Summary---Topic 6: Hypothesis Testing

A statistical hypothesis is a claim about the <u>population parameter</u>
E.g. population mean, population standard deviation, or population proportion, etc.

Hypothesis Test for the Population Mean

- ❑ Critical Value Approach
- ❑ p-Value Approach

Hypothesis Testing Procedure

Step 1: Define hypotheses
Step 2: Collect the data and identify the rejection region(s)
Step 3: Compute test statistic
Step 4: Make statistical decision

Step 1: Define Hypotheses

- Always about a population parameter (μ, σ), rather than a sample statistic ($\bar{X}$, s)
- Null hypothesis, H_0: Always contains the "=" sign
- Alternative hypothesis, H_1: Never contains the "=" sign (Mutually exclusive and collectively exhaustive from H_0)
- Always assumed H_0 is true at start (i.e. assume the hypothesis regarding to population parameter is true at start), and then use sample statistics to assess the strength of the evidence against H_0 so as to determine whether H_0 should be rejected or not
- Three sets of hypotheses to be tested

Two-tail test	Lower-tail test	Upper-tail test
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$

Critical Value Approach--Summary

<table>
<tr><th>Case</th><th>Population Distribution</th><th>Sample Size(n)</th><th>Sampling Distribution of Mean($\bar{X}$)</th><th>σ known, use Z distribution, $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$</th><th>σ unknown, use t distribution, $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$</th></tr>
<tr><td>1</td><td>Normal</td><td>Whatever</td><td>Normal</td><td rowspan="3">Two-tail test
$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$
Reject H_0 if:
$Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$

Lower-tail test
$H_0: \mu \geq \mu_0$
$H_1: \mu < \mu_0$
Reject H_0 if:
$Z < -Z_{\alpha}$

Upper-tail test
$H_0: \mu \leq \mu_0$
$H_1: \mu > \mu_0$
Reject H_0 if:
$Z > Z_{\alpha}$</td><td rowspan="3">Two-tail test
$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$
Reject H_0 if:
$t < -t_{\alpha/2, (n-1)}$ or $t > t_{\alpha/2, (n-1)}$

Lower-tail test
$H_0: \mu \geq \mu_0$
$H_1: \mu < \mu_0$
Reject H_0 if:
$t < -t_{\alpha, (n-1)}$

Upper-tail test
$H_0: \mu \leq \mu_0$
$H_1: \mu > \mu_0$
Reject H_0 if:
$t > t_{\alpha, (n-1)}$</td></tr>
<tr><td>2</td><td>Unknown / not normal</td><td>n ≥ 30</td><td>By CLT, sample mean $\bar{X}$ is approximately normally distributed</td></tr>
<tr><td>3</td><td>Unknown</td><td>n < 30</td><td>Not normal; We need additional assumption. (population distribution is normal, then go to case 1)</td></tr>
</table>

- **Reject H_0 ➔ There is sufficient evidence that the H_1 is true.**
- **Do not reject H_0 ➔ There is insufficient evidence that the H_1 is true.**

p-value Approach--Summary

<table>
<tr><th>Case</th><th>Population Distribution</th><th>Sample Size(n)</th><th>Sampling Distribution of Mean($\bar{X}$)</th><th>σ known, use Z distribution, $z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$</th><th>σ unknown, use t distribution, $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$</th></tr>
<tr><td>1</td><td>Normal</td><td>Whatever</td><td>Normal</td><td rowspan="3">Two-tail test
$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

p-value= $\mathrm{P}(Z \leq$</td><td rowspan="3">Two-tail test
$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

p-value= $\mathrm{P}(t_{n-1} \leq -|t|) + \mathrm{P}(t_{n-1} \geq |t|)$

Lower-tail test
$H_0: \mu \geq \mu_0$
$H_1: \mu < \mu_0$
p-value= $\mathrm{P}(t_{n-1} \leq t)$

Upper-tail test
$H_0: \mu \leq \mu_0$
$H_1: \mu > \mu_0$
p-value= $\mathrm{P}(t_{n-1} \geq t)$</td></tr>
<tr><td>2</td><td>Unknown / not normal</td><td>n ≥ 30</td><td>By CLT, sample mean $\bar{X}$ is approximately normally distributed</td></tr>
<tr><td>3</td><td>Unknown</td><td>n < 30</td><td>Not normal;
We need additional assumption. (population distribution is normal, then go to case 1)</td></tr>
</table>

- **If p-value < α➔Reject H_0 ➔ There is sufficient evidence that the H_1 is true.**
- **if p-value ≥ α➔Do not reject H_0 ➔ There is insufficient evidence that the H_1 is true.**

Errors in Decision Making

- Type I error: Reject H_0 **given** H_0 is true

→ P(Type I error) = level of significance = α

- Type II error: Do not reject H_0 **given** H_0 is false

→ P(Type II error) = β

Decision	**The Truth**	
	$\boldsymbol{H_0}$ True	$\boldsymbol{H_0}$ False
Do not reject $\boldsymbol{H_0}$	Right decision Confidence (1- α)	Wrong decision Type II Error (β)
Reject $\boldsymbol{H_0}$	Wrong decision Type I Error (α)	Right decision Power (1- β)

- There would be a tradeoff between type I error and type II error. When α ↓, β ↑
- To decrease both errors, we need increase the sample size n.

Exercises and Solutions

Q3. A manufacturer of chocolate candies uses machines to package candies as they move along a filling line. Although the packages are labeled as 8 ounces, the company wants the packages to contain a mean of 8.17 ounces so that virtually none of the packages contain less than 8 ounces. A sample of 50 packages is selected periodically, and the packaging process is stopped if there is evidence that the mean amount packaged is different from 8.17 ounces. Suppose that in a particular sample of 50 packages, the **mean amount dispensed is 8.159 ounces**, with a **sample standard deviation of 0.051 ounce**.

a) Is there evidence that the population mean amount is different from 8.17 ounces? (Use a 0.05 level of significance.)

b) Compute the p-value and interpret its meaning.

Solution: a) Let μ be the mean amount of the package

$$H_0: \mu = 8.17 \text{ ounces}$$
$$H_1: \mu \neq 8.17 \text{ ounces}$$

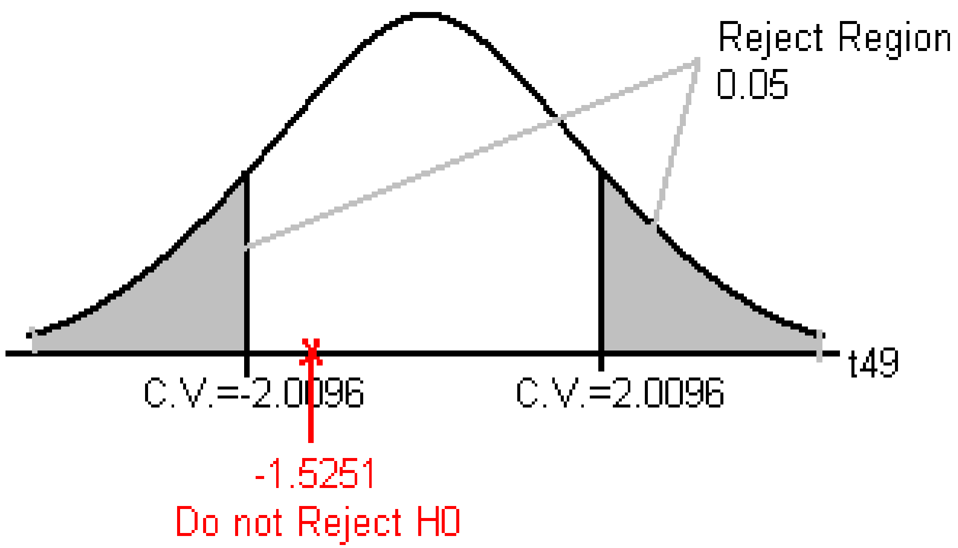
Since n = 50 > 30 from unknown population distribution, by Central Limit Theorem, the sampling distribution of $\bar{X}$ is approximately normal. Furthermore, σ is unknown, thus t-test should be used (two-tail test).

Then we compute the test statistic, $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{8.159-8.17}{0.051/\sqrt{50}} = -1.5251$,

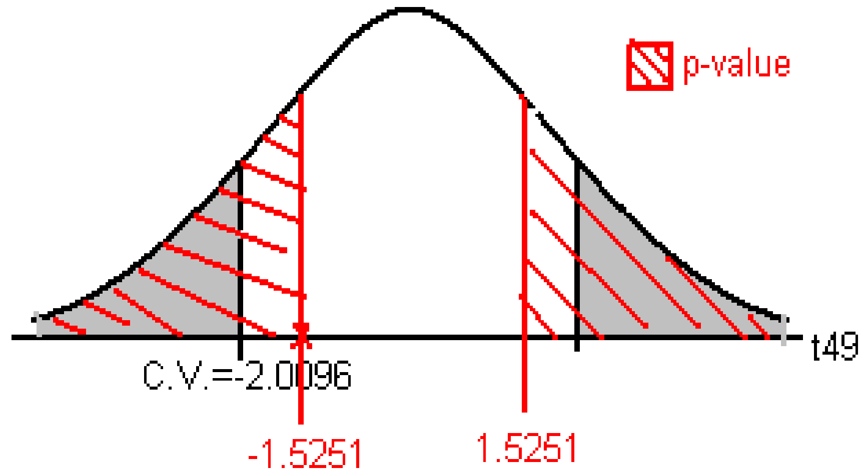
and the critical value with significance level $\alpha = 0.05$ is
$\pm t_{\alpha/2,n-1} = \pm t_{0.025,49} = \pm 2.0096$.

Since $-2.0096 < t < 2.0096$, we do not reject H_0 at $\alpha = 0.05$.

There is insufficient evidence that population mean amount is different from 8.17 ounces.



Q3. A manufacturer of chocolate candies
move along a filling line. Although the pa
company wants the packages to contain a
of the packages contain less than 8 ounce
periodically, and the packaging process is
amount packaged is different from 8.17 o
of 50 packages, the mean amount dispen
deviation of 0.051 ounce.



a) Is there evidence that the populatior
ounces? (Use a 0.05 level of significance.)

b) Compute the p-value and interpret its meaning.

Solution:

Since the test statistic $t = -1.5251$,

p-value$= \mathrm{P}(t_{n-1} \leq -|t|) + \mathrm{P}(t_{n-1} \geq |t|) = \mathrm{P}(t_{49} \leq -1.5251) + \mathrm{P}(t_{49} \geq$

<u>Interpretation</u>:

Probability of obtaining a test statistics 1.5251 or more or -1.5251 or less is between 0.1 and 0.2 exclusively, given H_0 is true.

Q4. The Glen Valley Steel Company manufactures steel bars. If the production process is working properly, it turns out that steel bars are **normally distributed** with **mean length of at least 2.8 feet**. Longer steel bars can be used or altered, but shorter bars must be scrapped. You select a **sample of 25 bars**, and the **mean length is 2.73 feet** and the **sample standard deviation is 0.20 feet**. Do you need to adjust the production equipment?

a) If you test the null hypothesis at the 0.05 level of significance, what decision do you make using the critical value approach to hypothesis testing?

b) If you test the null hypothesis at the 0.05 level of significance, what decision do you make using the p-value approach to hypothesis testing?

c) Interpret the meaning of the p-value in this problem.

d) Compare your conclusions in (a) and (b).

Solution: a) Let μ be the mean length of the steel bar

$$H_0: \mu \geq 2.8 \text{ feet}$$
$$H_1: \mu < 2.8 \text{ feet}$$

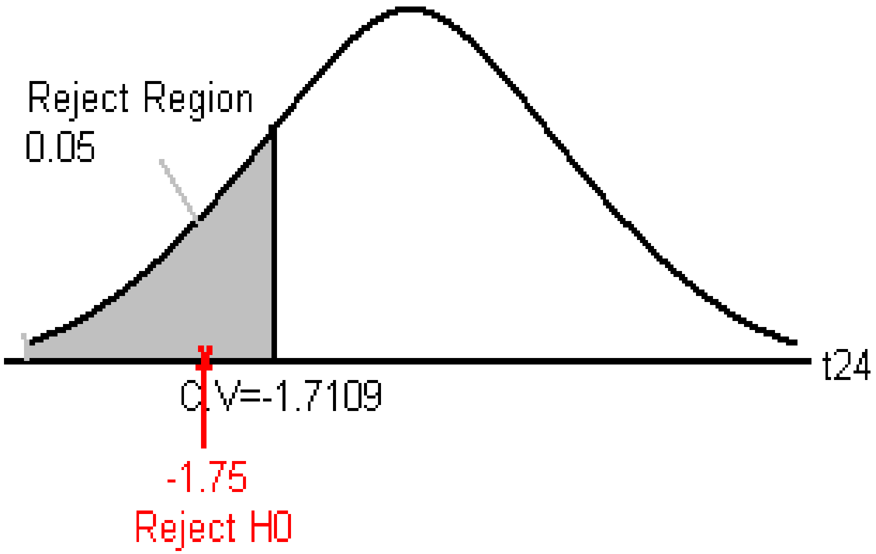
Since the population is normal distribution, the sampling distribution of $\bar{X}$ is normal distribution. Furthermore, σ is unknown, thus t-test should be used (**lower-tail test**).

Then we compute the test statistic, $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{2.73-2.8}{0.2/\sqrt{25}} = -1.75$,

and the critical value with significance level $\alpha = 0.05$ is
$-t_{\alpha,n-1} = -t_{0.05,24} = -1.7109$.

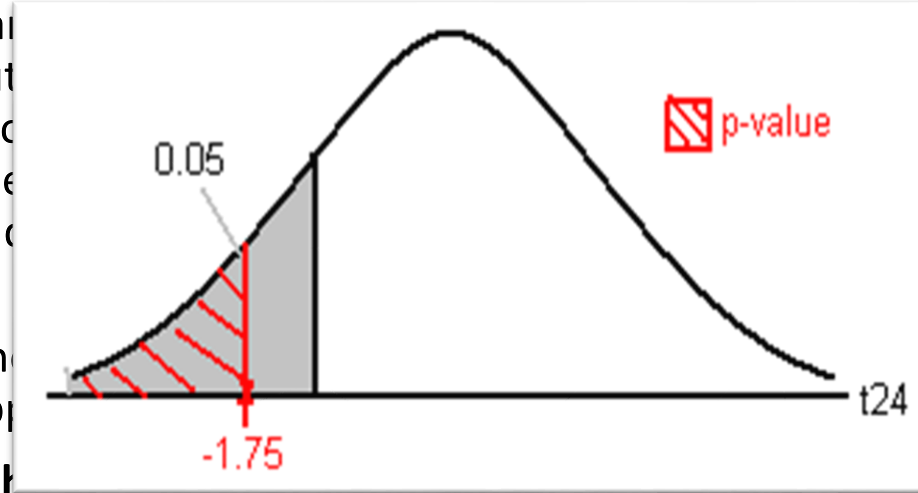
Since $t < -1.7109$, we reject H_0 at $\alpha = 0.05$.

There is sufficient evidence that the production equipment needs adjustment.



Q4. The Glen Valley Steel Company ma
process is working properly, it turns out
with mean length of at least 2.8 feet. Lo
shorter bars must be scrapped. You sele
length is 2.73 feet and the sample stand
adjust the production equipment?

a) If you test the null hypothesis at th
do you make using the critical value ap



b) If you test the null hypothesis at th
do you make using the p-value approach to hypothesis testing?

c) Interpret the meaning of the p-value in this problem.

d) Compare your conclusions in (a) and (b).

Solution:

b) Since the test statistic $t = -1.75$,
p-value= $\mathrm{P}(t_{n-1} \le t) = \mathrm{P}(t_{24} \le -1.75) = \mathrm{P}(t_{24} \ge 1.75) = (0.025, 0.05) < \alpha$.
Thus we reject H_0 at $\alpha = 0.05$.
There is sufficient evidence that the production equipment needs adjustment.

c) Interpretation:
Probability of obtaining a test statistics -1.75 or less is between 0.025 and 0.05 exclusively, given H_0 is true.

d) The conclusions are the same.

Q5. A bank branch located in a commercial district of a city has developed an improved process for serving customers during the 12:00 to 1 p.m. peak lunch period. The waiting time in minutes (operationally defined as the time the customer enters the line to the time he or she is served) of all customers during this hour is recorded over a period of a week. A random **sample of 15 customers** is selected, and the results are as follows:

4.21	5.55	3.02	5.13	4.77	2.34	3.54	3.20
4.50	6.10	0.38	5.12	6.46	6.19	3.79	

At the 0.05 level of significance, is there evidence that the average waiting time at a bank branch in a commercial district of the city is less than five minutes during the lunch period?

Solution: Let μ be the mean waiting time

$$H_0: \mu \geq 5 \text{ mins}$$
$$H_1: \mu < 5 \text{ mins}$$

Since n = 15 < 30 from unknown population distribution, the sampling distribution of $\bar{X}$ is not normal. **We need to assume the population distribution is normal**. In this way, the sampling distribution of $\bar{X}$ is normal. Furthermore, $\boldsymbol{\sigma}$ **is unknown**, thus t-test should be used (lower-tail test).

Then we need compute the test statistic, $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$,

where $\bar{X} = \frac{4.21+5.55+\cdots+6.19+3.79}{15} = 4.286667$,

and $s = \sqrt{\frac{(4.21-4.2867)^2+\cdots+(3.79-4.2867)^2}{15-1}} = 1.637985.$

Thus we obtain that

$$t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{4.286667-5}{1.637985/\sqrt{15}} = -1.6867.$$

And the critical value with significance level $\alpha = 0.05$ is

$-t_{\alpha,n-1} = -t_{0.05,14} = -1.7613.$

Since $t > -1.7613$, we do not reject H_0 at $\alpha = 0.05$.

There is insufficient evidence that the population average waiting time is less than 5 mins.

Q6*. A television documentary on over-eating claimed that Americans are about 10 pounds overweight on average. To test this claim, **18 randomly selected individuals were examined**, and their **average excess weight was found to be 12.4 pounds**, with a **sample standard deviation of 2.7 pounds**.

a) What assumption(s) is(are) required for performing the hypothesis testing in (ii) below?

b) At a significance level of 0.01, is there any reason to doubt the validity of the claimed 10-pound value?

c) Define the probability of type I error α and that of type II error β according to the context of this part.

Solution:

a) As n=18<30 from unknown distribution, **we need to assume the population distribution is normal**.

b) Let μ be the population mean of overweight

$$H_0: \mu = 10$$
$$H_1: \mu \neq 10$$

σ is unknown, thus t-test should be used (two-tail test).

Then we compute the test statistic, $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{12.4-10}{2.7/\sqrt{18}} = 3.7712$,

and the critical value with significance level $\alpha = 0.01$ is
$\pm t_{\alpha/2,n-1} = \pm t_{0.005,17} = \pm 2.8982$.

Since $t > 2.8982$, we reject H_0 at $\alpha = 0.01$.

There is sufficient evidence that the population mean overweight is not 10 pounds.

c)
Type I error (α)
= Pr(do not agree the claim of 10-pound overweight when in fact the claim is true)

Type I error (β)
=Pr(Agree the claim of 10-pound overweight when in fact the claim is false)

Q7. A management consultant has introduced new procedures to a reception office. He claims that the **receptionist should not do more than 10 minutes** of paperwork in each hour. A check is made on **40 random hours** of operation. The sample mean and sample standard deviation of the time spent on paperwork are found. Based on these figures, the null hypothesis that the **new procedures meet specifications is rejected** at a 1% level of significance.

a) After the consultant has asked the data entry clerk to show him the original data, he finds that the sample size should be 41, instead of 40. Should the null hypothesis that the new procedures meet specifications be rejected? Why or why not?

b) Peter, the manager of the reception office, asks the consultant to test the same hypothesis with a new level of significance of 5%. Should the null hypothesis that the new procedures meet specifications be rejected? Why of why not?

Solution: a) Let μ be the mean time spent on paperwork

$$H_0: \mu \leq 10 \text{ mins}$$
$$H_1: \mu > 10 \text{ mins}$$

Since n=40>30, the sampling distribution of $\bar{X}$ is approximately normal distribution(CLT). Furthermore, σ is unknown, thus t-test should be used (upper-tail test).

Then we need compute the test statistic, $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$

and the critical value with significance level $\alpha = 0.01$ is
$t_{\alpha,n-1} = t_{0.01,39} = 2.4258.$

<u>Based on these figures, the null hypothesis that the new procedures meet specifications is rejected at a 1% level of significance.</u>
➔ $t > 2.4258$

Now sample size n ↑ , n'=41, hence the critical value becomes

$$t_{\alpha,n'-1} = t_{0.01,40} = 2.4233,$$

and $s/\sqrt{n}\downarrow$, as a **result the new test statistic** $\boldsymbol{t' = \frac{\bar{X}-\mu_0}{s/\sqrt{n'}}}$ **will increase**.

➔ $t' > t > 2.4258 > 2.4233 = t_{\alpha,n'-1}$
➔ H_0 is still rejected.

Q7. A management consultant has introduced new procedures to a reception office. He claims that the receptionist should not do more than 10 minutes of paperwork in each hour. A check is made on 40 random hours of operation. The sample mean and sample standard deviation of the time spent on paperwork are found. Based on these figures, the null hypothesis that the new procedures meet specifications is rejected at a 1% level of significance.

b) Peter, the manager of the reception office, asks the consultant to test the same hypothesis with a new level of significance of 5%. Should the null hypothesis that the new procedures meet specifications be rejected? Why of why not?

Solution: b) Now the significance level $\alpha' = 0.05$, the critical value

$$t_{\alpha', n'-1} = t_{0.05,40} = 1.6839.$$

➔ $t' > t > 2.4258 > 1.6839 = t_{\alpha', n'-1}.$

Therefore, H_0 is still rejected.