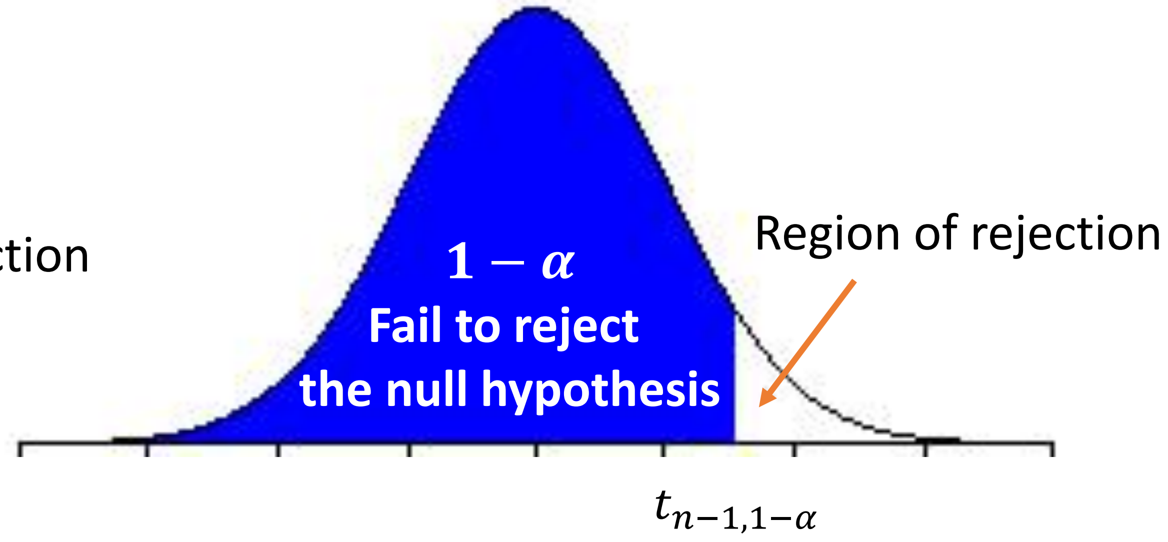


Critical Value (+)

(Right) One tailed test

Method 1: Critical Value Method for the T-test





Two tailed test:

$$H_0: \mu = \mu_0$$
$$H_a: \mu \neq \mu_0$$

The null hypothesis is rejected if

|t-value| > $\boldsymbol{t_{n-1,1-\frac{\alpha}{2}}}$

Right One tailed test:

$$H_0: \mu \leq \mu_0$$
$$H_a: \mu > \mu_0$$

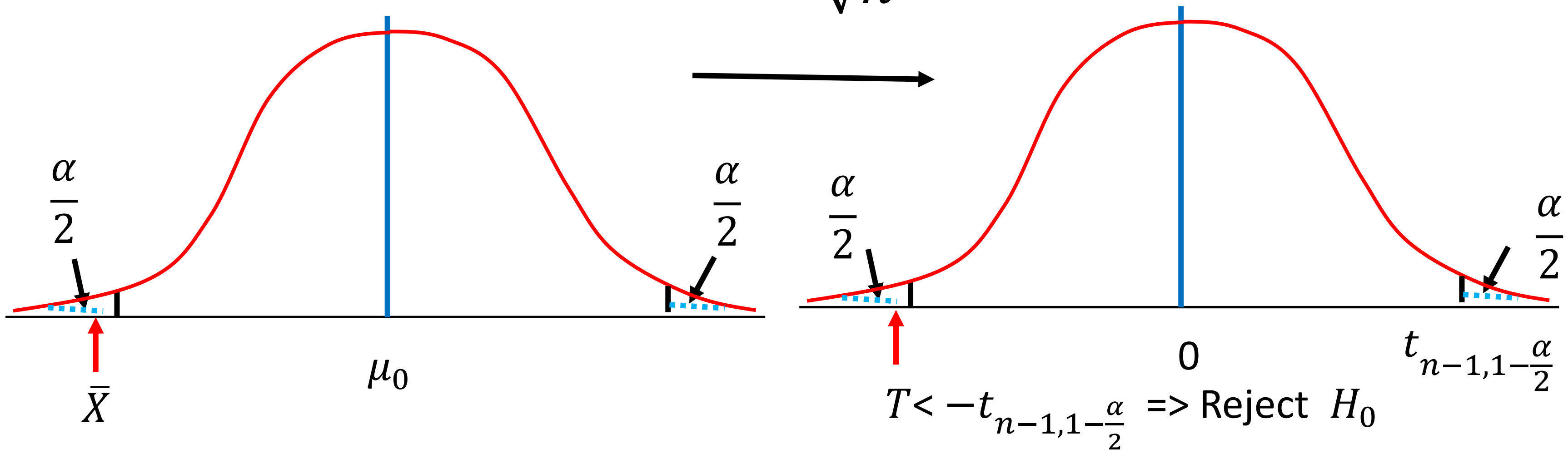
The null hypothesis is rejected if

t-value > $\boldsymbol{t_{n-1,1-\alpha}}$

Two Tailed Hypothesis Test for the Mean

H_0: $\mu = \mu_0$, H_a: $\mu \neq \mu_0$

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$



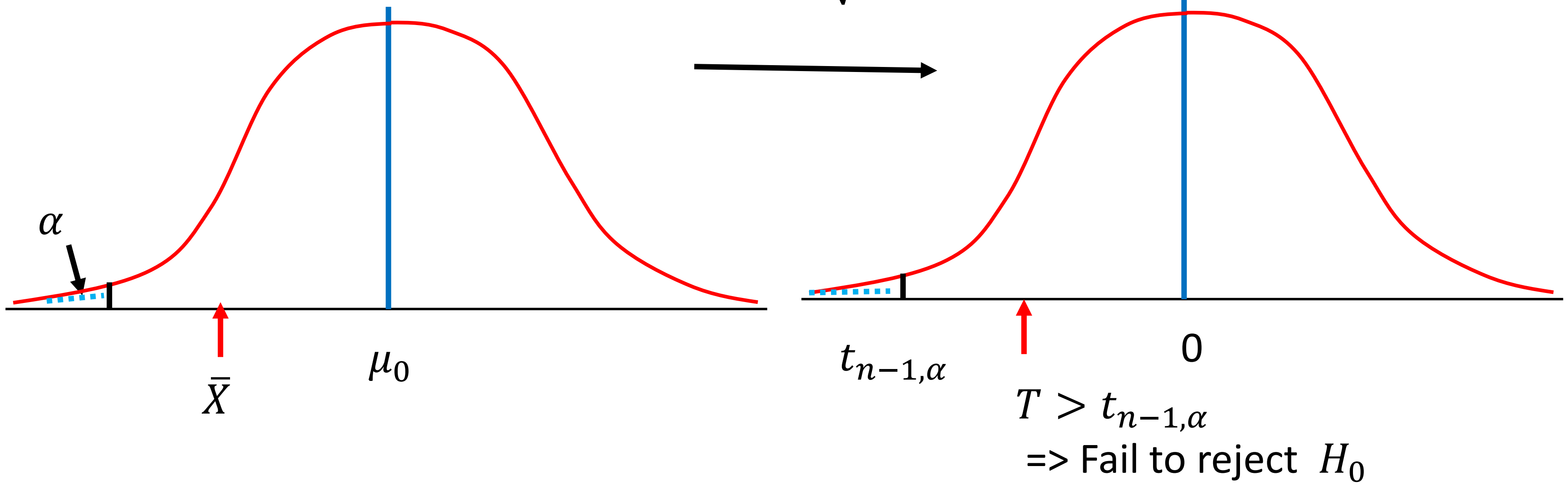
Distribution of $\bar{X}$ under H_0: $N(\mu_0, \sigma/\sqrt{n})$

t-distribution with df=n-1

Left One Tailed Hypothesis Test for the Mean

H_0: $\mu >= \mu_0$, H_a: $\mu < \mu_0$

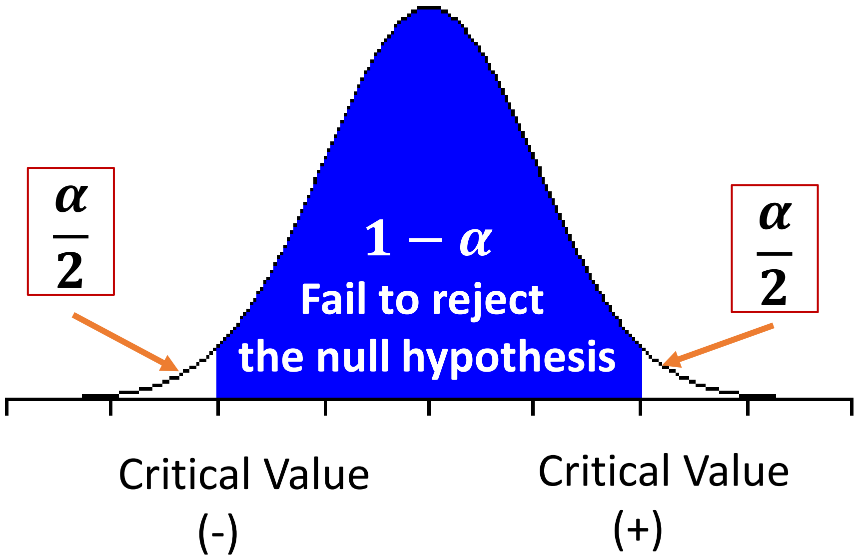
$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

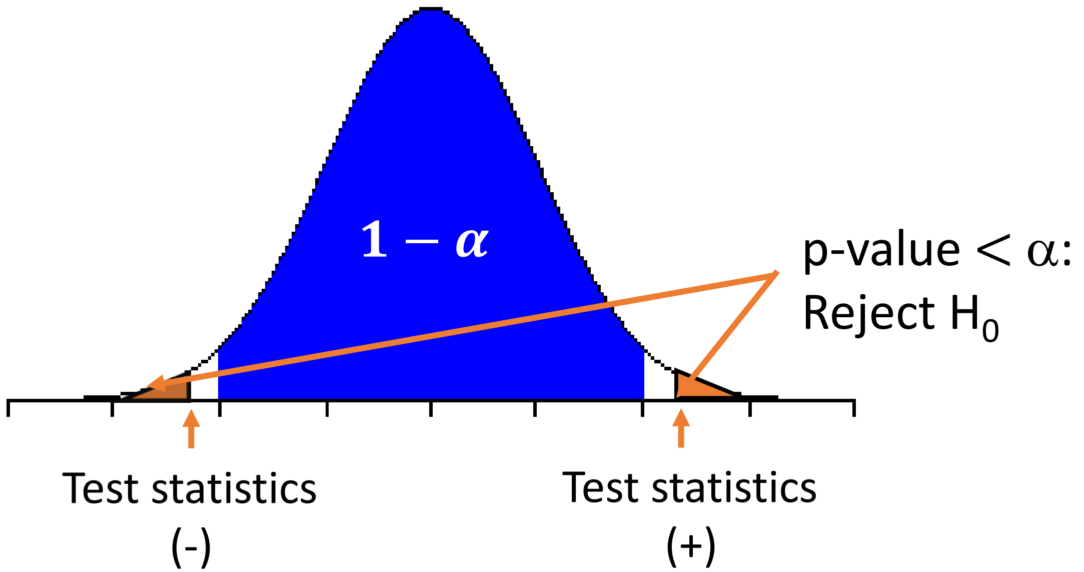


Distribution of $\bar{X}$ under H_0: $N(\mu_0, \sigma/\sqrt{n})$ t- distribution with df=n-1

Method 2: p-value

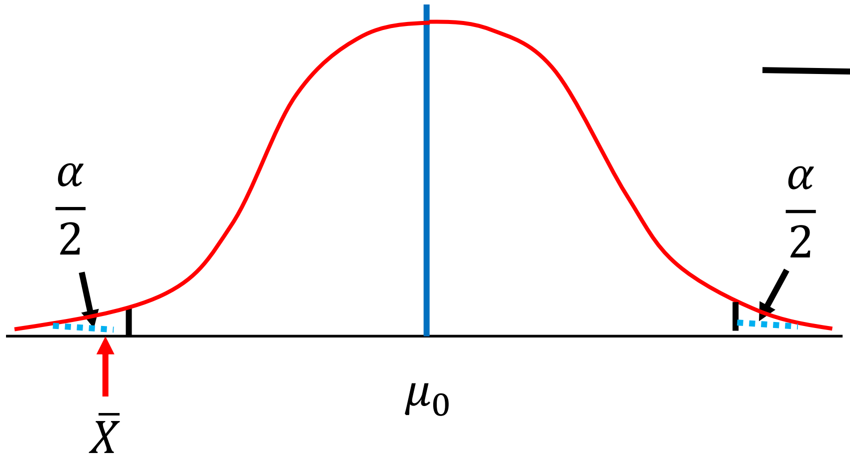
- The p-value is the probability of observing something as extreme as the test statistic assuming the null hypothesis is true
- We compute the p-value based on the test statistic and number of tails.
- If the p-value $< \alpha$, reject H_0
- If the p-value $\geq \alpha$, do not reject H_0

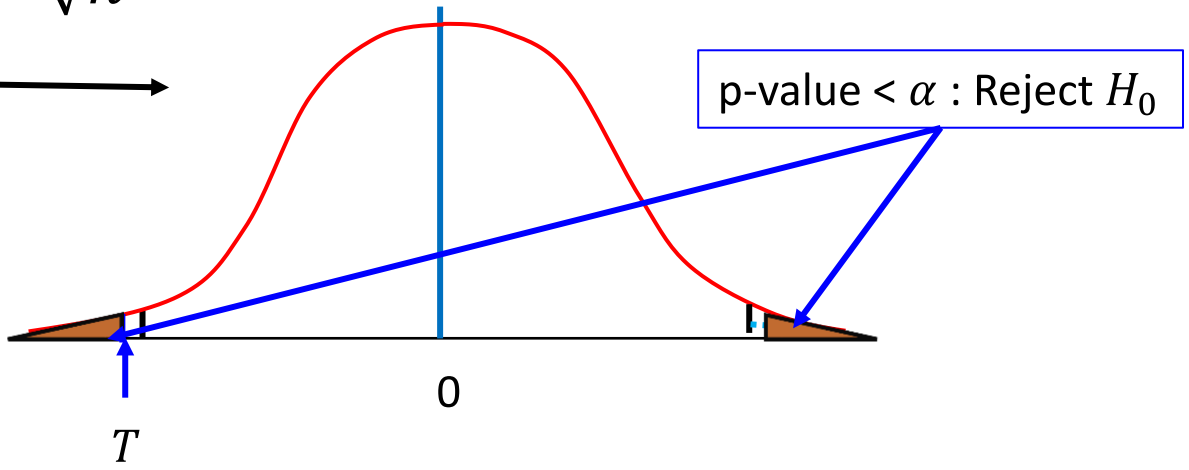




Two Tailed Hypothesis Test for the Mean

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$



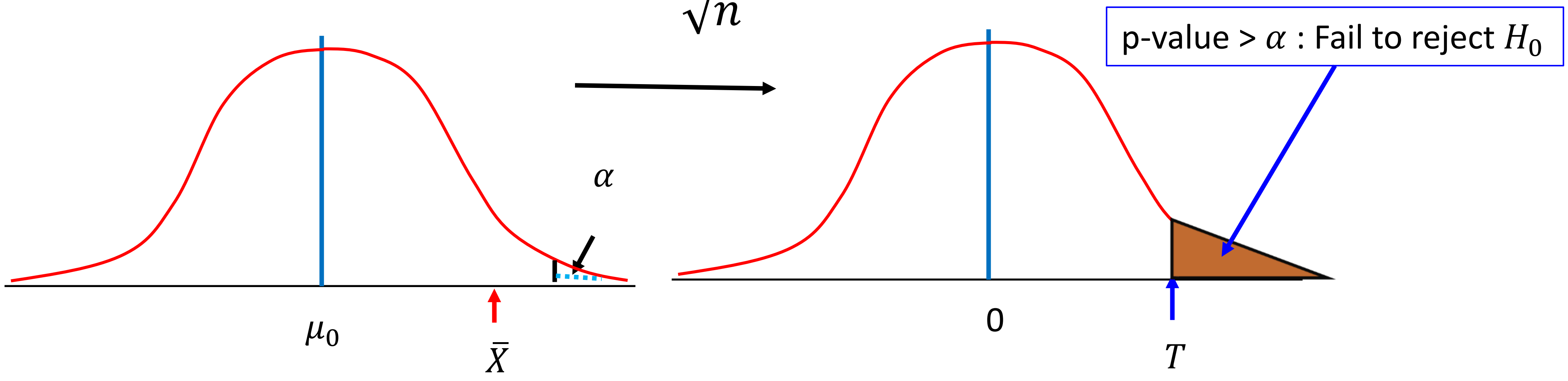


Distribution of $\bar{X}$ under H_0: $N(\mu_0, \sigma/\sqrt{n})$

t-distribution with df=n-1

Right One Tailed Hypothesis Test for the Mean

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$



Distribution of $\bar{X}$ under H_0: $N(\mu_0, \sigma/\sqrt{n})$ t-distribution with df=n-1

Interpretation of p-value

p-value means:

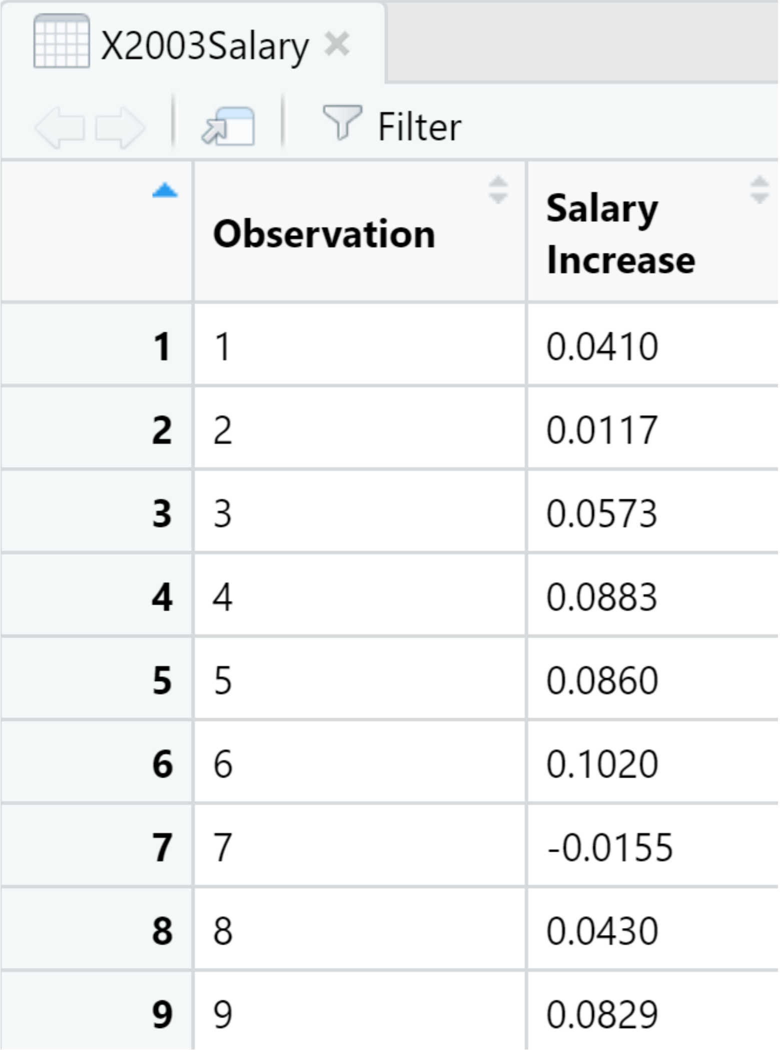
1. Assuming that the null hypothesis is true
2. there is a 100(p-value) percent chance
3. that we would get a result this extreme (point estimate), or more

35

Example: CEO Salary

- Suppose that average annual percentage salary increase for CEOs of mid-size corporations was 7% from 1999 to 2002
- For the 2003, due to a worsening economic situation, we hypothesize that the average salary increase was lower than in the previous years

- Use dataset 'From Excel' in R
- Import '2003Salary.xlsx'
- Summary statistics of 'Salary Increase':
  - Sample size (n)= 9
  - Sample mean $(\bar{X})$ = 0.055 = 5.5%
  - Sample stdev. (s)= 0.039 = 3.9%



X2003Salary

Filter

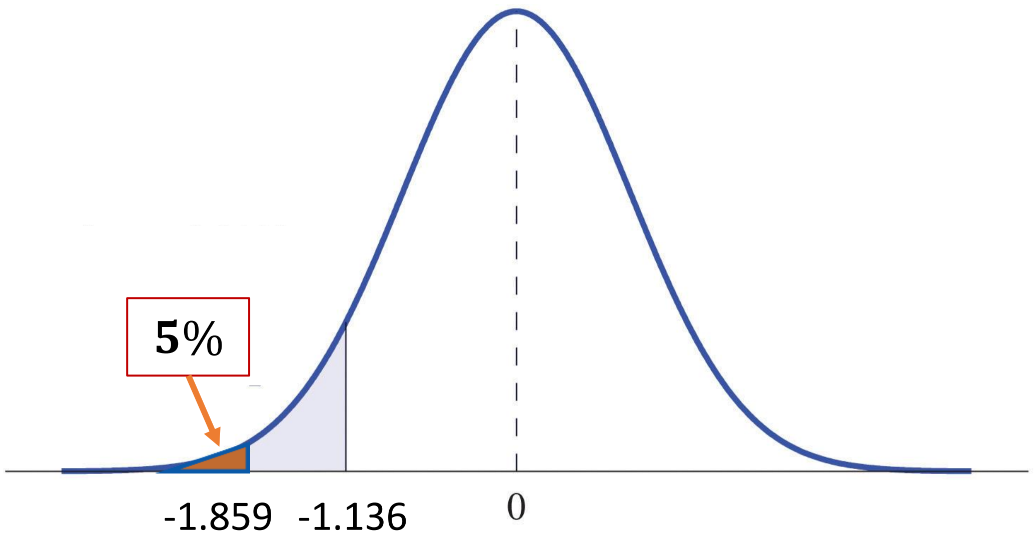
	Observation	Salary Increase
1	1	0.0410
2	2	0.0117
3	3	0.0573
4	4	0.0883
5	5	0.0860
6	6	0.1020
7	7	-0.0155
8	8	0.0430
9	9	0.0829

Example: CEO Salary

- Step 1: We have identified the null hypothesis and the alternative hypothesis
  - H_0: $\mu \geq \ 7\%$
  - H_a: $\mu < 7\%$
  - We have identified that the hypothesis test is a one-sided test
  - Set the significance level α to 5%
- Compute t- value, critical value, and p-value and make a decision
  - Note: we assume that the random variable Salary Increase follows a normal distribution.

Example: CEO Salary

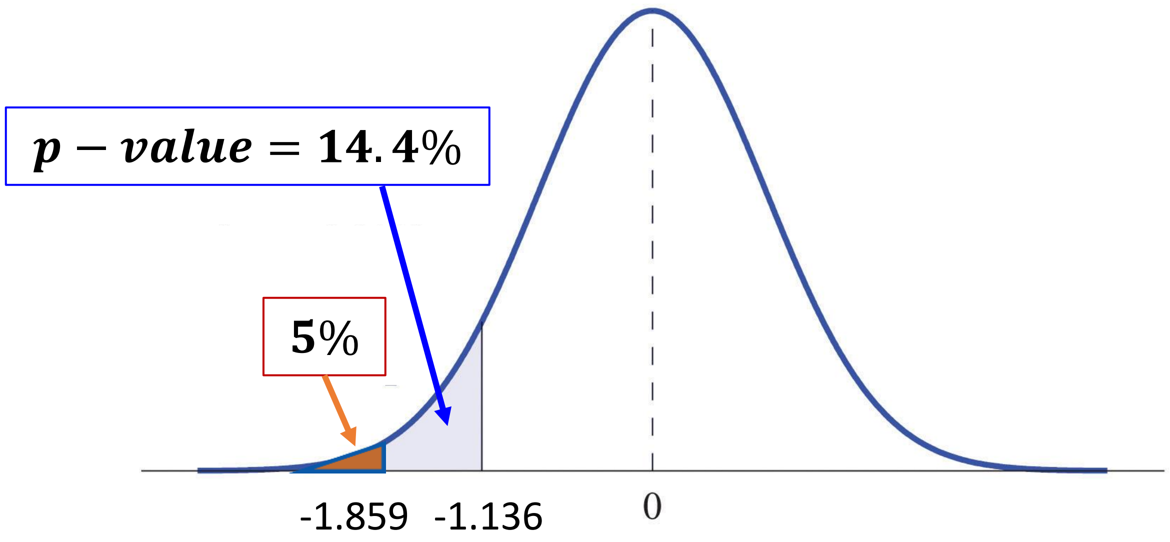
- Step 2: Compute the t-value
  - Sample size (n)= 9,
  - Sample mean $(\bar{X})$ = 0.055 = 5.5%
  - Sample stdev (s)= 0.039 = 3.9%
  - Hypothesized mean (μ_0) =0.07 = 7%
  - $T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{5.5\% - 7\%}{\frac{3.9\%}{\sqrt{9}}} = -1.136$



- Step 3: Make a decision
  - Method 1: Critical value method:
    - Critical value = T.INV(0.05,8) or qt(0.05,8) = -1.859 < -1.136 (t-value)
    - Conclusion: We do not reject the null hypothesis, that is, we do not have sufficient evidence for rejecting the null hypothesis

Example: CEO Salary

- Step 3: Make a decision
  - Method 2: p - value method by hand
  - p-value = T.DIST(-1.136,8,1) or pt(-1.136,8) = 0.144 = 14.4% > 5%: p-value is greater than α
  - Conclusion: We do not reject the null hypothesis



Example: CEO Salary

Method 2: p-value method using t.test() function in R

t.test(**data**, **mu** = mu_0, **alternative** = "two.sided", "less“ or "greater", **conf.level** = 0.95, …)

```
> t.test(X2003Salary$`Salary Increase`,mu=0.07,alternative = 'less', conf.level = 0.95)
        One Sample t-test
data:  X2003Salary$`Salary Increase`
t = -1.1356, df = 8, p-value = 0.1445
alternative hypothesis: true mean is less than 0.07
95 percent confidence interval:
        -Inf 0.07944177
sample estimates:
 mean of x
0.05518889
```

> 0.05 : **We do not reject the null**

Example: CEO Salary

Step 4: Interpret the results

In the CEO example our conclusion is that …

1. …the average salary increases in 2003 did not fall below the 7%
2. … the average salary increases in 2003 was below 7%
3. …the average salary increases in 2003 was the usual 7%

Interpretation of p-value

The p-value of 0.145 means that :

1. Assuming that the true average percentage salary increase is 7%
2. there is a 14.5% chance
3. that we get a mean of 5.5% or less on a sample of 9.

Example: Aluminum Sheets

- Use dataset 'AluminumSheet.xlsx'
- Summary statistics of 'Thickness':
  - Sample size (n)= 100
  - Sample average $(\bar{X})$= 0.04802 inches
  - Sample standard deviation (s) = 0.00873 inches
- Compute t- value, critical value, and p-value and make a decision

	A
1	Thickness
2	0.04904
3	0.054092
4	0.048577
5	0.050288
6	0.045012
7	0.045299
8	0.054733
9	0.048369
10	0.052945
11	0.043319
12	0.040177
13	0.034675
14	0.038976
15	0.04643

Example: Aluminum Sheets

- Step 1: Identify the null hypothesis and the alternative hypothesis

Let μ be the mean thickness of aluminum sheets.

H_0: μ = 0.05

H_a: μ ≠ 0.05

- The hypothesis test is a two-tailed test
- Set the significance level α to 5%

- Step 2: Compute the t-value

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0.04802 - 0.05}{\frac{0.00873}{\sqrt{100}}} = -2.2682$$

Example: Aluminum Sheets

Step 3: Make a decision

Method 1: Critical value method by hand

- Critical value = T. INV$(0.975, 99)$ or $\text{qt}(0.975,99)$ = 1.98 < |-2.2682(t-value)|
- Conclusion: We reject the null hypothesis

Method 2: p-value method by hand

- T. DIST(-2.2682$, 99,1)$ or $\text{pt}(-2.2682, 99) = 0.0127$
- p-value = 2*0.0127= 0. 0254 < 0.05
- Reject H_0 at α= 0.05 & 0.1 *Do not buy sheets from the supplier!*
- What about α= 0.01? → Do not reject! -> *Hire the supplier!*

Example: Aluminum Sheets

Method 2: p-value method using t.test function

```
> t.test(AluminumSheet$Thickness,mu=0.05, conf.level = 0.95, alternative = 'two.sided')
        One Sample t-test
data:  AluminumSheet$Thickness
t = -2.2682, df = 99, p-value = 0.02549
alternative hypothesis: true mean is not equal to 0.05
95 percent confidence interval:
 0.04628847 0.04975212
sample estimates:
mean of x
0.0480203
```

< 0.05 : **We reject the null**

Interpretation of p-value

The p-value of 0.02549 means that :

1. Assuming that the true average thickness is 0.05 inches
2. there is a 2.5% chance
3. that we get a mean of 0.04802" or more extreme on a sample of 100 sheets.

- Our sample was 0.00198" away from the goal. If the true mean was 0.05", there's a 2.5% chance that their sheets would be more than 0.00198" away from the target thickness.

More Examples on Hypothesis Test for the Population Mean

Meal Service

Overfilling Beverage Bottles

Example: Meal Service (10/11(M))

- A government contractor provided services to the military in a troubled region.
  - Average of 10,000 daily meals provided.
  - Operations lasted 300 days
  - Cost: $10/meal
  - Total charged: $30 million

- The government believes that the charges of the contractor are too high.
- The government obtains a random sample of 30 days
  - Average number of meals for 30 days: 8,983 Meals Served
  - Suppose that **population standard deviation** is 1643.17 meals per day

Meal Service: Hypothesis Test

- What is the government trying to determine?
- Construct the hypothesis for the government.
- What is the risk of Type I and Type II error?
- Make a decision for $\alpha = 1\%, 5\%$, and 10%, using two methods.
- Interpret the p-value.

Meal Service: Hypothesis Test

- What is the government trying to determine?

  To determine whether the contractor's charges are accurate or not.

- Construct the hypothesis for the government.
  - Let μ be the mean number of meals served per day.
  - *H_0: The contractor's charges are accurate, $\mu \geq 10000$* $(\mu_0 = 10000)$
  - *H_1: The contractor overcharges the government, $\mu < 10000$*
  - *This is a one sample, one-tailed Z-test*

- What is the risk of Type I and Type II error?
  - *Type I error: Fire the contractor when the contractor's charges are correct.*
  - *Type II error: Do not fire the contractor when the contractor overcharges the government.*

Meal Service: What is your conclusion?

- Set up a test statistic and compute the value.
  - *Given n = 30,* $\bar{X}$ *= 8,983, s =1643.17*
  - $Z = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{8983-10000}{\frac{1643.17}{\sqrt{30}}} = -3.39$
- Make a decision for $\alpha = 1\%, 5\%$, and 10%, using two methods.
  - *Critical Value Method:*
    - *Reject* H_0 *if* $Z = -3.39 < Z_{1,\alpha}$
    - $Z_{0.1} = -1.28$, $Z_{0.05} = -1.64$, $Z_{0.01} = -3.09$
  - *p-value = NORM.S.DIST(-3.39, 1) or pnorm(-3.39) = 0.0003*
  - *Decision: Reject* H_0 *=> Yes, the contractor overcharged the government*

Example: Overfilling Beverage Bottles

- Quality control example
  - Goal is to ensure that the mean fill level is 12 oz.
  - If the process is overfilling, it costs $5,000 to stop and correct the process
  - If the process overfills bottles, it costs $2,000 in daily losses
- On one particular day, a sample of 32 bottles yields a mean of 12.100 oz., and a standard deviation of 0.149 oz.
  - Should the production manager conclude that the process is systematically overfilling?

Overfilling: Hypothesis Test

- What is the manager trying to determine?
- Construct the hypothesis for the manager.
- What is the risk of Type I and Type II error? Which one is worse?
- Make a decision for $\alpha = 1\%, 5\%$, and 10%, using two methods.
- Interpret the p-value.

Overfilling: Formulating the Hypothesis Test

- Hypotheses
  - H_0: The process is not overfilling bottles or $\mu \leq 12$ $(\mu_0 = 12)$
  - H_1: The process is overfilling bottles or $\mu > 12$
  - This is a one sample, one-tailed T- test

- Potential Errors
  - Type I: Stop the process when it is not overfilling bottles
  - Type II: Do not stop the process when it is overfilling bottles
- Which is worse?
  - Type II error (assuming that the process is inspected every week)
- How does this affect our choice of α?
  - Set α to be high, say 10%

Overfilling: "By Hand"

- Compute test statistic
  - *Given n = 32,* $\bar{X}$ *= 12.1 oz., s = 0.149363907oz.*
  - $T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{12.1 - 12}{\frac{0.149363907}{\sqrt{32}}} = 3.60976713$
- *Critical value method*:
  - *Reject* H_0 *if* $T = 3.6098 > t_{n-1,1-\alpha}$ =T.INV(1-α, 31) or qt(1-α, 31)
  - $t_{31,\,0.9} = 1.31$, $t_{31,\,0.95} = 1.70$, $t_{31,\,0.99} = 2.46 \ll 3.6098$
- p-value = *T.DIST.RT(3.6098, 31) or 1- pt (3.6098, 31) = 0.000533 << α*

- Decision: Reject H_0 => Stop the process!

Example: Overfilling “using R”

```
> t.test(Beverage_Bottling$Fill, mu=12, conf.level = 0.95, alternative = "greater")
```

One Sample t-test

data: Beverage_Bottling$Fill

$\boxed{\text{t = 3.6098, df = 31, p-value = 0.0005331}}$ $< \alpha$: **We reject the null**

alternative hypothesis: true mean is greater than 12

95 percent confidence interval:

12.05054   Inf

sample estimates:

mean of x

12.09531

Interpretation of p-value

p-value means:

1. Assuming that the mean fill level is 12 oz (the machine is working fine)
2. there is a 0.05% chance
3. that the machine would fill 32 bottles to an average of 12.1 ounces or more

Really strong evidence against the null!

Next …

- Hypothesis test for the proportion