情報連携学部

INIAD

Statistics and data analysis II

Week13

「Hypothesis test for mean」

Takashi Sano, Hirotada Honda

INIAD

Lecture plan

Week1: Introduction of the course and some mathematical preliminaries
Week2: Overview of statistics, One dimensional data(1): frequency and histogram
Week3: One dimensional data(2): basic statistical measures
Week4: Two dimensional data(1): scatter plot and contingency table
Week5: Two dimensional data(2): correlation coefficients, simple linear regression and concepts of Probability /
Probability(1):randomness and probability, sample space and probabilistic events
Week6:Probability(2): definition of probability, additive theorem, conditional probability and independency
Week7:Review and exam(i)
Week8: Random variable(1): random variable and expectation
Week9: Random variable(2): Chebyshev's inequality, Probability distribution(1):binomial and Poisson distributions
Week10: Probability distribution(2): normal and exponential distributions
Week11: From descriptive statistics to inferential statistics -z-table and confidwncw interval-
Week12: Hypothesis test(1) -Introduction, and distributions of test statistic (t-distribution)-
Week13: Hypothesis test(2) -Test for mean-
Week14: Hypothesis test(3) -Test for difference of mean-

※ Might be changed!

Copyright © 2021 by INIAD

2. Hypothesis test for mean

Copyright © 2021 by INIAD

Agenda

- Hypothesis test for mean
- In case population S.D. is known
- In case population S.D. is unknown
- Exercises

Copyright © 2021 by INIAD

2-1. Hypothesis test for mean

Copyright © 2021 by INIAD

Hypothess test for mean

- In the inferential statistics, we estimate the characteristics of the population from observed samples.

- Especially, if we validate the value of the population mean, it is called as the hypothesis test for mean.

Copyright © 2021 by INIAD

Usage

- By using samples, validate whether the product specification is correctly applied.

Copyright © 2021 by INIAD

Take care…

①Population variance is known？Unknown？

→ If known, apply z-dist., otherwise, t-dist.

②Two-sided？One -sided？

→ The p-value python returns depends on cases.
Is it two-sided p-val.? Or one-sided?

③The row data is given? Just some statistical indicators (sample mean, unbiased SD) are given?

Copyright © 2021 by INIAD

2-2. Hypothesis test for mean (In case the population S.D. is known)

Copyright © 2021 by INIAD

Hypothesis test for mean①
(In case the population S.D. is known)

- Let the population mean and S.D. of a normal distribution be μ and σ, respectively. Then, the n samples extracted from them also follows the normal distribution.
  - The sample mean remains as μ, but the S.D. of the samples reduces to $\frac{\sigma}{\sqrt{n}}$

- 95% C.I. of the normal distribution:

$$-1.96 \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96$$

Copyright © 2021 by INIAD

Hypothesis test for mean①
(In case the population S.D. is known)

・In case σ^2 is known, we use the following quantity as the test statistic:

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- Then, what is the p.d.f. of the test statistic?

N(0,1)!

Copyright © 2021 by INIAD

Hypothesis test for mean①
(In case the population S.D. is known)

- vi) For the significance level a, set the rejection region R that satisfies

$$P(T(X_1,X_2,\ldots,X_N)\in R)=a.$$

・In case of the two-sided test, it's out of the upper / lower 2.5-percentiles (=±1.96) of z-distribution.

・ In case of the two-sided test, it's out of the upper / lower 5-percentiles (=1.64).

Copyright © 2021 by INIAD

Example①

In a certain maker of a part (named as “M”) of computer, its diameter is described as 1.54[cm] in its product specification. In a certain sample survey, they extracted 8 samples randomly, and observed the following data of measured diameter [cm]:

1.53 1.57 1.54 1.57 1.53 1.55 1.56 1.53

It’s known that the population variance is σ^2=0.0001. Then, can you say that this part follow its product specification? Test with the significance level of 5%.

Copyright © 2021 by INIAD

Flow of hypothesis test

- i) Set the population (☚Similar to confidence interval)
- ii) Set the null hypothesis H_0.
- iii) Extract samples $x_1, x_2, \ldots, x_N$ from the population.
- iv) Find a statistics $T(x_1, x_2, \ldots, x_N)$ from the sample above.
- v) Calculate the probability density of the statistics $T(X_1, X_2, \ldots, X_N)$ for r.v.s $X_1, X_2, \ldots, X_N$.

Copyright © 2021 by INIAD

Flow of hypothesis test

- vi) For a certain significance level a, find a region R, where

  Usually, a=0.01 or 0.05

  $$P(T(X_1,X_2,\ldots,X_N)\in R)=a$$

  holds(This region R is called as the critical region)

- vii) If $T(x_1,x_2,\ldots,x_N)\in R$, reject the null hypothesis H_0 / otherwise, H_0 cannot be rejected.

Copyright © 2021 by INIAD

Example①【answer】

Flow of hypothesis test.

・Null hypothesis H_0 "The diameter of part "M" is1.54cm"

Alternative hypothesis H_1:" The diameter of part "M" is not 1.54cm"
(No good with too large nor too small results.)

・Find the test statistic $T(x_1,x_2,\ldots,x_N)$.

・Since the population variance is known, we use

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Copyright © 2021 by INIAD

Example①【answer】

- Find the p.d.f. of the test statistic t.

Copyright © 2021 by INIAD

Example①【answer】

- If X follows N(μ, σ^2), then

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

follows N(0,1).

Copyright © 2021 by INIAD

Example①【answer】

- Find the p.d.f. of the test statistic t.

  ⇒ z-distribution.

Copyright © 2021 by INIAD

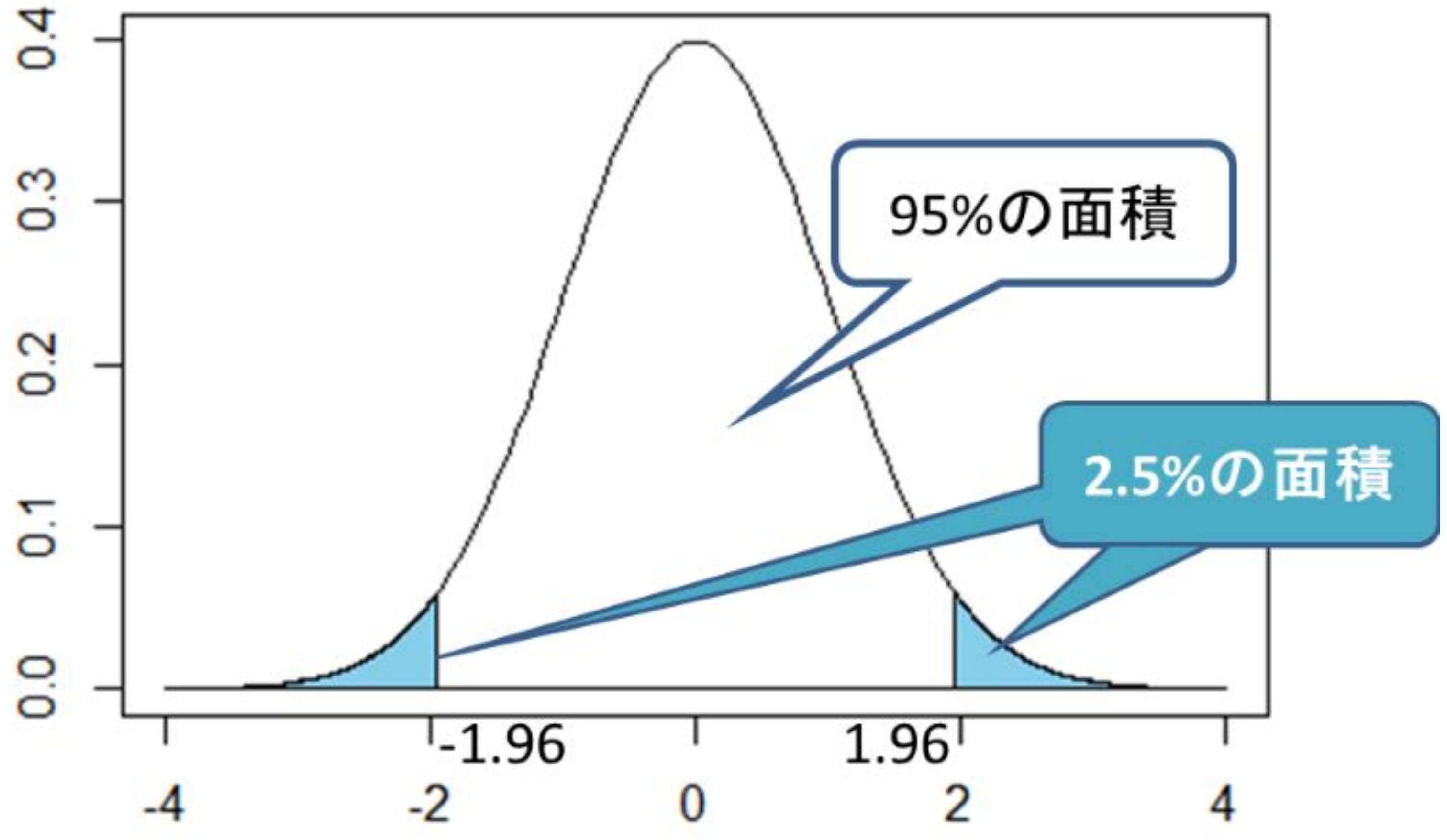
Example①【answer】

・Determine the rejection region.

⇒ Two-sided test from the form of the alternative hypothesis.

The rejection region is as follows（out of upper /lower2.5-percentiles of z-distribution）.

・|t|>1.96 is the rejection region.



Copyright © 2021 by INIAD

Example①【answer】

- Find the value of $T(x_1, x_2, \ldots, x_N)$:

$$\bar{x} = \frac{1.53 + 1.57 + \ldots + 1.53}{8} = 1.5475$$

$$\mu = 1.54 \qquad \sigma = \sqrt{0.0001} = 0.01$$

Thus, we have

$$t = \frac{1.5475 - 1.54}{\frac{0.01}{\sqrt{8}}} = 2.121$$

- Belong to the rejection region $|z|>1.96$!

Copyright © 2021 by INIAD

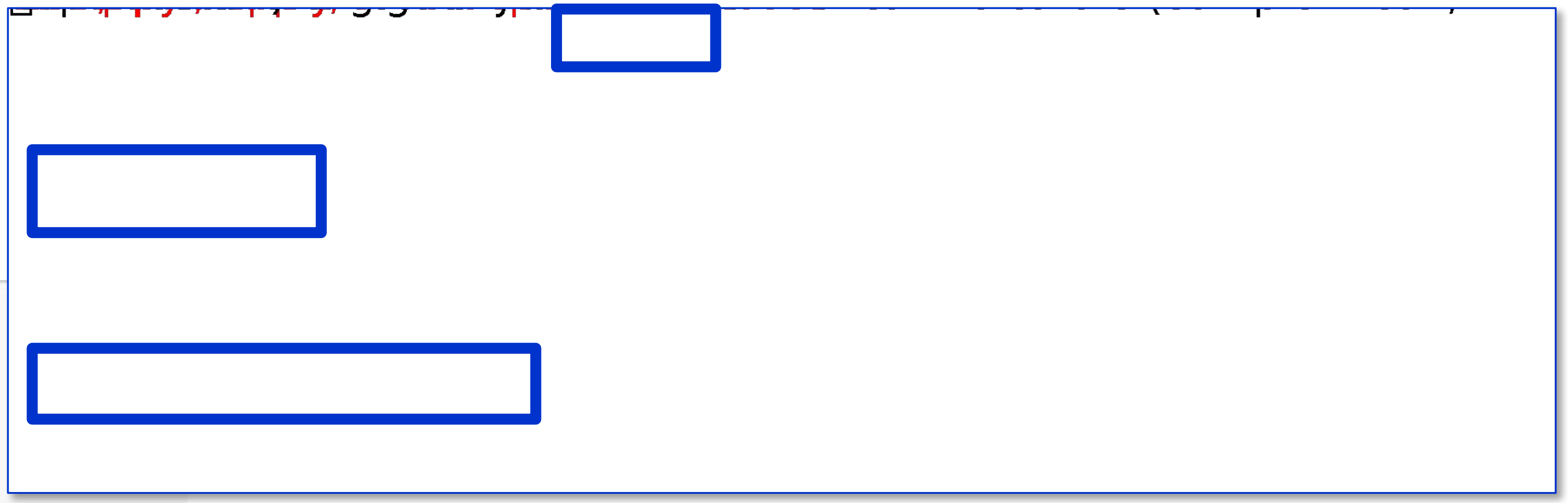
Example①【answer】

Reject H0 under the significance level of 5%. Employ the alternative hypothesis.

That is, “the diameter of the part M is not 1.54cm.”

Copyright © 2021 by INIAD

Example①【answer】: Using Python



```
import numpy as np
from scipy import stats
from scipy.stats import norm

X = np.array([1.53, 1.57, 1.54,1.57, 1.53, 1.55, 1.56, 1.53])
mu_0=1.54
avg = X.mean()
std = 0.01#X.std()
N=X.size
#print sample mean.
print(avg)
z = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(z)
p = norm.cdf(- np.abs(z), 0, 1) * 2
print(p)
```

Copyright © 2021 by INIAD

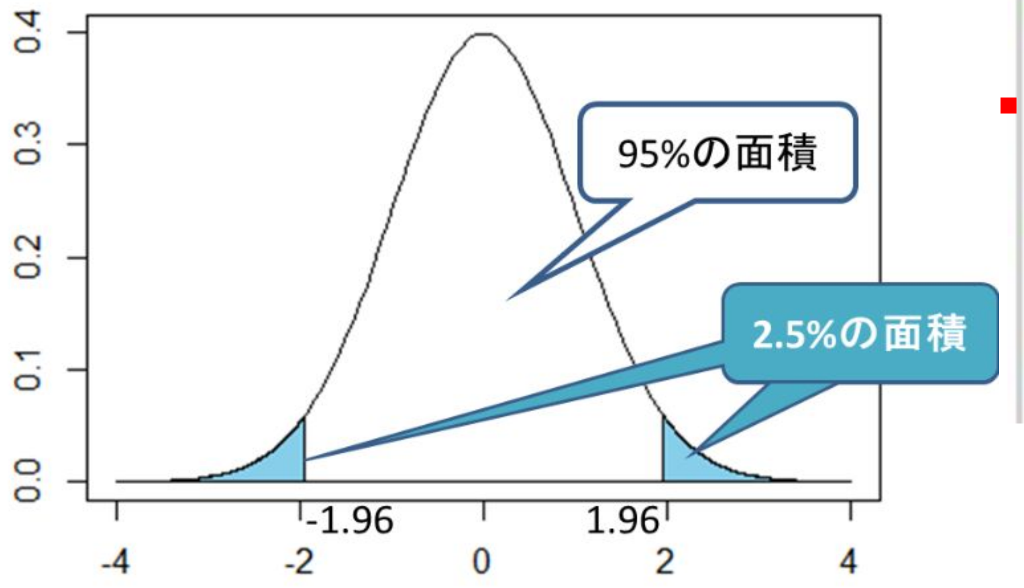
Example①【answer】: Using Python

```
import numpy as np
from scipy import stats
from scipy.stats import norm

X = np.array([1.53, 1.57, 1.54,1.57, 1.53, 1.55, 1.56, 1.53])
mu_0=1.54
avg = X.mean()
std = 0.01#X.std()
N=X.size
#print sample mean.
print(avg)
z = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(z)
p = norm.cdf(- np.abs(z), 0, 1) * 2
print(p)
```

Null hypothesis

Z-statistic

・To find the p-value, you should make it twice
in case of the two-sided test.



Copyright © 2021 by INIAD

Example①【answer】: Using Python

[Output]

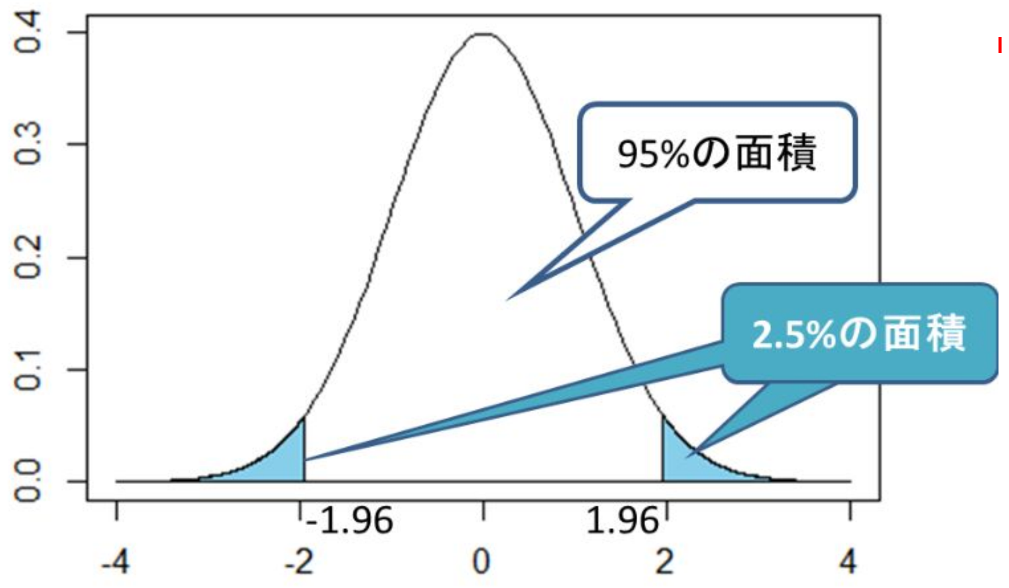
```
import numpy as np
from scipy import stats
from scipy.stats import norm

X = np.array([1.53, 1.57, 1.54,1.57, 1.53, 1.55, 1.56, 1.53])
mu_0=1.54
avg = X.mean()
std = 0.01#X.std()
N=X.size
#print sample mean.
print(avg)
z = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(z)
p = norm.cdf(- np.abs(z), 0, 1) * 2
print(p)
```

```
1.5475
2.1213203435596606
0.033894853524687726
```

Null hypothesis

Z-statistic

P-val.<5%
In the rejection region.



Answer.

“The diameter is not 1.54cm”

or

“Reject H0.”

※Assuming you can answer H0 correctly!

Copyright © 2021 by INIAD

2-3. Hypothesis test for mean (In case population S.D. is unknown)

Copyright © 2021 by INIAD

Test for mean②(population SD is unknown)

- In method ①, the population SD σ was known
  What about σ is unknown 、、？

- You should replace σ by the unbiased SD, S.

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}}$$

Follows the t-distribution of df = (N-1)

Copyright © 2021 by INIAD

Test for mean②(population SD is unknown)

- vi) For a, find a rejection region that meets

  $P(T(X_1,X_2,\ldots,X_N)\in R)=a.$

・For two-sided test, out of upper/ lower 2.5-percentiles(=$\pm t_{N-1}(0.025)$)

・For one-sided test, above the upper 5-percentile (=$t_{N-1}(0.05)$) or below the lower 5-percentile(=$-t_{N-1}(0.05)$)

Copyright © 2021 by INIAD

Example②

In a certain factory, they make a component of a compter, named A. After we measured the length of 10 samples of this A, the mean and unbiased variance were 7.2[cm] and 0.04, respectively.

If the length of A follows the normal distribution, can we state that the length of A is 7.0cm on average? Conduct the hypothesis test with the significance level of 5%.

Copyright © 2021 by INIAD

Flow of hypothesis test

- i) Set the population (☚Similar to confidence interval)
- ii) Set the null hypothesis H_0.
- iii) Extract samples $x_1, x_2, \ldots, x_N$ from the population.
- iv) Find a statistics $T(x_1, x_2, \ldots, x_N)$ from the sample above.
- v) Calculate the probability density of the statistics $T(X_1, X_2, \ldots, X_N)$ for r.v.s $X_1, X_2, \ldots, X_N$.

Copyright © 2021 by INIAD

Flow of hypothesis test

- vi) For a certain significance level a, find a region R, where

  $$P(T(X_1,X_2,\ldots,X_N)\in R)=a$$

  holds(This region R is called as the critical region)

Usually, a=0.01 or 0.05

- vii) If $T(x_1,x_2,\ldots,x_N)\in R$, reject the null hypothesis H_0 / otherwise, H_0 cannot be rejected.

Copyright © 2021 by INIAD

Example②【Answer】

Flow of hypothesis test:

・Null hypothesis H_0:"The length of A is 7.0cm."

Alternative hypothesis H_1:"The length of A is not 7.0cm."

・Find the test statistic $T(x_1,x_2,\ldots,x_N)$.

・Since the population variance is unknown, we take

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}}$$

Copyright © 2021 by INIAD

Example②【Answer】

- Find the p.d.f. of the test statistic t.

Copyright © 2021 by INIAD

Example②【Answer】

- If X follows N (μ, σ^2), then, the test statistic

$$t = \frac{X - \mu}{\frac{s}{\sqrt{n}}}$$

follows the t-distribution with df=(n-1). Here,

$$s = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n-1}}$$

Copyright © 2021 by INIAD

Example②【Answer】

- Find the p.d.f. of the test statistic t.
- ⇒t-distribution with df=9.

Copyright © 2021 by INIAD

Example②【Answer】

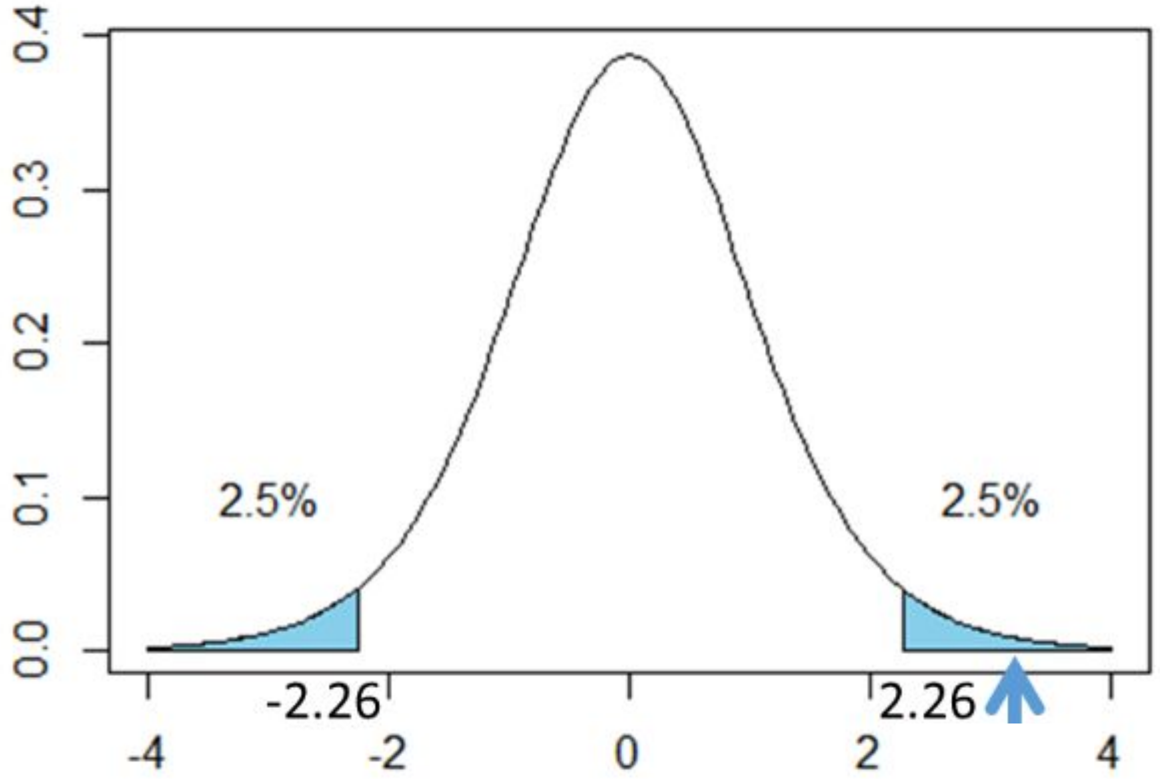
- Determine the rejection region.

  ⇒ Two-sided test.

The rejection region is outside of upper/lower 2.5-percentiles of the t-distribution.

$$t_9(0.05/2) = 2.26$$

- The region $|t| > 2.26$ is the rejection region.

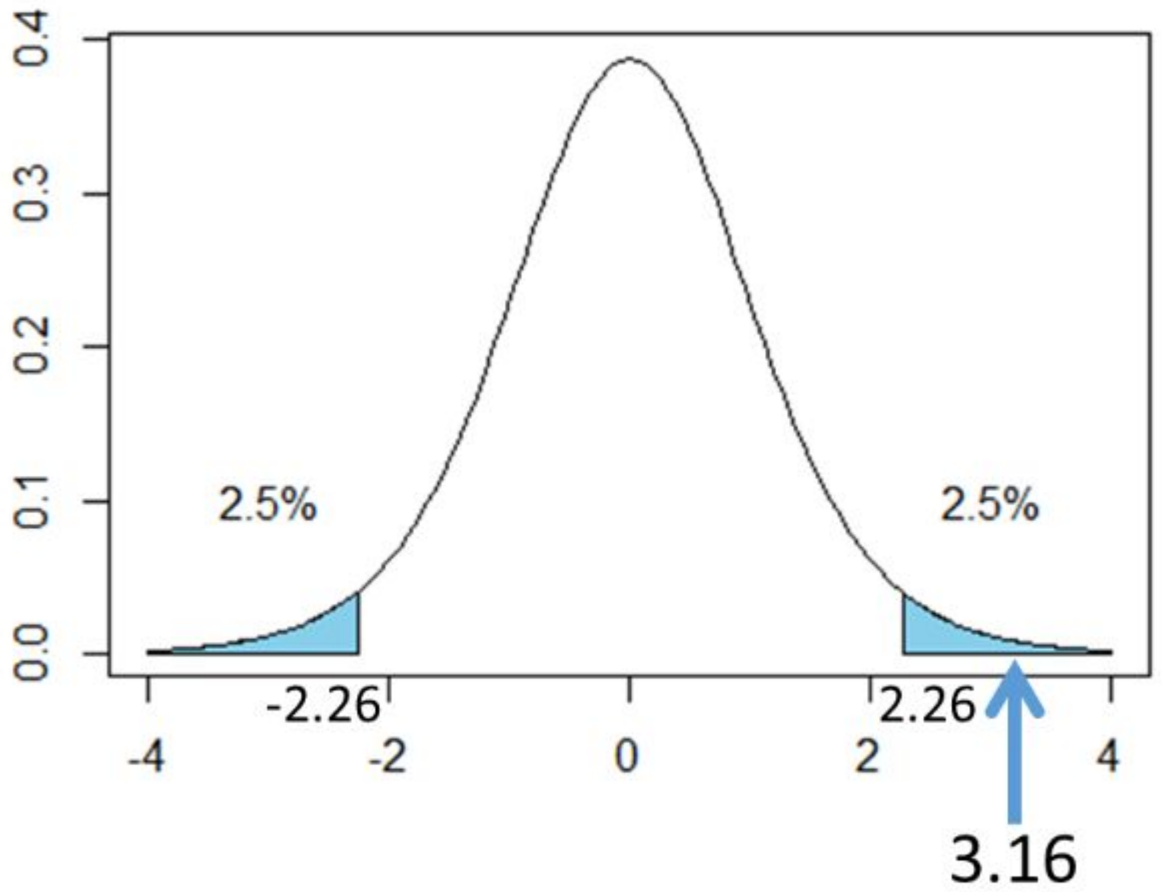


Example②【Answer】

- Find the value of the test statistic $T(x_1, x_2, \ldots, x_N)$:

$$t = \frac{7.2 - 7.0}{\frac{0.2}{\sqrt{10}}} = 3.16$$

- Belongs to the rejection region |t|>2.26 !



Copyright © 2021 by INIAD

Example②【Answer】

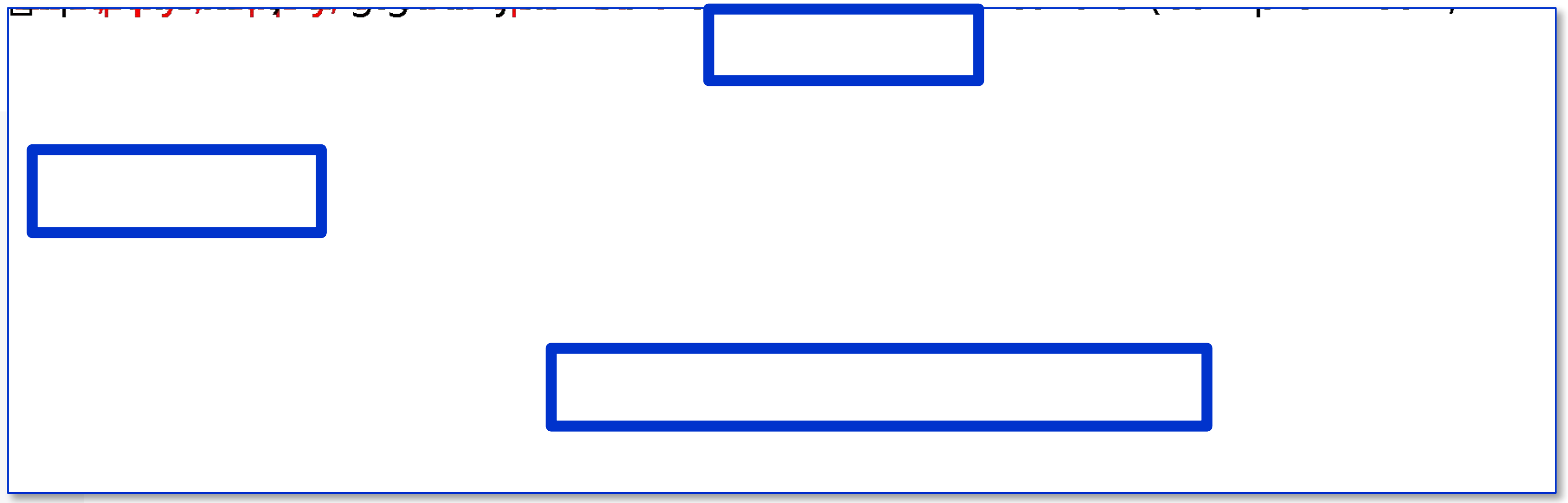
Reject H0 under the significance level of 5%.

“The length of A is not 7.0cm.”

Copyright © 2021 by INIAD

Example②【Answer】Using Python

```
import numpy as np
from scipy import stats
from scipy.stats import t

mu_0=7.0
avg = 7.2
std = 0.2#X.std()
N=10 #X.size
#print sample mean.
stat_t = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(stat_t)
p = t.cdf(-np.abs(stat_t), df=N-1) * 2
print(p)
```



Copyright © 2021 by INIAD

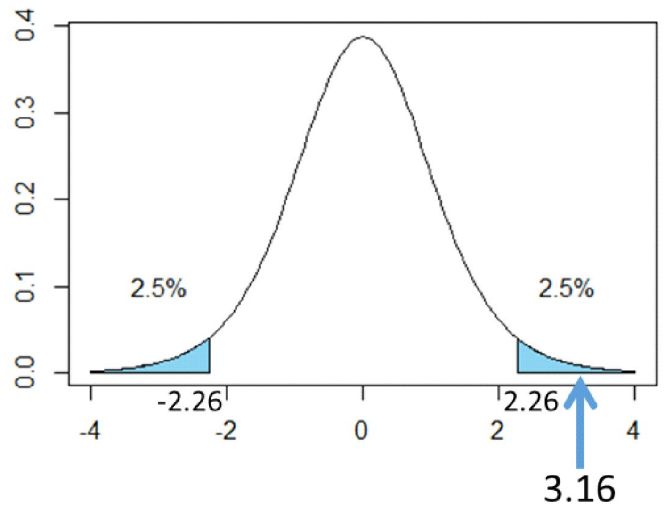
Example②【Answer】Using Python

```
import numpy as np
from scipy import stats
from scipy.stats import t

mu_0=7.0
avg = 7.2
std = 0.2#X.std()
N=10 #X.size
#print sample mean.
stat_t = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(stat_t)
p = t.cdf(-np.abs(stat_t), df=N-1) * 2
print(p)
```

Null hypothesis

Test statistic

・To find the p-value, you should make it twice in case of the two-sided test.



Answer

“The diameter is not 7.0cm /significantly different from 7.0cm”

Or

“Reject H0”

※Assuming you can answer H0 correctly!

Copyright © 2021 by INIAD

Hypothesis test for mean (In case the population S.D. is unknown / one-sided test)

Copyright © 2021 by INIAD

Example ③

A certain maker states that the lifetime of their light bulb is 2000 hours. To validate this statement, we bought 15 samples and tested their lifetime.

Then, the mean and unbiased S.D. were 1900 and 150 hours, respectively.

If the lifetime follows the normal distribution, can we say that the statement of this maker is correct?
Conduct the hypothesis test with the significance level of 5%.

Copyright © 2021 by INIAD

Flow of hypothesis test

- i) Set the population (☚Similar to confidence interval)
- ii) Set the null hypothesis H_0.
- iii) Extract samples $x_1,x_2,\ldots,x_N$ from the population.
- iv) Find a statistics $T(x_1,x_2,\ldots,x_N)$ from the sample above.
- v) Calculate the probability density of the statistics $T(X_1,X_2,\ldots,X_N)$ for r.v.s $X_1,X_2,\ldots,X_N$.

Copyright © 2021 by INIAD

Flow of hypothesis test

- vi) For a certain significance level a, find a region R, where

  $$P(T(X_1,X_2,\ldots,X_N) \in R) = a$$

  holds(This region R is called as the critical region)

Usually, a=0.01 or 0.05

- vii) If $T(x_1,x_2,\ldots,x_N) \in R$, reject the null hypothesis H_0 / otherwise, H_0 cannot be rejected.

Copyright © 2021 by INIAD

Example ③【Answer】

・Null hypothesis H_0:"The lifetime is 2000 hours."

Alternative hypothesis H_1:"The lifetime is less than 2000 hours." (No problem if it's longer)

・Find the test statistic $T(x_1,x_2,\ldots,x_N)$.

・Since the population variance is unknown, we take:

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}}$$

Copyright © 2021 by INIAD

Example ③【Answer】

- Find the p.d.f. of the test statistic t.

Copyright © 2021 by INIAD

Example ③【Answer】

- If X follows N (μ, σ²), then the test statistic

$$t = \frac{X - \mu}{\frac{s}{\sqrt{n}}}$$

follows the t-distribution with df=(n-1). Here,

$$s = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n-1}}$$

Copyright © 2021 by INIAD

Example ③【Answer】

- Find the pd.f. of the test statistic t.

  ⇒ follows the t-distribution with df= 14.

Copyright © 2021 by INIAD

Example ③【Answer】

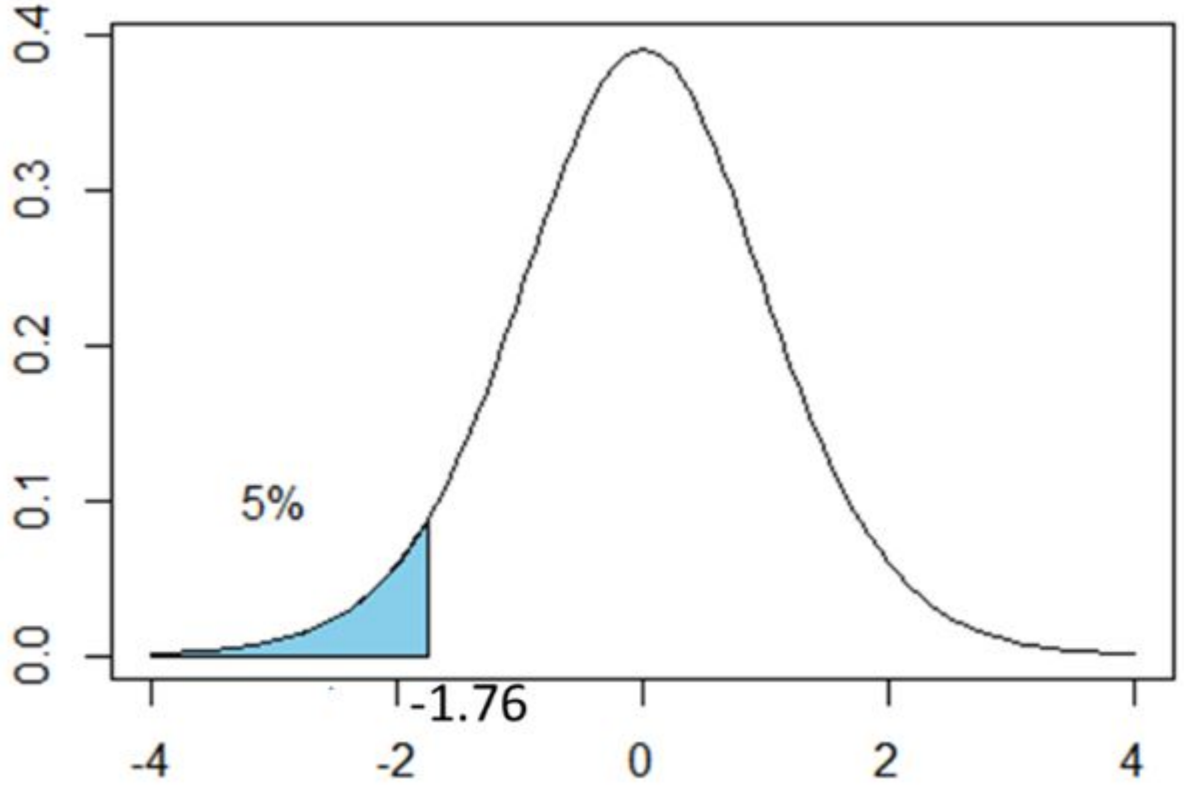
- Determine the rejection region.

  ⇒one-sided test.

The rejection region is left side of the lower 5-percentile of the ti-distribution.

$$t_{14}(0.05) = -1.76$$

- The region of t<-1.76 is the rejection region.



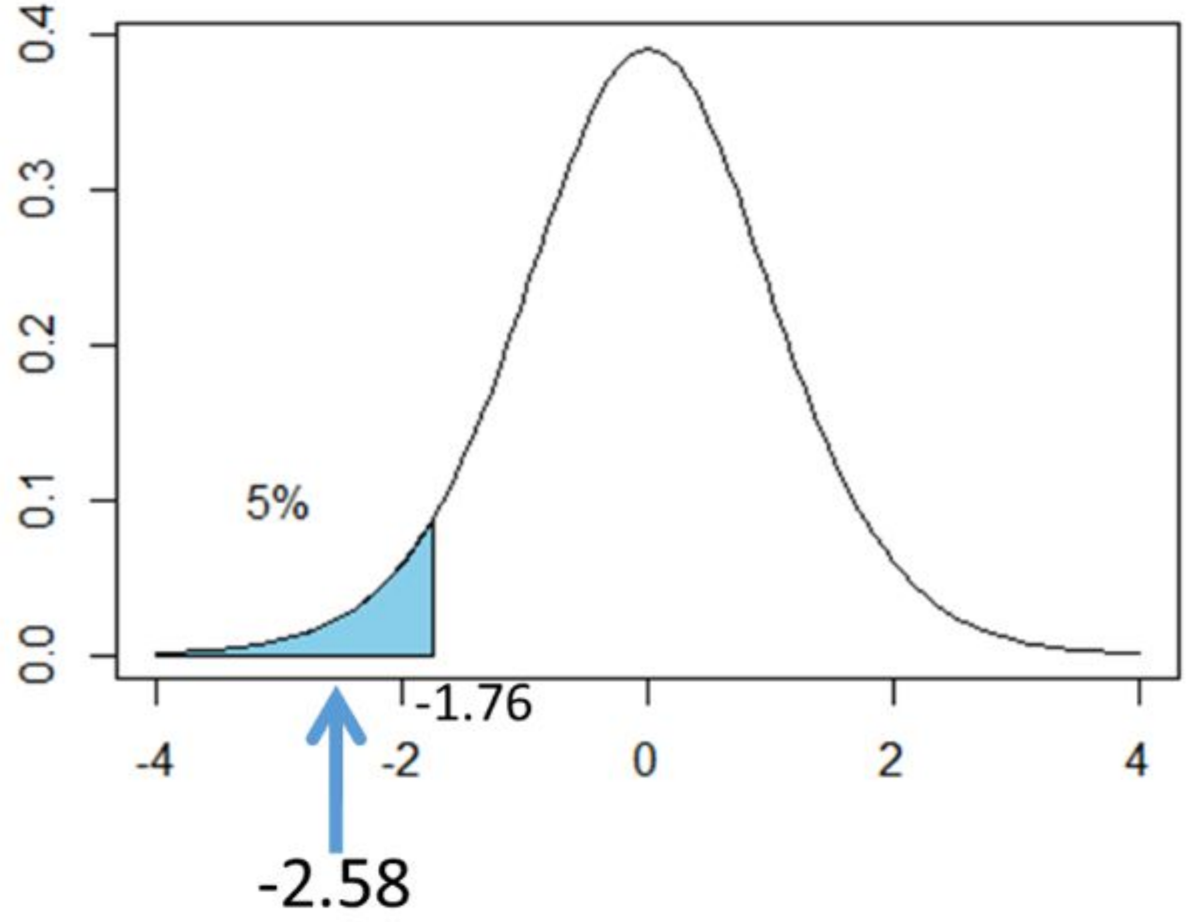
Copyright © 2021 by INIAD

Example ③【Answer】

- Find the value of the test statistic $T(x_1, x_2, \ldots, x_N)$:

$$t = \frac{1900 - 2000}{\frac{150}{\sqrt{15}}} = -2.58$$

- Belong to the rejection region t<-1.76 !



Copyright © 2021 by INIAD

Example ③【Answer】

We reject H0 under the significance level of 5%.

Thus, "The lifetime of ther light bulb is shorter than 2000 hours. Their statement should be corrected."

Copyright © 2021 by INIAD

Copyright © 2021 by INIAD

Example ③【Answer】Using Python

```
import numpy as np
from scipy import stats
from scipy.stats import t

mu_0=2000
avg = 1900
std = 150 #X.std()
N=15 #X.size
#print sample mean.
stat_t = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(stat_t)
p = t.cdf(-np.abs(stat_t), df=N-1)
print(p)
```

Null hypothesis

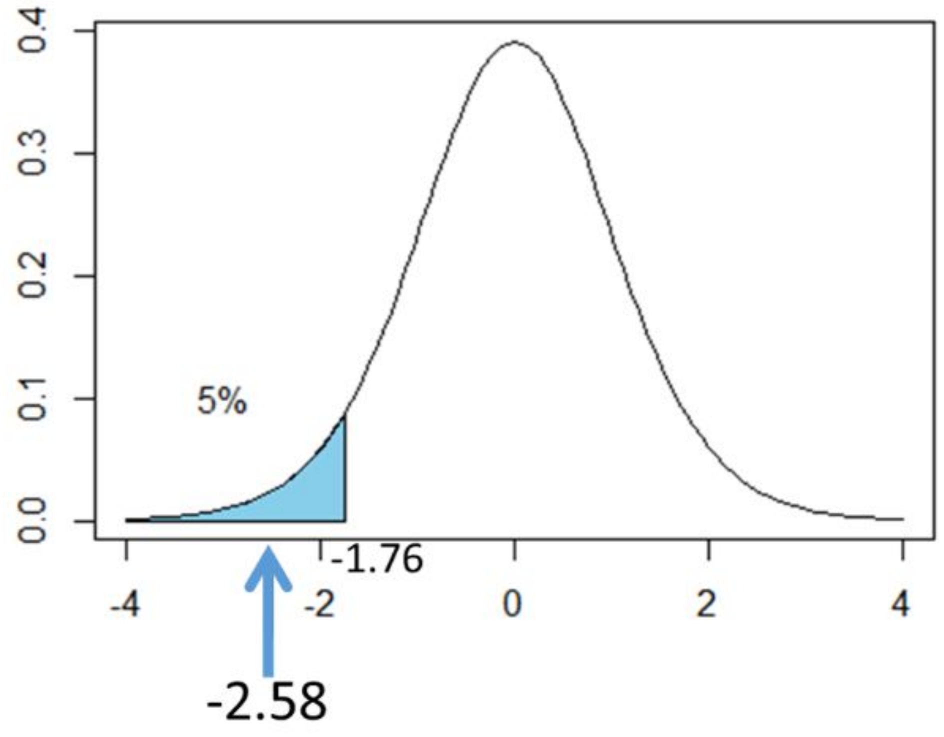
Test statistic

・Find the p-value. One-sided test this case!
(No need to make twice).

Example ③【Answer】Using Python

[出力]

```
import numpy as np
from scipy import stats
from scipy.stats import t

mu_0=2000
avg = 1900
std = 150 #X.std()
N=15 #X.size
#print sample mean.
stat_t = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(stat_t)
p = t.cdf(-np.abs(stat_t), df=N-1)
print(p)
```




```
-2.581988897471611
0.010862621599030588
```

P-value < 5%
In the rejection region.

Copyright © 2021 by INIAD

Cautions in one-sided test

- In case of the one-sided test, if the sample mean belongs to the reverse side of the null hypothesis, we can promptly stop test by concluding “we cannot reject H_0”.
- ※In case of the prevous example, if the samples mean is less than 2000, we can promptly terminate the test.

Copyright © 2021 by INIAD

Exercises

Copyright © 2021 by INIAD

Exercise①

In a certain farm, they developed a new fertilizer. Then, after we measured the their yield in 6 farms where they used this new fertilizer, we obtained the following data of the yield per unit area.

42.9 43.7 43.2 40.8 42.8 44.2   [kg]

The average yield with the conventional fertilizer per unit area was 41.4kg. Then, can we say that this new fertilizer improves the yield? Do the hypothesis test with known S.D. σ=3.5[kg] under the significance level of 5%.

Copyright © 2021 by INIAD

Exercise①【Answer】

Let us denote the average yield with the new fertilizer as μ.

Null hypothesis H_0:"μ=41.4[kg]"

Alternative hypothesis H_1:" μ>41.4[kg]"

Since the population variance is known, we take:

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Copyright © 2021 by INIAD

Exercise①【Answer】

We have: $\bar{x} = 42.93$

Thus,

$$t = \frac{42.93 - 41.4}{\frac{3.5}{\sqrt{6}}} = 1.071$$

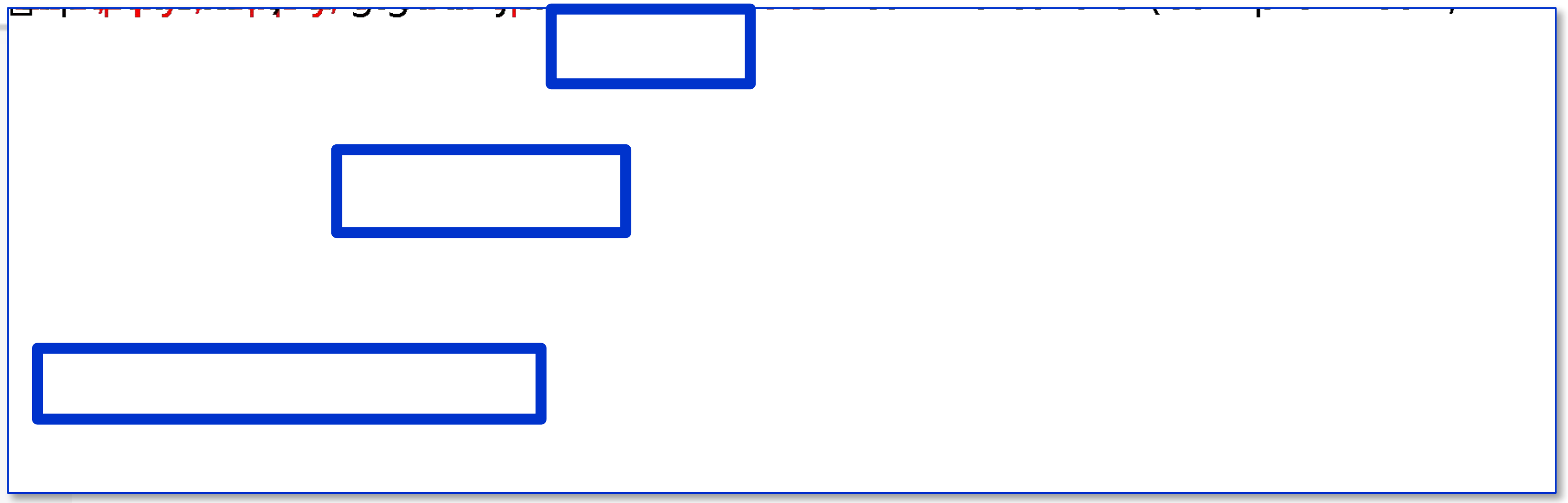
Does not satisfy t>1.64. It does not belong to the significance level. So we can't reject H_0.

⇒ "We can't say that the new fertilizer has a significant effect"

Copyright © 2021 by INIAD

Exercise①【Answer】Using Python

```
#Exercise 4.
import numpy as np
from scipy import stats
from scipy.stats import norm

X = np.array([42.9, 43.7, 43.2, 40.8, 42.8, 44.2])
mu_0=41.4
avg = X.mean()
std = 3.5 #X.std()
N=X.size
#print sample mean.
print(avg)
z = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(z)
p = norm.cdf( -np.abs(z), 0, 1)
print(p)
```



Copyright © 2021 by INIAD

Exercise①【Answer】Using Python

```
#Exercise 4.
import numpy as np
from scipy import stats
from scipy.stats import norm

X = np.array([42.9, 43.7, 43.2, 40.8, 42.8, 44.2])
mu_0=41.4
avg = X.mean()
std = 3.5 #X.std()
N=X.size
#print sample mean.
print(avg)
z = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(z)
p = norm.cdf( -np.abs(z), 0, 1)
print(p)
```

・Find the p-value. One-sided test this case!
(No need to make twice).

Copyright © 2021 by INIAD

Exercise①【Answer】Using Python

[Output]

```
#Exercise 4.
import numpy as np
from scipy import stats
from scipy.stats import norm

X = np.array([42.9, 43.7, 43.2, 40.8, 42.8, 44.2])
mu_0=41.4
avg = X.mean()
std = 3.5 #X.std()
N=X.size
#print sample mean.
print(avg)
z = (avg -mu_0)/ (std / np.sqrt(N))
#print z-value.
print(z)
p = norm.cdf( -np.abs(z), 0, 1)
print(p)
```

```
42.93333333333334
1.073109792076434
0.14161092902072725
```

Copyright © 2021 by INIAD

Exercise②

In a certain maker of a part (named as “M”) of computer, its diameter is described as 1.54[cm] in its product specification. In a certain sample survey, they extracted 8 samples randomly, and observed the following data of measured diameter [cm]:

1.53 1.57 1.54 1.57 1.53 1.55 1.56 1.53

The population variance is unknown. Then, can you say that this part follow its product specification? Test with the significance level of 5%.

Copyright © 2021 by INIAD

Exercise② 【Answer】

Null hypothesis H_0:"μ=1.54[cm]"

Alternative hypothesis H_1:" μ≠1.54[cm]"

Since the population variance is known, we take:

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}}$$

Copyright © 2021 by INIAD

Exercise② 【Answer】

We have $\bar{x} = 1.5475$, and the unbiased S.D. is

$$S^2 = \frac{(1.53 - 1.5475)^2 + (1.57 - 1.5475)^2 + \ldots + (1.53 - 1.5475)^2}{8 - 1} = 0.000307$$

$$t = \frac{1.5475 - 1.54}{\frac{0.0175}{\sqrt{8}}} = 1.212$$

The upper 2.5-percentile of the t-distribution with df=(8-1=7) is

$$t_7(\frac{0.05}{2}) = 2.365$$

Since |t| < 2.365, it doesn't belong to the rejection region. We cannot reject H0.

"We cannot recognize the different under the significance level of 5%."

Copyright © 2021 by INIAD

Exercise② 【Answer】Using Python

```
#Exercise 5.
import numpy as np
from scipy import stats
from scipy.stats import t

X = np.array([1.53, 1.57, 1.54,1.57, 1.53, 1.55, 1.56, 1.53])
mu_0=1.54
avg = X.mean()
std = np.std(X, ddof=1)
N=X.size
#print sample mean.
#print(avg)
stats_t = (avg -mu_0)/ (std / np.sqrt(N))
#print stats_t-value.
print(stats_t)
p = t.cdf( -np.abs(stats_t), df=N-1) * 2
print(p)
```

To find test statistic, you should make it twice in case of the two-sided test.

Copyright © 2021 by INIAD

Exercise②【Answer】Using Python

[Output]

```
#Exercise 5.
import numpy as np
from scipy import stats
from scipy.stats import t

X = np.array([1.53, 1.57, 1.54,1.57, 1.53, 1.55, 1.56, 1.53])
mu_0=1.54
avg = X.mean()
std = np.std(X, ddof=1)
N=X.size
#print sample mean.
#print(avg)
stats_t = (avg -mu_0)/ (std / np.sqrt(N))
#print stats_t-value.
print(stats_t)
p = t.cdf( -np.abs(stats_t), df=N-1) * 2
print(p)
```

```
1.2104198771789023
0.2653980394260665
```

P-value >5%
Cannot reject the null hypothesis.

Copyright © 2021 by INIAD

Exercise③

A catalog of a climbing shop I states that the breaking strength of their 15mm rope is 4500kg.
Now, after we conducted the sample measurement of its strength by using 50 samples, the mean and unbiased S.D. were 4450[kg] and 120[kg], respectively.

Then, can we state that the stated strength 4500[kg] is satisfied on average? Conduct the hypothesis test with the significance level of 5%.

Copyright © 2021 by INIAD

Exercise③【Answer】

Let μ denote its mean.

Null hypothesis H_0:"μ=4500"

Alternative hypothesis H_1:"μ<4500" (No problem in case it's larger.)

Since the population variance is unknown, we take:

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}}$$

Copyright © 2021 by INIAD

Exercise③【Answer】

Sample mean and unbiased S.D. were 4450 and 120, respectively. So,

$$t = \frac{4450 - 4500}{\frac{120}{\sqrt{50}}} = -2.946$$

This follows the t-distribution with df= 50-1=49. But its lower 5-percentile is

$-t_{49}(0.05)$ =-1.68.

Now, since t < -1.68, it belongs to the regnificance region, and H0 is rejected.

We can say "The mean of samples are different from the one stated in their catalog under the significance level of 5%."

Copyright © 2021 by INIAD

Exercise③【Answer】Using Python

```python
#Exercise 6.
import numpy as np
from scipy import stats
from scipy.stats import t

#X = np.array([1.53,  1.57,  1.54, 1.57,  1.53,  1.55,  1.56,  1.53])
mu_0=4500
avg =4450# X.mean()
std = 120#np.std(X,  ddof=1)
N=50 #X.size
stats_t = (avg -mu_0)/ (std / np.sqrt(N))
#print stats_t-value.
print(stats_t)
p = t.cdf( -np.abs(stats_t), df=N-1)
print(p)
```

```
-2.946278254943948
0.0024555744280253798
```

・Find the p-value. One-sided test this case!
(No need to make twice).

Copyright © 2021 by INIAD

Exercise③【Answer】Using Python

[Output]

```
#Exercise 6.
import numpy as np
from scipy import stats
from scipy.stats import t

#X = np.array([1.53, 1.57, 1.54, 1.57, 1.53, 1.55, 1.56, 1.53])
mu_0=4500
avg =4450# X.mean()
std = 120#np.std(X, ddof=1)
N=50 #X.size
stats_t = (avg -mu_0)/ (std / np.sqrt(N))
#print stats_t-value.
print(stats_t)
p = t.cdf( -np.abs(stats_t), df=N-1)
print(p)
```

```
-2.946278254943948
0.0024555744280253798
```

P-value<5%
Reject the null hypothesis.

Copyright © 2021 by INIAD

Exercise④

For a certain product, we measured its diameter of 5 samples, and observed:
36.3, 35.7, 35.9, 37.1, 36.1 [mm].

The spec of this product states that it should be 35.5m.
Then, check if we can state that the actual diameter of this product is significantly larger than the spec or not.

Assume the normality of population and apply the hypothesis test for mean under the significance level of 5%.

Copyright © 2021 by INIAD

演習④【解答】

```
import numpy as np
from scipy import stats
from scipy.stats import t

X=np.array([36.3, 35.7, 35.9, 37.1, 36.1])
mu_0=35.5
X_mean=X.mean()
X_sd=np.std(X,ddof=1)
N=X.size

stats_t = (X_mean -mu_0)/ (X_sd / np.sqrt(N))
p_val = t.cdf( -np.abs(stats_t),df=N-1)

print("p-value is")
print(p_val)

if p_val<0.05:
    print("帰無仮説棄却")
else:
    print("帰無仮説棄却できない")
```

```
p-value is
0.020381913442855257
帰無仮説棄却
```

Copyright © 2021 by INIAD