2019 INIAD StatI
Week15
「Hypothesis test (3)」

2019 23rd, July

Hirotada Honda

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: Review and exam(ii)
Week12: From descriptive statistics to inferential statistics -z-table and confidwncw interval-
Week13: Hypothesis test(1) -Introduction, and distributions of test statistic (t-distribution)-
Week14: Hypothesis test(2) -Test for mean-
Week15: Hypothesis test(3) -Test for difference of mean-
Week16: Review and exam(3)

Summary so far ：
Review on hypothesis test for mean

What is ‘significant difference (i.e., rejecting H0)’?

What is ‘significant difference (i.e., rejecting H0)’?

Not within the range of “noise”

What is ‘significant difference (i.e., rejecting H0)’?

Not within the range of “noise”

Obviously different.

What is ‘significant difference (i.e., rejecting H0)’?

H0 “$\mu=\mu_0$” seems obviously false, and is rejected.

What is ‘significant difference (i.e., rejecting H0)’?

H0 “$\mu=\mu_0$” seems obviously false, and is rejected.

Two-sided:

What is ‘significant difference (i.e., rejecting H0)’?

H0 “$\mu = \mu_0$” seems obviously false, and is rejected.

Two-sided:

H1: $\mu \neq \mu_0$

is employed.

What is ‘significant difference (i.e., rejecting H0)’?

H0 “$\mu=\mu_0$” seems obviously false, and is rejected.

One-sided:

What is ‘significant difference (i.e., rejecting H0)’?

H0 “$\mu=\mu_0$” seems obviously false, and is rejected.

One-sided:

H1: $\mu > \mu_0$ or

$\mu < \mu_0$

is employed.

Example①

Mr.A, a junior high-school student, is concerned whether he becomes taller than the last year. The record of the last year was 165.4cm.

The measured value of this year was 165.6cm.

Example ①

Mr.A, a junior high-school student, is concerned whether he becomes taller than the last year. The record of the last year was 165.4cm.

The measured value of this year was 165.6cm.

Is this within the rage of noise?
Or obviously （i.e., significantly）
he is taller than the last year?

Example ①

Mr.A, a junior high-school student, is concerned whether he becomes taller than the last year. The record of the last year was 165.4cm.

The measured value of this year was 165.6cm.

Vote your opinion：
- Within the noise
- Significantly taller

Example ①

With the terms of hypothesis test now.

Example ①

Population： Possible measured values of the height of this year
（it’s a r.v. Its set is population.）

Now, suppose that the actual height μ is the population mean.

Sample set： The measured values of N times measurements.

Example ①

H0 : $\mu = 165.4$

That is, “Not changed from the last year.”

Example ①

H0 : $\mu = 165.4$

H1 : $\mu > 165.4$

That is, ‘Taller than the last year.’ （He is interested in one-direction: taller or not.）

Example ①

H0 :

$$\mu = 165.4$$

H1 :

$$\mu > 165.4$$

That is, ‘Taller than the last year.’ （He is interested in one-direction: taller or not.）

One-sided !

One population so far.

How about the case of 2 populations ？

How about the case of 2 populations？

Example②： Students in junior-highschools in Japan and US.
Different averages of height？

Example ②

Students in Japan and US. Samples are:

Jap. :   165.2, 165.4, 166.7, 168.1 [cm]

US :   166.2, 168.4, 170.3 [cm]

Example ②

Students in Japan and US. Samples are:

Jap. : 165.2, 165.4, 166.7, 168.1 [cm]

US : 166.2, 168.4, 170.3 [cm]

The diff. are within the range of noise?
Or, seems significantly different?

Example ②

Students in Japan and US. Samples are:

Jap. : 165.2, 165.4, 166.7, 168.1 [cm]

US : 166.2, 168.4, 170.3 [cm]

Vote your opinion :
- Within the noise.
- Significantly different.

Example ②

With the terms of hypothesis test now.

Example ②

Population :
“Japanese students” と “US students”

Population means: actual average height,
μ1 and μ2, resp.

Sample sets :
Measured height of Japanese and US students, N_1 and N_2, resp.

Example ②

H0 : $\mu_1 = \mu_2$

"The average height of Japanese students is the same as that of US students"

Example ②

H0 : $\mu_1 = \mu_2$

H1 : $\mu_1 \neq \mu_2$

“There is a diff. between the Japanese and US students”

Example ②

H0 : $\mu_1 = \mu_2$

H1 : $\mu_1 \neq \mu_2$

“There is a diff. between the Japanese and US students”

Two-sided !

So far, we have seen the ‘test for the diff. of mean’ in case of two populations.

【Ref.】 How about the case of 3 or more populations？

【Ref.】 ANOVA （Autumn semester）

H0： $\mu_1 = \mu_2 = \mu_3$

【Ref.】 ANOVA （Autumn semester）

H0： $$\mu_1 = \mu_2 = \mu_3$$

「The mean of 3 populations are the same.」

1-1. Two-sample test for mean

Two-sample test for mean

Assume there are two independent samples that follow the normal distribution.

When we extract samples from them, two-sample t-test enables you to check that if there is a significant difference between the mean of these two populations.

Ex) Is there any difference between the scores of 2 classes?

Is the new medicine actually effective?

Two-sample test for mean

Two ways depending on:

- There is a one-to-one relationship between 2 samples;

  ⇒ Paired t-test

-The 2-samples are independent.

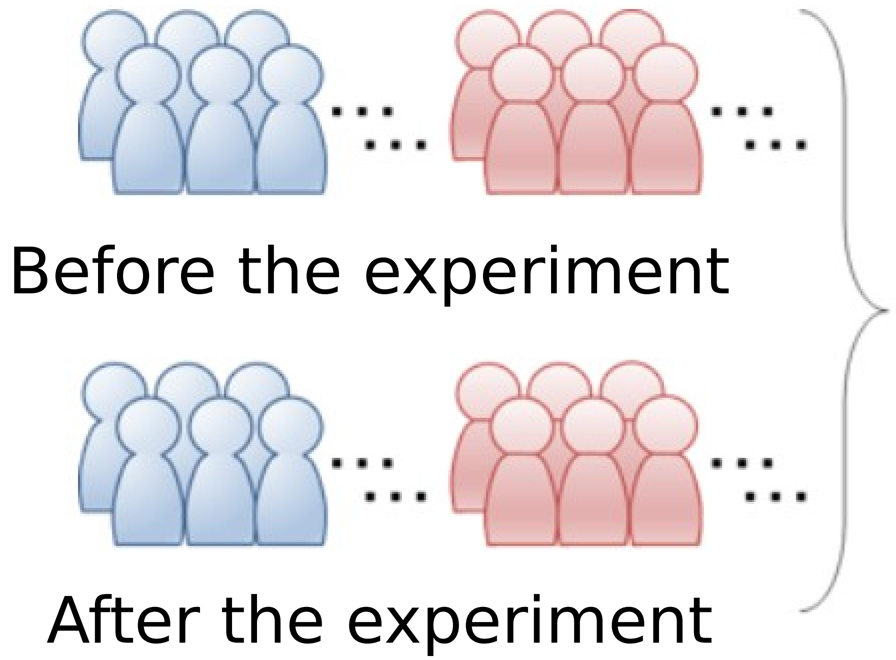
  ⇒ Independent sample t-test

Paired t-test

Paired test

The 2 samples are extracted from one population.

Ex)A test of the new medicine for the blood pressure.
After providing the medicine to subjects for a while,
you measure their blood pressure, and compare the values with those before providing the medicine.



Samples are same people.
→”Paired test”

Paired t-test

By taking the differences of 2 samples, it’s reduced to the usual t-test.

→ You can apply the usual t-test (already learned it).

Null hypothesis H_0:”No difference; the difference $\mu=0$”

Example④

We validate a new medicine that reduces the blood pressure. The measured data before and after providing it so 5 subjects are shown below.

Is this medicine actually effective? Do the hypothesis testing with the significance level of 5%.

Subject ID	Before[mmHg]	After[mmHg]
1	180	150
2	130	135
3	165	145
4	155	150
5	140	140

Example④ 【Answer】

In the paired t-test, you should validate whether the difference of 2 samples vanishes or not. In this case, we focus on the difference of the bllod pressure before and after the experiment.

First, take the difference and find the average of the difference.

Example④ 【Answer】

Difference of before and after.

Subject ID	Before[mmHg]	After[mmHg]	Difference [before-after]
1	180	150	30
2	130	135	-5
3	165	145	20
4	155	150	5
5	140	140	0

Regard this column as a sample X.

Example④ 【Answer】

Null hypothesis H0:”The difference μ=0”
(i.e., this medicine is not effective.)

Alternative hypothesis H1:”μ>0”
→ One-sided test!

Now, define the test statistic:

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

Example④ 【Answer】

Find the p.d.f of the test statistic:

Example④ 【Answer】

Find the p.d.f of the test statistic:

→ Follows the t-distribution with df = (5-1) = 4.

Example④ 【Answer】

Define the rejection region.
Since this is the one-sided test, it is out of the upper 5-percentile of the t-distribution.
※) Note that we are interested in whether it's positive or not.

By noting

$$t_4(0.05) = 2.132$$

$t > 2.132$ is the rejection region.

Example④ 【Answer】

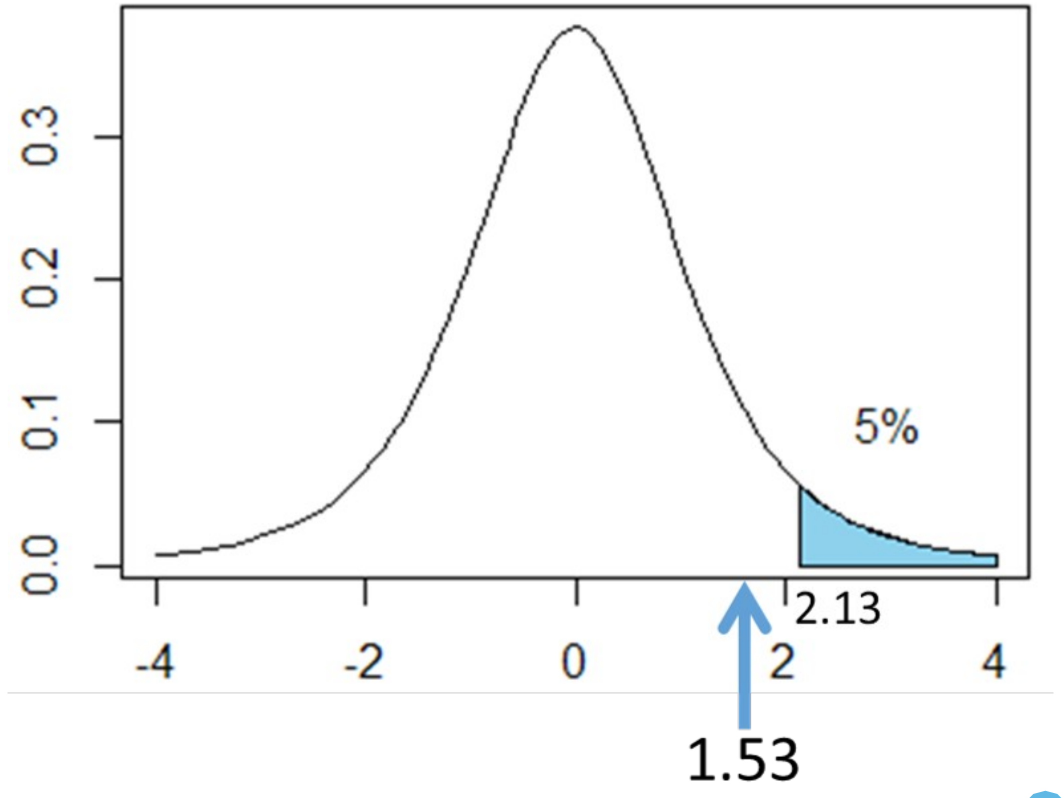
・Find the value of test statistics $T(x_1,x_2,\ldots,x_N)$ on the basis of H0. In this case, in the definition of

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

μ=0. We also have

$$s^2 = \frac{1}{5-1} \times \{(30-10)^2 + (-5-10)^2 + (20-10)^2 + (5-10)^2 + (0-10)^2\} = 212.5$$

$$t = \frac{10-0}{\sqrt{\frac{212.5}{5}}} \fallingdotseq 1.53$$



Not in the rejection region!

Example④ 【Answer】

We cannot reject H0 under the significance level of 5%.

“We cannot state that this medicine actually reduces the blood pressure.”

Example④ 【Answer】

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

X = np.array([30, -5, 20, 5, 0])
mu_0=0
avg = X.mean()
std = np.std(X, ddof=1)

N=X.size
stats_t = (avg -mu_0)/ (std / np.sqrt(N))
#print t-value
print(stats_t)
p = t.cdf( -np.abs(stats_t),df=N-1)
print(p)
```

```
1.533929977694741
0.09991459276886
```

Mean and unbiased S.D.

Test statistic

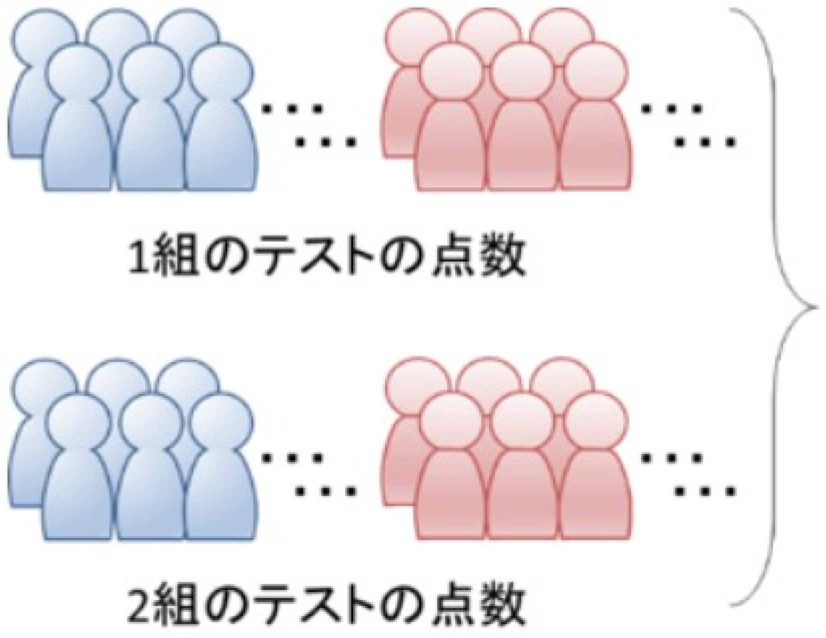
P-value.
Not made twice.

P-value >5 %
Not reject H0.

Independent samples t-test

Independent samples t-test

No mapping between 2 samples.
Ex) Scores of an exam in classes A and B.
Different students in classes A and B.



Independent samples
→” Independent samples t-test”

Independent samples t-test

We further classify it into 3 cases;

i) Population variances σ_1^2 , σ_2^2 are both known;

ii) Population variances σ_1^2 , σ_2^2 are unknown, but we can assume the equality: $\sigma_1^2 = \sigma_2^2$;

iii) Population variances σ_1^2 , σ_2^2 are unknown.

Independent sample t-test

We further classify it into 3 cases;

i) Population variances σ_1^2 , σ_2^2 are both known;

ii) Population variances σ_1^2 , σ_2^2

Not popular recently.

we can assume the equality: $\sigma_1^2 = \sigma_2^2$;

iii) Population variances σ_1^2 , σ_2^2 are unknown.

Independent sample t-test

In this course, we recommend iii) except for the cases when both variances are known. But we just refer to ii) also.

i) Population variances σ_1^2 , σ_2^2 are both known;

ii) Population variances σ_1^2 , σ_2^2 are unknown, but we can assume the equality: $\sigma_1^2 = \sigma_2^2$;

iii) Population variances σ_1^2 , σ_2^2 are unknown.

i) Population variances σ_1^2 , σ_2^2 are both known;

Independent sample t-test

i) Population variances σ_1^2 , σ_2^2 are both known;

Let us denote:

i) The sample mean, population mean, population variance and sample size of the first sample are

$$\bar{x}_1,\ \mu_1,\ \sigma_1^2,\ n_1.$$

ii) The sample mean, population mean, population variance and sample size of the second sample are

$$\bar{x}_2,\ \mu_2,\ \sigma_2^2,\ n_2.$$

Define the test statistic: $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Follows z-distribution!

Example⑤

The table below shows the data of the measured total cholesterol value [mg/dl] of i) 15 patients of alcoholic fatty liver, and ii) 20 healthy people.

We know that the population variances of i) and ii) are $\sigma_1^2 = 30.3^2$ and $\sigma_2^2 = 25.8^2$, respectively.

Then, is there any significant difference between these samples?

Do the hypothesis testing with the significance level of 5%.

Alcoholic patients	Healthy people
253 224 258 246 294	194 154 176 176 180
265 242 300 276 262	160 174 193 207 170
233 212 273 228 308	161 206 184 182 165
	172 184 205 176 180

Example⑤ 【Answer】

Null hypothesis H0:"There is no difference."
Alternative hypothesis H1:"There is some difference."

Define the test statistics. Since we know the population variances and

$$\bar{x}_1 = 258.3,\ \bar{x}_2 = 179.8$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{258.3 - 179.8}{\sqrt{\frac{30.3^2}{15} + \frac{25.8^2}{20}}} = 8.076$$

Example⑤ 【Answer】

Find the p.d.f of the test statistic:

Example⑤ 【Answer】

Find the p.d.f of the test statistic:

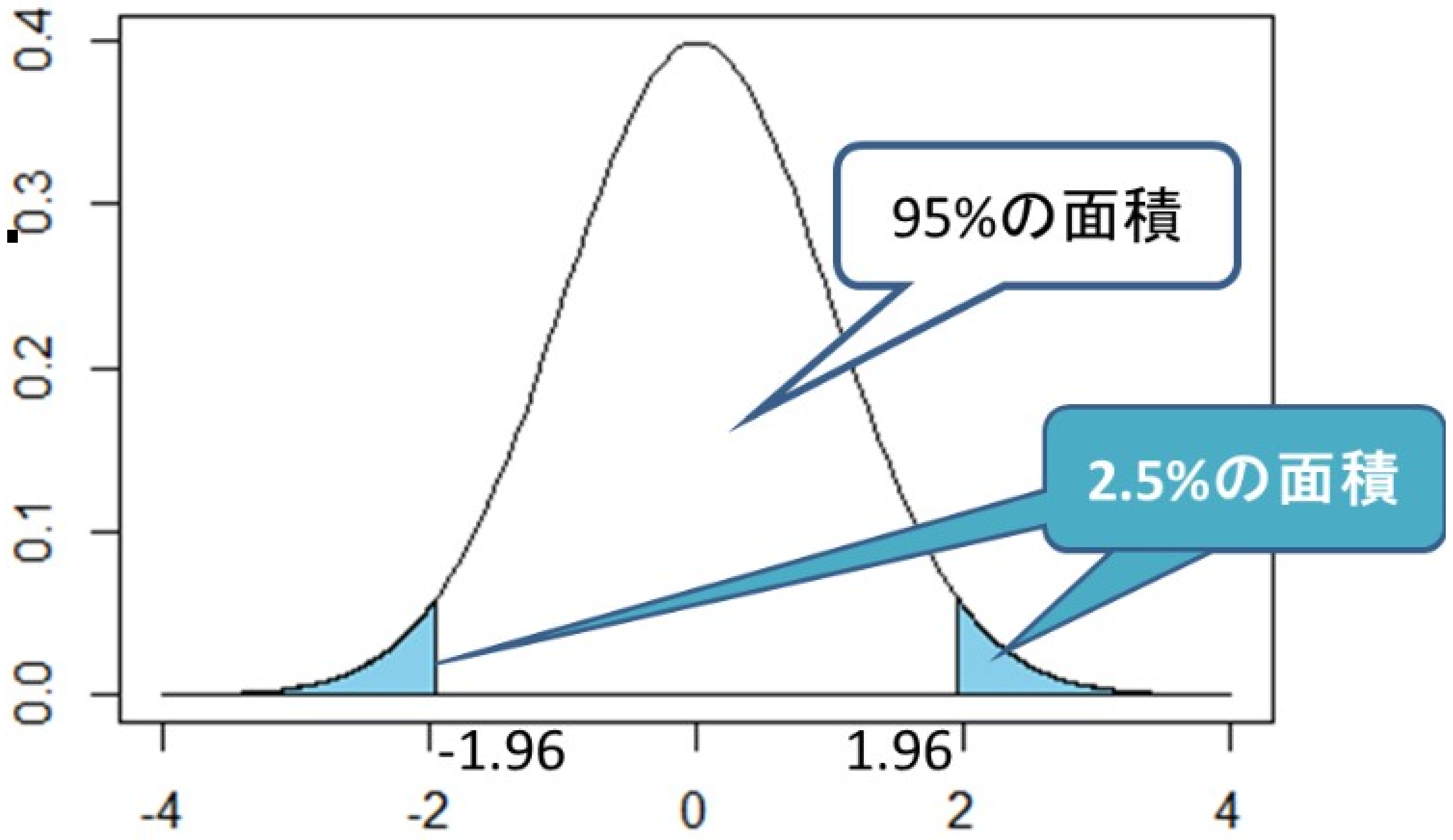
→ Follows the z-distribution

Example⑤ 【Answer】

Find the rejection region.
Two-sided test in this case.

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

$|t| > 1.96$ is the rejection region.



Example⑤ 【Answer】

- Find the value of test statistics $T(x_1,x_2,...,x_N)$ on the basis of H0. In this case,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{258.3 - 179.8}{\sqrt{\frac{30.3^2}{15} + \frac{25.8^2}{20}}} = 8.076$$

is in the rejection region.

Example⑤ 【Answer】

We reject the null hypothesis H0 under the significance level of 5%.

“There is a significant difference between the 2 samples.”

Example⑤ 【Answer】

```
#Example 8.
import numpy as np
from scipy import stats
from scipy.stats import norm

X1 = np.array([253, 224, 258, 246, 294, 265, 242, 300, 276, 262, 233, 212, 273, 228, 308])
X2 = np.array([194, 154, 176, 176, 180, 160, 174, 193, 207, 170, 158, 206, 184, 182, 165, 172, 184, 205, 176, 180])


avg_1 = X1.mean()
avg_2 = X2.mean()
std_1 = 30.3
std_2 = 25.8
N1=X1.size
N2=X2.size
#print sample mean.
stat_t = (avg_1 - avg_2)/  np.sqrt(std_1*std_1/N1 + std_2*std_2/N2)
#print z-value.
print(stat_t)
p =  norm.cdf(-np.abs(stat_t),0,1)*2
print(p)
```

```
8.072292227445995
6.899050964791441e-16
```

- p-value<5%

Reject H0

ii) Population variances σ_1^2, σ_2^2 are unknown, but we can assume the equality: $\sigma_1^2 = \sigma_2^2$;

※Actually, we should test whether the equivalence of variance holds or not.
（test for equality of variance）

Out of coverage.

Independent samples t-test

ii) Population variances σ_1^2 , σ_2^2 are unknown, but we can assume the equality: $\sigma_1^2 = \sigma_2^2$;

Define the test statistic by

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Follows the t-distribution of df=(n1+n2-2)

Independent sample t-test

Here,

$$s^2 = \frac{(n_1 - 1) \times {s_1}^2 + (n_2 - 1) \times {s_2}^2}{n_1 + n_2 - 2}$$

$s_1^2,\ s_2^2$ are unbiased sample variances of 2 samples.

Under the null hypothesis H0:"there is no differecnce ", it is reduced to

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example⑥

There are 2 classes. Class A and B consist of 30 and 32 students, respectively. We compare the result of an exam.

Now, the mean and unbiased S.D. of the scores of class 1 were 75 and 5, respectively. On the other hand, the mean and unbiased S.D. of the scores of class 2 were 70 and 8, respectively.

Under the equivalence of variance, can you state that there is any difference between these classes? Answer with the significance level of 5%.

Example⑥ 【Answer】

Null hypothesis H0:”No difference”.
Alternative hypothesis H1: “There is some difference.”

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

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example⑥ 【Answer】

Find the p.d.f of the test statistic:

Example⑥ 【Answer】

Find the p.d.f of the test statistic:

→ Follows the t-distribution of df = (30+32-2=60).

Example⑥ 【Answer】

Find the rejection region.
Two-sided test in this case.

The rejection region is outside of the upper/lower 2.5-percentiles of t-distribution. $t_{60}\left(\frac{0.05}{2}\right) = 2.0$

$|t| > 2.0$ is the rejection region.

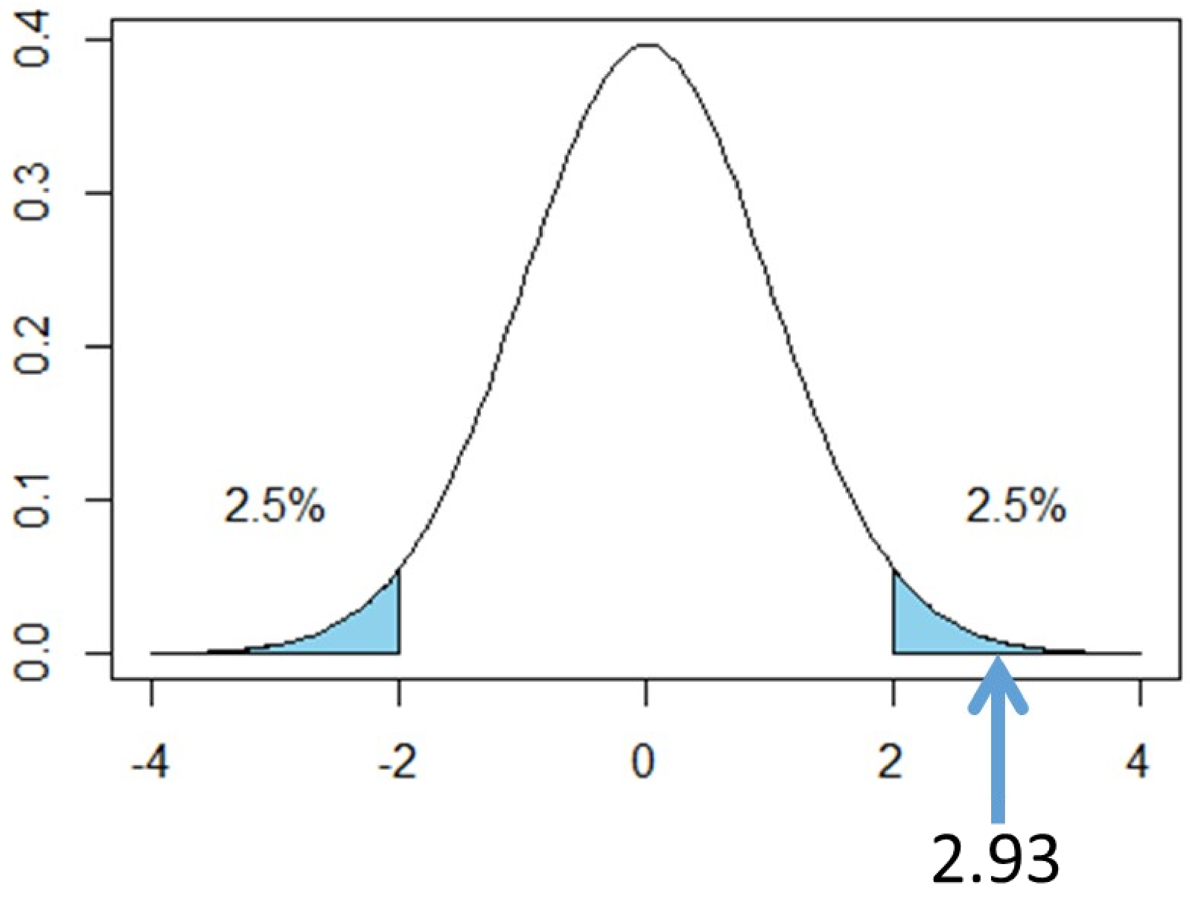
Example⑥ 【Answer】

・Find the value of test statistics $T(x_1,x_2,...,x_N)$ on the basis of H0. In this case,

$$s^2 = \frac{(30-1)\times 5^2 + (32-1)\times 8^2}{30+32-2} = 45.15$$

$$t = \frac{75-70}{\sqrt{45.15\times\left(\frac{1}{30}+\frac{1}{32}\right)}} \fallingdotseq 2.93$$

is in the rejection region.



Example⑥ 【Answer】

We reject the null hypothesis H0 under the significance level of 5%.

“There is a significant difference between the 2 classes.”

iii) Population variances σ_1^2 , σ_2^2 are unknown.

※Called as the Welch’s test

Independent sample t-test

iii) Population variances σ_1^2, σ_2^2 are unknown.

$$\boldsymbol{x}_1 = \{x_{11}, x_{12}, \ldots, x_{1n_1}\} \qquad \boldsymbol{x}_2 = \{x_{21}, x_{22}, \ldots, x_{2n_2}\}$$

- The sample mean and the population mean of the first sample are denoted as $\bar{x}_1$ and μ_1, respectively.

The sample mean and the population mean of the second sample are denoted as $\bar{x}_2$ and μ_2, respectively.

Independent sample t-test

iii) Population variances σ_1^2, σ_2^2 are unknown.

Null hypothesis H0:”No difference”

→ test statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

where

$$s_1^2 = \frac{(x_{11} - \bar{\boldsymbol{x}}_1)^2 + \ldots + (x_{1N_1} - \bar{\boldsymbol{x}}_1)^2}{N_1 - 1}$$

$$s_2^2 = \frac{(x_{21} - \bar{\boldsymbol{x}}_2)^2 + \ldots + (x_{2N_2} - \bar{\boldsymbol{x}}_2)^2}{N_2 - 1}$$

Independent sample t-test

iii) Population variances σ_1^2, σ_2^2 are unknown.

The test statistic $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Approximately follows the t-distribution of df = m.
Here,

m is an integer closest to

$$\frac{\left(\dfrac{s_1^2}{N_1} + \dfrac{s_2^2}{N_2}\right)^2}{\dfrac{s_1^4}{N_1^2(N_1-1)} + \dfrac{s_2^4}{N_2^2(N_2-1)}}$$

Example⑦

The tables below show the measured body temperature

By using i)electronic thermometer, and ii) mercury thermometer. In general, it is said that the electronic one shows higher values than the mercury thermometer.

Now, validate this statement by the hypothesis testing

with the significance level of 5%.

Sample	Electronic thermometer	Mercury thermometer
1	37.1	36.8
2	36.7	36.6
3	36.6	36.5
4	37.4	37.0
5	36.8	36.7

Sample	Electronic thermometer	Mercury thermometer
6	36.7	36.5
7	36.9	36.6
8	37.4	37.1
9	36.6	36.4
10	36.7	36.7

Example⑦ 【Answer】

Null hypothesis H0:”No difference”.

Alternative hypothesis H1;”Electronic is higher”.

One-sided!

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

Example⑦ 【Answer】

Find the p.d.f of the test statistic:

First, we should find the unbiased S.D.s.

$$\bar{x}_1 = 36.89,\ \bar{x}_2 = 36.69$$

$$s_1^2 = 0.0939,\ s_2^2 = 0.0498$$

The test statistic in the Welch' test is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Example⑦ 【Answer】

Find the p.d.f of the test statistic:

$$\frac{\left(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2(N_1-1)}+\frac{s_2^4}{N_2^2(N_2-1)}}=\frac{\left(\frac{0.0939}{10}+\frac{0.0498}{10}\right)^2}{\frac{(0.0939)^2}{10^2\times(10-1)}+\frac{(0.0498)^2}{10^2\times(10-1)}}=16.45$$

$\Rightarrow$ t-distribution of df=m=16

Find the rejection region.
One-sided test in this case.

The rejection region is outside of the upper5-percentile of t-distribution.

$$t_{16}(0.05) = 1.746$$

$|t| > 1.746$ is the rejection region.

- Find the value of test statistics $T(x_1,x_2,\ldots,x_N)$ on the basis of H0. In this case,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{36.89 - 36.69}{\sqrt{\frac{0.0939}{10} + \frac{0.0498}{10}}} = 1.668$$

is not in the rejection region.

We cannot reject H0 under the significance level of 5%.

“We cannot state that there is a significant difference between
These 2 types of thermometers”.

stats.ttest_ind() does the Welch's test.

Note that ttest_ind() returns the *two-sided* p-value!

```
import numpy as np
from scipy import stats


X1=np.array([37.1,36.7,36.6,37.4,36.8,36.7,36.9,37.4,36.6,36.7])
X2=np.array([36.8,36.6,36.5,37.0,36.7,36.5,36.6,37.1,36.4,36.7])

stat_t, p_val = stats.ttest_ind(X1,X2,equal_var=False)

p_val=p_val/2
print("p-value is")
print(p_val)

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

Euqal_val=False
means Weltch's test.
(True as default.)

stats.ttest_ind() does the Welch’s test.
Note that ttest_ind() returns the *two-sided* p-value!

```
import numpy as np
from scipy import stats


X1=np.array([37.1,36.7,36.6,37.4,36.8,36.7,36.9,37.4,36.6,36.7])
X2=np.array([36.8,36.6,36.5,37.0,36.7,36.5,36.6,37.1,36.4,36.7])

stat_t, p_val = stats.ttest_ind(X1,X2,equal_var=False)

p_val=p_val/2
print("p-value is")
print(p_val)

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is
0.05738829046148806
Cannot reject H0.
```

Should be made half because one-sided test here.

In case of two-sided test

Ex⑦ 【Python】

Apply *stats.ttest_ind()*
Its p-value means the one for the two-sided test.

```
import numpy as np
from scipy import stats


X1=np.array([37.1,36.7,36.6,37.4,36.8,36.7,36.9,37.4,36.6,36.7])
X2=np.array([36.8,36.6,36.5,37.0,36.7,36.5,36.6,37.1,36.4,36.7])

stat_t, p_val = stats.ttest_ind(X1,X2,equal_var=False)

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

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is
0.11477658092297612
Cannot reject H0.
```

Euqal_val=False
Weltch’s test.
(Default is True)

Need not make p_val half now.

Test for equality of variance

Test for equality of variance

How we check "the equality of variance"?

- Assume 2 sampels:

$x_1 = \{x_{11}, x_{12}, \ldots, x_{1N1}\}$, $x_2 = \{x_{21}, x_{22}, \ldots, x_{2N2}\}$

Then, how we measure the ratio of their variances?

Test for equality of variance

Define

$$T(s_1, s_2) = s_1^2 / s_2^2$$

,where

$$s_1^2 = \frac{(x_{11} - \bar{\boldsymbol{x}}_1)^2 + \ldots + (x_{1N_1} - \bar{\boldsymbol{x}}_1)^2}{N_1 - 1}$$
$$s_2^2 = \frac{(x_{21} - \bar{\boldsymbol{x}}_2)^2 + \ldots + (x_{2N_2} - \bar{\boldsymbol{x}}_2)^2}{N_2 - 1}$$

Follows the F-distribution of df=(N_1-1, N_2-1): $F_{(N_1-1, N_2-1)}$

Test for equality of variance

Two-sided test.

Null hypothesis H0:"No difference between the variances", i.e., $\sigma_1=\sigma_2$.

Alternative hypothesis H1:" $\sigma_1 \neq \sigma_2$ "

Then, we check whether $T(s_1, s_2) \in R$ holds or not.

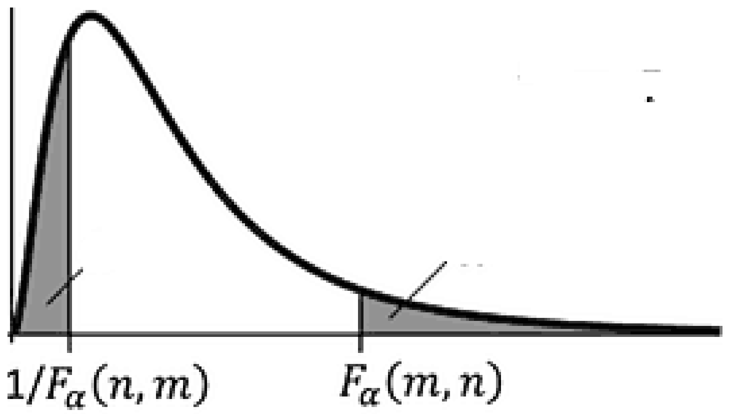
Equivalenylly: reject H0 if

$$s_1^2/s_2^2 \leq F_{(N_1-1,N_2-1)}(1-0.05/2) = \frac{1}{F_{(N_2-1,N_1-1)}(0.05/2)}$$

Or

$$s_1^2/s_2^2 \geq F_{(N_1-2,N_2-1)}(0.05/2)$$

holds.



Example

- Do the test for equality of vairiance to the samples below.

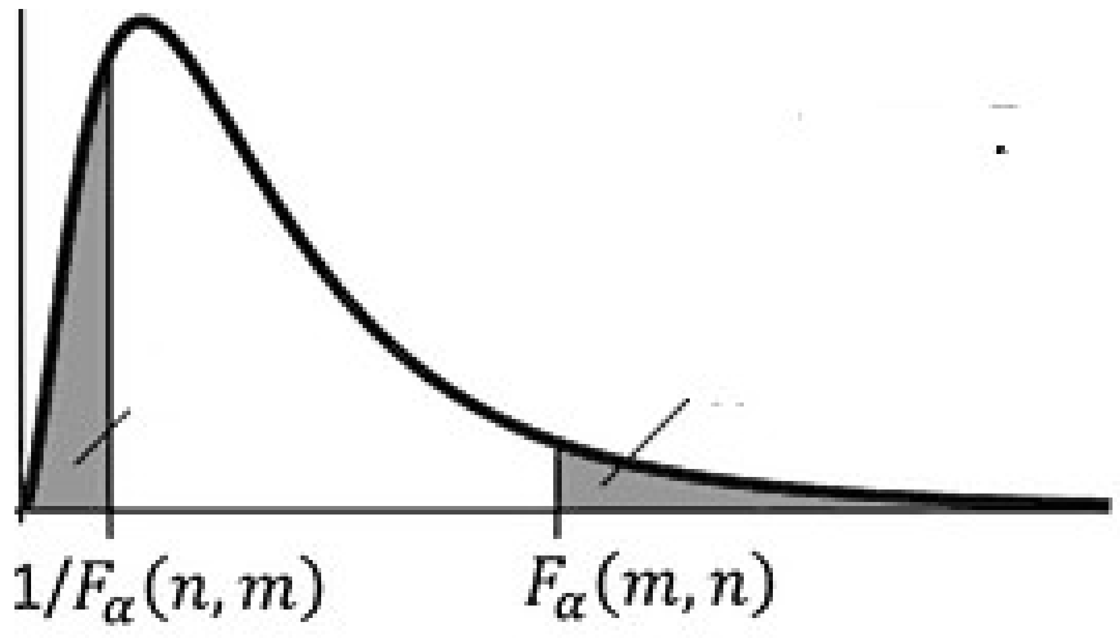
Group- A	Group- B
10	9
9	10
9	8
9	10
7	8
8	7
6	7
7	5
5	7

Example 【Answer】

$$T(s_1, s_2) = s_1^2 / s_2^2$$

Follows the F-distribution of df=(8,8).

$$T(s_1, s_2) = s_1^2 / s_2^2 = 1.0319$$



0.0226

4.433

Example 【Answer】

We cannot reject H0. That is, no significant difference.
→The equality of variance holds.

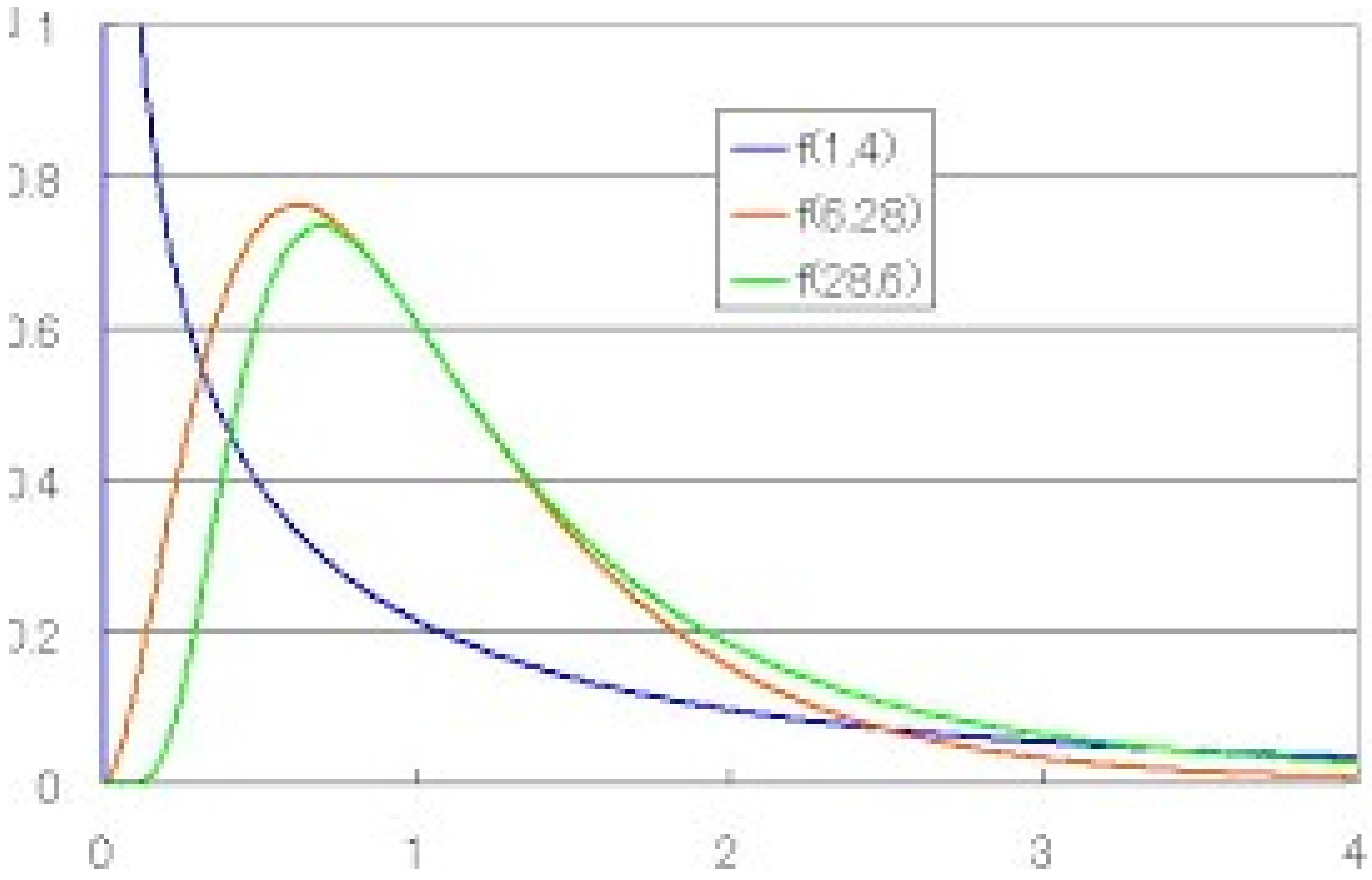
F-distribution

- As we have seen before, F-distribution has two dfs!
- Applied to the test for equivalence of variance.
- 【pdf】
- The F-distribution with df=(n_1, n_2):

$$f(x; n_1, n_2) = \frac{\Gamma(\frac{n_1+n_2}{2})\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})\left(1+\frac{n_1}{n_2}x\right)^{\frac{n_1+n_2}{2}}} \quad (0 < x < \infty)$$

Pdf of F-distribution.

- With various dfs.



F-table

http://www.biwako.shiga-u.ac.jp/sensei/mnaka/ut/fdisttab.html

（Upper 2.5-percentile）

α =0.025	自由度 m 1	2	3	4	5	6	7	8	9	10	11	12
自由度 n												
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	973.03	976.71
2	38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387	39.398	39.407	39.415
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419	14.374	14.337
4	12.218	10.649	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.7935	8.7512
5	10.007	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.5678	6.5245
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.4098	5.3662
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.7095	4.6658
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572	4.2951	4.2434	4.1997
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.9121	3.8682
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.6649	3.6209
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879	3.5257	3.4737	3.4296
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358	3.3736	3.3215	3.2773
13	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	3.3880	3.3120	3.2497	3.1975	3.1532
14	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	3.2853	3.2093	3.1469	3.0946	3.0502
15	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227	3.0602	3.0078	2.9633
20	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365	2.7737	2.7209	2.6758

Exercises

Exercise①

Check whether the capacity of beer bottle is smaller than 633mlon average or not. Set the significance level at 5%.

No.	Capacity[ml]
1	632.9
2	633.1
3	633.2
4	632.3
5	633.1
6	634.7
7	633.6
8	633.0
9	632.4
10	633.6

Exercise① 【Answer】

Null hypothesis H0:"The mean μ is 633ml."
Alternative hypothesis H1:" The mean μ is smaller than 633ml."

Define the test statistic $T(x_1,x_2,\dots,x_N)$:

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

Exercise① 【Answer】

- Since $\bar{x} = 633.19 > 633$ holds,
  we can terminate the test.

"We cannot reject H0". That is,
"We cannot state that the capacity is smaller than 633ml."

Exercise②

The table below shows the weight of 20 bags that encloses the bread flour of 25.5kg.
Check that we can say that the average weight is 25.5.kg or not. Set the significance level at 5%.

	Data
Sample size	20
Mean	25.29
Unbiased variance	2.23

Exercise② 【Answer】

Null hypothesis H0:"The mean μ is 25.5."
Alternative hypothesis H1:" The mean μ is not 25.5."

- Find the value of test statistics $T(x_1,x_2,\ldots,x_N)$ on the basis of H0. In this case,

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}} = \frac{25.29 - 25.5}{\sqrt{\frac{2.23}{20}}} \fallingdotseq -0.63$$

Exercise② 【Answer】

Find the p.d.f of the test statistic:
→ t-distribution of df = 20-1 = 19.

Determine the rejection region.
→ Two-sided test, so outside of upper/lower 2.5-percentiles.

$$t_{19}\left(\frac{0.05}{2}\right)$$

- $|t| > 2.093$ is the rejection region.

Exercise② 【Answer】

$$t = \frac{\overline{x} - \mu}{\sqrt{\frac{s^2}{n}}} = \frac{25.29 - 25.5}{\sqrt{\frac{2.23}{20}}} \fallingdotseq -0.63$$

Not in the rejection region |t|> 2.093!
W cannot reject H0.

We cannot state that the mean is not 25.5kg.

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

mu_0=25.5
X_mean = 25.29
X_sd=np.sqrt(2.23)
N=20
stats_t = (X_mean - mu_0)/(X_sd/np.sqrt(N))
p_val=t.cdf(-np.abs(stats_t),df=N-1)*2
print("p-value is")
print(p_val)

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is
0.5369027187014798
Cannot reject H0.
```

Exercise③

A researcher surveyed the number of a certain birds in a island. As a result of the 10 times' survey, they found the mean and unbiased S.D. were 25 and 3, resp.

Then, can we state that "there exist more than 21 of this species of birds"? Check with it the significance level of 5%.

Exercise3 【Answer】

Null hypothesis H0:”the number of this species is 21.”
Alternative H1:”The number is larger than 21”.

Define the test statistic $T(x_1,x_2,\ldots,x_N)$:

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

Exercise3 【Answer】

Find the p.d.f of the test statistic:
→ t-distribution of df = 10-1 = 9.

Determine the rejection region.
→ One-sided test, so outside of upper 5-percentile.

$$t_9(0.05) = 1.833$$

- t> 1.833 is the rejection region.

Find the value of the test statistic:
On the basis of H0,

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}} = \frac{25 - 21}{\sqrt{\frac{3^2}{10}}} \fallingdotseq 4.22$$

In t> 1.833 !
We reject H0:

“The number is larger than 21.”

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

mu_0=21
X_mean = 25
X_sd = 3

N=10
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("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is
0.0011257483136483594
Reject H0.
```

Exercise④

We collected 16 patients of a certain disease of thyroid gland, and measured the concentration of the calcium. Then, the sample mean was 7.4[mg/dl].

On the other hand, it is known that the value of healthy people follows the normal distribution with its mean of 9.8[mg/dl].
In addition, the S.D. of the concentration has the S.D. of 0.5[mg/dl].

Now, can we state that the concentration of 16 patients significantly differs from that of healthy people?
Set the significance level at 5%.

Exercise④

We collected 16 patients of a certain disease of thyroid gland, and measured the concentration of the calcium. Then, the sample mean was 7.4[mg/dl].

On the other hand, it is known that the value of healthy people follows the normal distribution with its mean of 9.8[mg/dl].
In addition, the S.D. of the concentration has the S.D. of 0.5[mg/dl].

Apply the usual t-test.

Now, can we state that the concentration of 16 pa
significantly differs from that of healthy people?
Set the significance level at 5%.

Exercise④ 【Answer】

Null hypothesis H0:”The concentration is 9.8[mg/dl].”
Alternative H1:” The concentration is not 9.8[mg/dl].”.

Define the test statistic $T(x_1,x_2,...,x_N)$. Since S.D. is known,

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

Exercise④ 【Answer】

Find the p.d.f of the test statistic:
→ z-distribution.

Determine the rejection region. Since two-sided test, outside of upper/lower 2.5-percentiles of z-dist. is the rejection region.

- $|t|> 1.96$ is the rejection region.

Exercise④ 【Answer】

Find the value of the test statistic:
On the basis of H0,

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{7.4 - 9.8}{\frac{0.5}{\sqrt{16}}} = -19.2$$

In |t|> 1.96 !
We reject H0,

"There exists a significant difference."

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

mu_0=9.8
X_mean = 7.4
X_sd = 0.5
N=16

z = (X_mean -mu_0)/ (X_sd / np.sqrt(N))
p_val = norm.cdf( -np.abs(z),0,1)*2

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

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is
3.7009201398846437e-82
Reject H0.
```

Q.5

In a certain bakery, they make two kinds of breads in factories A and B.
As a result of a sample survey, the mean and unbiased SD of those made in the factory A were 93 and 3.7, respectively.
On the other hand, the mean and unbiased SD of those made in the factory B were 87 and 3.9, respectively.
The sample sizes of A and B are 10 and 8, respectively.

Then, can you say there's a difference between the

A.5

- H_0:"No difference"

  H_1: "There's a difference"

   -> independent samples test. Two-sided test.

  Define the test statistic $T(x_1,x_2,...,x_N)$. Since SDs are
  unknown...
  T-distribution!

A.5

Find the df. Due to welch’s test,

$$\frac{\left(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2(N_1-1)}+\frac{s_2^4}{N_2^2(N_2-1)}}=14.76$$

$$t_{15}(0.05/2)=2.13$$

∴ The rejection region is $|t|>2.13$.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 3.32$$

- Belongs to the rejection region: $|t|>2.13$

Reject H0!

That is,

“There’s a difference.”

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

M1=93
M2=87

sd1=3.7
sd2=3.9

N1=10
N2=8

stat_t = (M2-M1)/np.sqrt( pow(sd1,2)/N1 + pow(sd2,2)/N2 )

m11= pow(sd1,2)/N1 + pow(sd2,2)/N2
m1 = pow(m11,2)
m2 = pow(sd1,4)/pow(N1,2)/(N1-1) + pow(sd2,4)/pow(N2,2)/(N2-1)
df_val=m1/m2

p_val = stats.t.cdf(-np.abs(stat_t),df=df_val)*2

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

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is
0.004772875717597794
Reject H0.
```

Q.6

As a result of a questionnaire to women concerning the height of their shoes, 24 women who answered as “I’m cool.” had the mean and unbiased SD of 3.67[cm] and 1.79[cm], resp.

On the other hand, the mean and unbiased SD of those 48 women who answered as “I’m not so cool.” were 2.77[cm] and 1.29[cm], resp.

Then, is there any difference between these two groups or not?
Set the significance level at 5%. (Do the two-sided

A.6

- H_0:"No difference"

  H_1: "There's a difference"

  -> independent samples test. Two-sided test.

  Define the test statistic $T(x_1, x_2, \ldots, x_N)$. Since population SDs are unknown...
  T-distribution!

A.6

Find the df. Due to welch's test,

$$\frac{\left(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2(N_1-1)}+\frac{s_2^4}{N_2^2(N_2-1)}}=35.33$$

$$t_{35}(0.05/2)=2.03$$

- The rejection region is $|t|>2.03$.

A.6

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 2.826$$

- Belongs to the rejection region: |t|>2.03

Reject H0!

That is,

"There's a difference."

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

M1=3.67
M2=2.77

sd1=1.79
sd2=1.29

N1=24
N2=48

stat_t = (M2-M1)/np.sqrt( pow(sd1,2)/N1 + pow(sd2,2)/N2 )

m11= pow(sd1,2)/N1 + pow(sd2,2)/N2
m1 = pow(m11,2)
m2 = pow(sd1,4)/pow(N1,2)/(N1-1) + pow(sd2,4)/pow(N2,2)/(N2-1)
df_val=m1/m2

p_val = stats.t.cdf(-np.abs(stat_t),df=df_val)*2

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

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is
0.034850398357228147
Reject H0.
```

Q.7

Due to the questionnaire to 70 male people of 20-59 years old
concerning the price of the shaver ,

・ The mean and unbiased SD of 25 people of 20-39 years old were \15400 and \2810, resp.
・ The mean and unbiased SD of 45 people of 40-59 years old were \18600 and \4720, resp.

Then, is there any difference between these two groups or not?
Set the significance level at 5%. (Do the two-sided test)

A.7

- H_0:”No difference”

  H_1: ”There’s a difference”

  -> independent samples test. Two-sided test.

  Define the test statistic $T(x_1,x_2,\dots,x_N)$. Since population SDs are unknown…
  T-distribution!

A.7

Find the df. Due to welch's test,

$$\frac{\left(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2(N_1-1)}+\frac{s_2^4}{N_2^2(N_2-1)}}=67.60$$

$$t_{68}(0.05/2)=2.00$$

- The rejection region is |t|>2.00.

A.7

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = -3.55$$

- Belongs to the rejection region: $|t|>2.00$

Reject H0!

That is,

"There's a difference."

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

M1=15400
M2=18600

sd1=2810
sd2=4720

N1=25
N2=45

stat_t = (M1-M2)/np.sqrt( pow(sd1,2)/N1 + pow(sd2,2)/N2 )

m11= pow(sd1,2)/N1 + pow(sd2,2)/N2
m1 = pow(m11,2)
m2 = pow(sd1,4)/(pow(N1,2)*(N1-1)) + pow(sd2,4)/(pow(N2,2)*(N2-1) )
df_val=m1/m2

p_val = stats.t.cdf(-np.abs(stat_t),df=df_val)*2

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

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is
0.0006985268118186147
Reject H0.
```

Q.8

For two machines A and B, that have almost the same specs,
We did a sample survey. Based on the samples of 18 and 9 companies that introduced A and B, resp., we calculated the
average error occurrence per year and per 10 machines, and summarized as below.
Then, is there any difference? Set the significance leve

	Mean	Unbiased SD
A	10.56	10.68
B	8.22	2.17

A.8

- H_0:"No difference"

  H_1: "There's a difference"

  -> independent samples test. Two-sided test.

  Define the test statistic $T(x_1,x_2,...,x_N)$. Since population SDs are unknown...
  T-distribution!

A.8

Find the df. Due to welch's test,

$$\frac{\left(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2(N_1-1)}+\frac{s_2^4}{N_2^2(N_2-1)}}=19.64$$

$$t_{20}(0.05/2)=2.06$$

- The rejection region is |t|>2.06.

A.8

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 2.56$$

- Belongs to the rejection region: $|t|>2.06$

Reject H0!

That is,

“There’s a difference.”

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

M1=10.56
M2=8.22

sd1=10.68
sd2=2.17

N1=18
N2=9

stat_t = (M1-M2)/np.sqrt( pow(sd1,2)/N1 + pow(sd2,2)/N2 )

m11= pow(sd1,2)/N1 + pow(sd2,2)/N2
m1 = pow(m11,2)
m2 = pow(sd1,4)/(pow(N1,2)*(N1-1)) + pow(sd2,4)/(pow(N2,2)*(N2-1) )
df_val=m1/m2
p_val = stats.t.cdf(-np.abs(stat_t),df=df_val)*2

print("p-value is",p_val)

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is 0.3824527698717164
Cannot reject H0.
```

Q.9

In a certain elementary school, they surveyed the pocket money of 50 boys and 52 girls. Is there any difference?

	Girls	Boys
Sample size	52	50
Mean[\]	25356	32430
Unbiased SD[\]	21171	24663

A.9

- H_0:"No difference"

H_1: "The mean of the patients is higher."

-> independent samples test. Two-sided test.

Define the test statistic $T(x_1,x_2,\ldots,x_N)$. Since population SDs are unknown…
T-distribution!

A.9

Find the df. Due to welch's test,

$$\frac{\left(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2(N_1-1)}+\frac{s_2^4}{N_2^2(N_2-1)}}=96.49$$

$$t_{96}(0.05/2)=1.985$$

- The rejection region is |t|>1.985.

A.9

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = -1.55$$

- Does not belong to the rejection region: $|t|>1.985$

Cannot reject H0.

That is,

“We cannot say that there’s a difference.”

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

M1=25356
M2=32430

sd1=21171
sd2=24663

N1=52
N2=50

stat_t = (M1-M2)/np.sqrt( pow(sd1,2)/N1 + pow(sd2,2)/N2 )

m11= pow(sd1,2)/N1 + pow(sd2,2)/N2
m1 = pow(m11,2)
m2 = pow(sd1,4)/(pow(N1,2)*(N1-1)) + pow(sd2,4)/(pow(N2,2)*(N2-1) )
df_val=m1/m2
p_val = stats.t.cdf(-np.abs(stat_t),df=df_val)*2
print("p-value is",p_val)

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is 0.12402045819753468
Cannot reject H0.
```

Q.10

The table below shows the result of the measured IgG (mg/100ml) of 9 patients of a certain disease and 7 healthy people.
Here, IgG is a sort of a protein.
You can state that the value of IgG of patients is higher than that of healthy people? Do the one-sided test with the significance level of 5%.

免疫グロブリンIgG値(mg/100ml)

No.	Patients	Healthy
1	1326	1220
2	1418	1080
3	1820	980
4	1516	1420
5	1635	1170
6	1720	1290
7	1580	1116
8	1452	
9	1600	

「新版 医学への統計学」(朝倉書店)

A.10

- H_0:"No difference"

H_1: "The mean of the patients is higher."

-> independent samples test. One-sided test.

Define the test statistic $T(x_1,x_2,\ldots,x_N)$. Since population SDs are unknown...
T-distribution!

	Patients	Healthy
Sampel mean	1563.0	1182.286
Unbiased SD	s_1^2=153.8116	s_2^2=144.6026
Sample size	N_1=9	N_2=7

A.10

Find the df. Due to welch's test,

$$\frac{\left(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2(N_1-1)}+\frac{s_2^4}{N_2^2(N_2-1)}} = 13.238$$

$$t_{13}(0.05) = 1.77$$

- The rejection region is |t|>1.77.

A.10

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 5.08$$

- Belongs to the rejection region: $|t| > 1.77$

Reject H0!

That is,

“Patients are higher.”

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

x1 = np.array([1326,1418,1820,1516,1635,1720,1580,1452,1600])
x2 = np.array([1220,1080,980,1420,1170,1290,1116])

stat_t,p = stats.ttest_ind(x1,x2,equal_var=False)
p_val = p/2
print("p-value is",p_val)

if p_val<0.05:
    print("Reject H0.")
else:
    print("Cannot reject H0.")
```

```
p-value is 9.565721758604052e-05
Reject H0.
```

summary

- Divide the two sample t-test into 2 parts.
- Divide the independent samples test into 3 parts.
- You can do the Welch’s test?