Biostatistics

Week #14

5/24/2022

Ch 11 – Comparison of Two Means

Introduction

- In previous chapter we learned how to compare the unknown mean of a ***single*** population to some fixed, known value μ_0.
- In practical applications, it is common to **compare the means of *two* populations where both means are not known**.
- ***Matched pair comparison*** is important in many scientific aspects.
- The idea is to draw a conclusion about their **similarities** or **differences**.

Example 1

- We are interested in knowing the relationship between the use of ***oral contraceptives (OC)*** and the level of ***blood pressure (bp)*** in women.
- Two different experiment designs can be used to assess this relationship – ***longitudinal*** study and ***cross-sectional*** study.

Longitudinal Study

- Identify a group of women who are not currently OC users and measure their bp.
- Rescreen these (**same**) women 1 year later and ascertain a subgroup who have become OC users. Measure their bp. This approach is often called a ***follow-up*** study.
- Compare these bps **for the same women**.

Cross-Sectional Study

- Identify both a group of OC users and a group of non-OC users. Measure their bps.
- Compare their bps.

Comments

- Longitudinal study uses <u>**paired samples**</u>. It is more <u>***definitive***</u>, since most confounding factors that influence the women’s bp will be present both in the screening and the follow-ups.
- Cross-sectional study uses independent samples, which is usually considered <u>***suggestive***</u> only, because of possible confounding factors.

11.1 Paired Samples

Example 2

- It was **suspected** that the amount of carbon monoxide CO (一氧化碳) in the air could increase the carboxyhemoglobin (一氧化碳血紅素) level for patients with coronary artery (冠狀動脈) disease, which might cause angina (心絞痛) .
- We wish to perform a hypothesis testing on the assumption that ***indeed the increased CO level would be a threat to patients with coronary artery disease***.

Study design

- Same individuals for taking the experiment (a longitudinal study).
  - ***Control group*** for having them breathe in normal air.
  - ***Experiment group*** for having them breathe in CO-rich air.
- Which group would have a ***quicker* onset** of angina. That is, the ***time elapse*** to angina onset is ***shorter***.

Note that ‘group’ here means different breathing conditions, not different people.

Cont'd

- 63 patients are randomly selected to measure ***the percent decreases in time to angina*** (our random variable) for **each** of the following two occasions. (See next slide)
- More decrease in time to angina means quicker onset, which is more threatening.

- Occasion #1 (control group):
  - On a given day, each individual exercises on a treadmill (跑步機) until the patient experiences angina, for which the onset time for angina t_1 was recorded.
  - **The same patient** (after experiencing angina at t_1) is exposed to **plain room air** for approximately 1 hour, followed by performing a second exercise test until the onset time for another angina t_2 was recorded.
  - *Usually t_2 will be smaller than t_1, so t_1-t_2 would be a positive number.*

- Occasion #2 (experiment group):
  - On *another* day, each individual again exercises on a treadmill (跑步機) until the patient experiences angina, for which the onset time for angina t_1 was recorded.
  - The same patient (after experiencing angina at t_1) is exposed to **room air abundant with carbon monoxide (CO)** for approximately 1 hour, followed by performing a second exercise test until the onset time for another angina t_2 was recorded.

- We are interested in the "percent **decrease** in time"
- For example, if a patient was recorded $t_1 = 983$ and $t_2 = 957$ seconds in control group, we will have this value

$$\frac{983 - 957}{983} = 0.026 = 2.6\%$$

- The same individual in experiment group has $t_1 = 991$ and $t_2 = 900$, then

$$\frac{991 - 900}{991} = 0.092 = 9.2\%$$

Larger means more threatening…

- It is also possible that the two random variables are (+,–), (–,+) and (–,–).
- For example, if a patient was recorded t_1 = 983 and t_2 = 957 seconds in control group, we will have this value

$$\frac{983-957}{983}=0.026=2.6\%$$

- The same individual in experiment group has t_1 = 991 and t_2 = 999(吸了CO變得更強!!), then

$$\frac{991-999}{991}=-0.008=-0.8\%$$

Larger means more threatening…

The matched pairs experiment

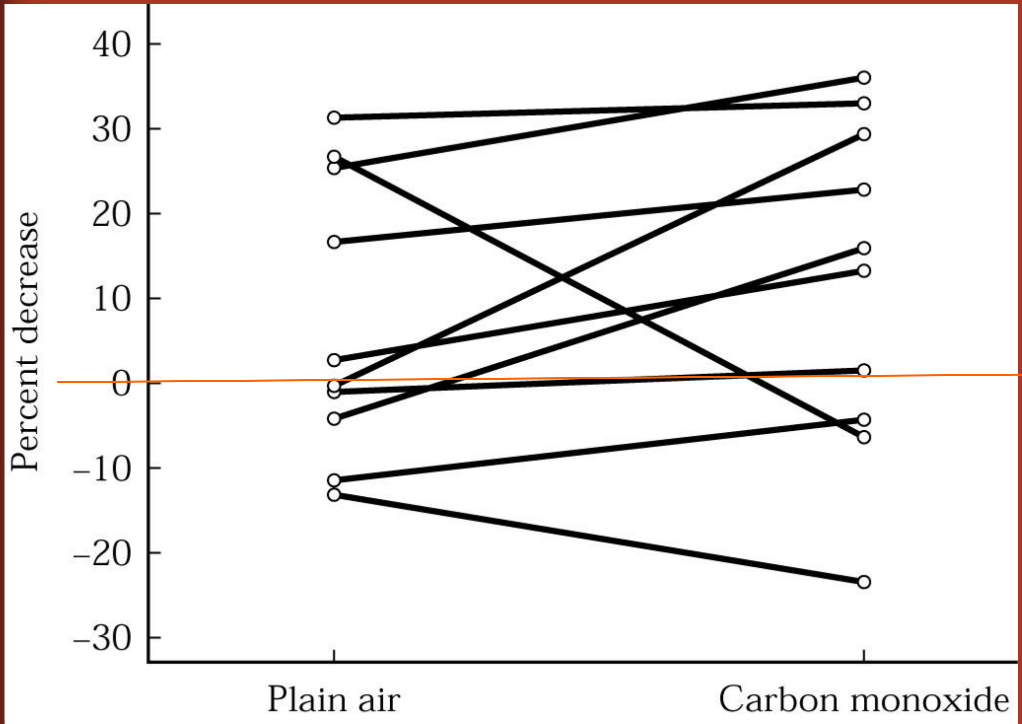


FIGURE 11.1
Percent decrease in time to angina on two different occasions for each of ten men with coronary artery disease

- It looks like most of the pairs have the "percent decrease" up (i.e., earlier onset of angina in CO environment) instead of down.
- ***Is such increase significant?***

Larger means more threatening…

Matched Pairs Experiment – initial statement

- Compare two populations of percent decreases in time to angina.
- The parameter tested is $\mu_1 - \mu_2$

Sample 1	Sample 2
x_{11}	x_{12}
x_{21}	x_{22}
x_{31}	x_{32}
$\vdots$	$\vdots$
x_{n1}	x_{n2}

Negative means threatening. More negative, more threatening...

μ_1 The population mean of percent decreases exposed to **plain** air

μ_2 The population mean of percent decreases exposed to **carbon Monoxide (CO)**

The Null Hypothesis

- We'd like to know whether indeed the increased CO level would be a threat to patients with coronary artery disease.
- A threat can be interpreted as **a quicker onset of second angina** (a larger "percent decrease in time", or μ_2 is greater than μ_1).
- **Let the null hypothesis be the following:**

$$H_0: (\mu_1 - \mu_2) \geq 0$$

1: plain air
2: carbon monoxide

Cont'd

- This null hypothesis states that observations in sample 1 would be in general matching or greater than in sample 2.
- Rejecting it would suggest otherwise. That is, sample 2 indeed gives **significantly larger values** of "percent decrease in time"

$$H_0: (\mu_1 - \mu_2) \geq 0$$
$$H_A: (\mu_1 - \mu_2) < 0$$

The problem is…

- ***How would we evaluate $\mu_1 - \mu_2$ as a random variable?***
- Computing μ_1 from sample #1 and μ_2 from sample #2 will not serve this purpose. [because it is a single observation, not a random variable]
- We need to define a random variable representing the difference of “percent decrease in time” between two samples.

Matched Pairs Experiment – changing statement

- Since the **difference of the means** is equal to the **mean of the differences**, we can rewrite the hypotheses in terms of δ (the mean of the differences) rather than in terms of $\mu_1 - \mu_2$.

Difference ***d***

$$d_1 = x_{11} - x_{12}$$
$$d_2 = x_{21} - x_{22}$$
$$d_3 = x_{31} - x_{32}$$
$$\ldots$$
$$d_n = x_{n1} - x_{n2}$$

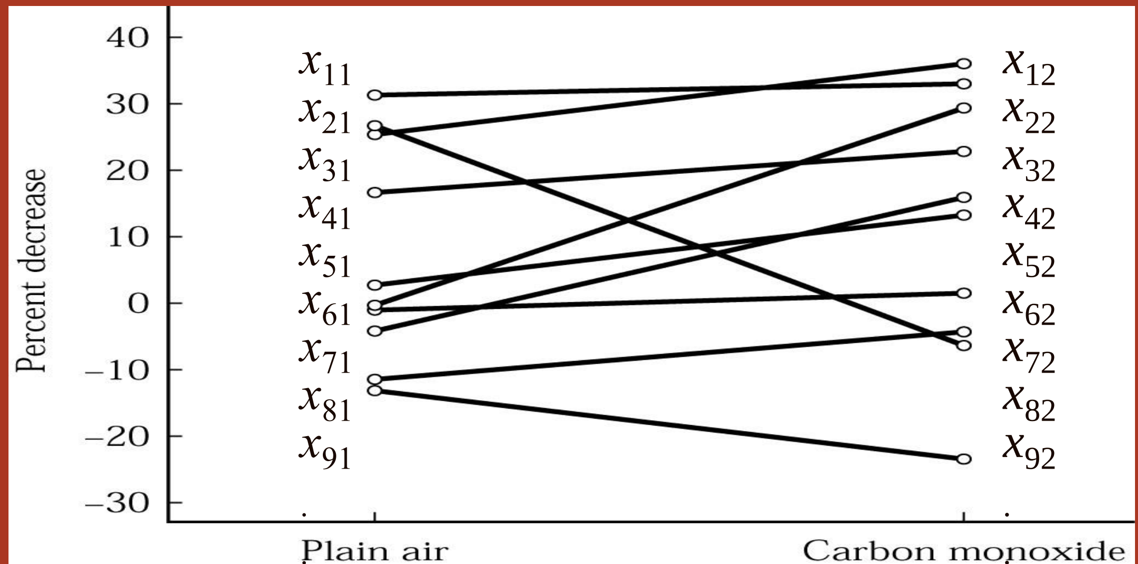


FIGURE 11.1
Percent decrease in time to angina on two different occasions for each of ten men with coronary artery disease

The paired t-test

- Solution
  - $\delta = \mu_1 - \mu_2$ is the "percent decrease in time"
  - The hypotheses:
    H_0: $\bar{d} = \delta (= 0)$ (i.e., no difference of angina onset)
    H_A: $\delta < 0$ (scenario #2 has bigger "percent decrease in time", or early onset of angina)
  - The *t* statistic:

$$t = \frac{\bar{d} - \delta}{s_d / \sqrt{n}}$$

Note that we changed $\delta \geqq 0$ to $\delta = 0$ here!

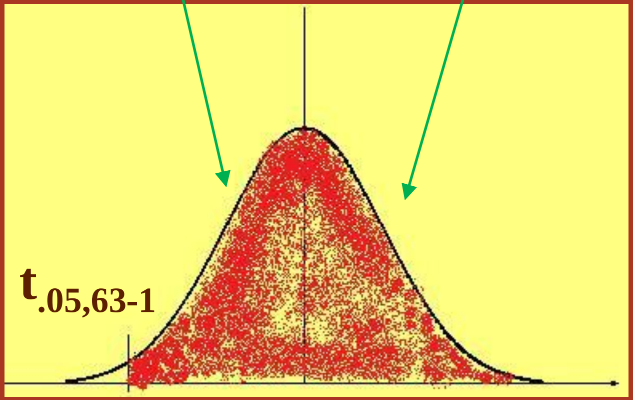
The paired *t*-test

The t-statistics

$$t = \frac{\bar{d} - \delta}{s_d / \sqrt{n}}$$

Mean values only slightly less than 0 would suggest μ_1 and μ_2 are comparable.

Mean values greater than 0 wouldn't serve our purpose to show CO threat.



For the null hypothesis to match $\delta = 0$, a negative mean value of ***d*** would result in a negative ***t*** towards the left tail. Further into the left tail would suggest to reject the null hypothesis that $\delta = 0$.

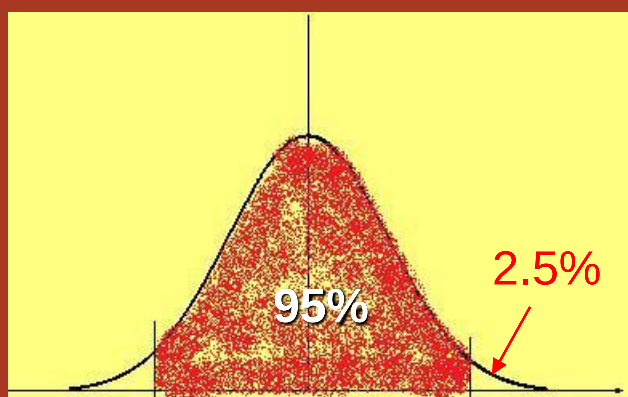
This is a one-sided test with a left-tail.

The rejection region is $t < t_{.05,63-1}$

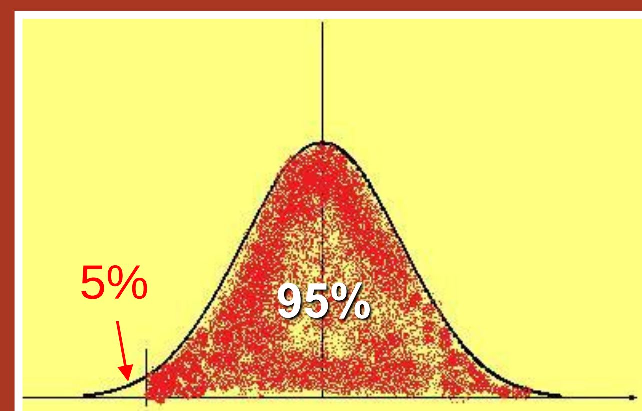
Degrees of freedom = n – 1

Computing $t_{.05,63-1}$

- Recall that earlier we had the MATLAB function TINV giving this value t.
- Note that we had a 2-tailed test then (left).
- Now we have a one-sided test (right).



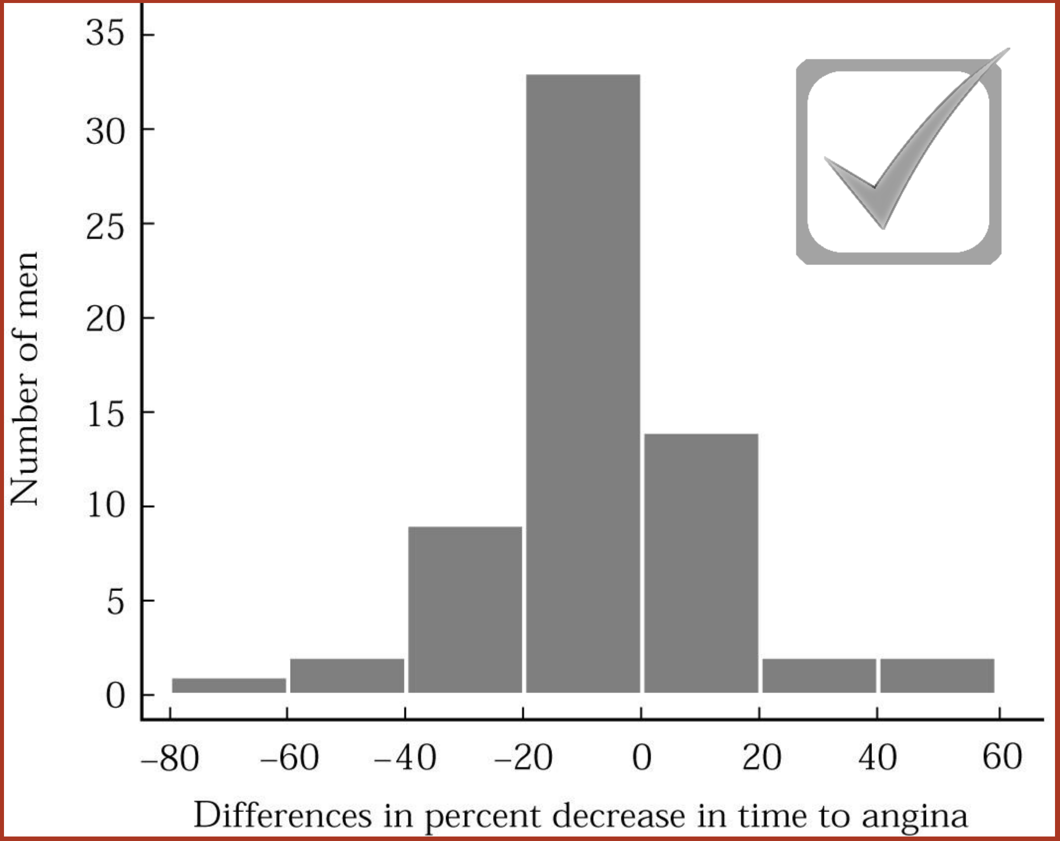

```
>> tinv(0.975,62)
ans = 1.9990
```




```
>> tinv(0.05,62)
ans =-1.6698
```

Checking the required conditions for the paired observations case

- The validity of the results depends on the normality of the differences. Indeed we have a fairly normal distribution here!



Matched Pairs Experiment – implementation

- Solution (given the following statistics)

$$\bar{d} = \frac{\sum_{i=1}^{63} d_i}{63} = -6.63$$

$$s_d = 20.29$$

It is a negative value, which is good. But is it negative enough to reject the null hypothesis, so that we may ascertain that CO-environment promotes angina onset?

  – Calculate t

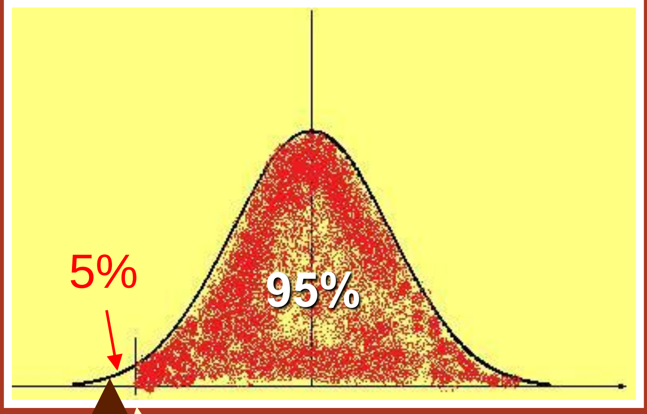
$$t = \frac{\bar{d} - \delta}{s_d / \sqrt{n}} = \frac{-6.63 - 0}{20.29 / \sqrt{63}} = -2.59$$

<u>Rejecting the null hypothesis, since it is further to the left tail of t=-1.6698.</u>

Computing p=P(t<–2.59)

- Although we can immediately know to **reject** the hypothesis when obtaining $t = -2.59$ which is smaller than $t_{.05,62} = -1.6698$, we can still use MATLAB to compute the p-value for this sample.
- Below we use MATLAB in performing this computation.

```
>> tcdf(-2.59,62)
ans =
   0.0060
```




```
>> tinv(0.05,62)
ans =-1.6698
```

Matched Pairs Experiment – Conclusion

- There is a significance difference of the mean percent decreases in time to angina between two exposures. [statistically different~~~]
- Sample 2 apparently delivers larger percent decreases in time, suggesting patients exposed to CO would likely to have a quicker onset of angina.

$$H_0: (\mu_1 - \mu_2) \geq 0$$

$$H_A: (\mu_1 - \mu_2) < 0$$

Example 3 – BP changes upon using OC?

N	Systolic bp with no OC (x_1)	Systolic bp with OC (x_2)	$d = x_2 - x_1$
1	115	128	13
2	112	115	3
3	107	106	-1
4	119	128	9
5	115	122	7
6	138	145	7
7	126	132	6
8	105	109	4
9	104	102	-2
10	115	117	2

```
>> x1=[115 112 107 119 115 138 126 105 104 115];
>> x2=[128 115 106 128 122 145 132 109 102 117];
>> d=x2-x1 =
   13    3   -1    9    7    7    6    4   -2    2
```

- The null hypothesis would be d=0.
- This is a ***2-tailed test***.
- This is a t-test since no population standard deviation is available.
- Degree of freedom is 10-1=9.
- Level of significant α = 0.05.

```
>> mu=mean(d)
mu =    4.8000

>> std=std(d)
std =    4.5656

>> t=(mu-0)/(std/sqrt(10))
t =    3.3247

>> 2*(1-tcdf(t,9))
ans =    0.0089
>>
```

Since p is smaller than 0.05, we conclude that bp will change significantly (regardless increasing or decreasing) when using OC.

Comments

- This example problem asked "whether BP **changes** upon using OC". This is a 2-sided test since we did not ask whether BP will drop or rise.
- If the problem asked "whether BP **rises** upon using OC", then this would be a 1-sided test with a p-value half of what we obtained earlier. Still we'd reject the null hypothesis that the BP change equals to zero.

A Brief Summary

- Here we covered "Chapter 11.1 – Paired Samples":
  - For **each** observation on the first group, there is a **corresponding** observation in the second group. So we may reduce the problem to a **one-sample** test problem (using one difference of the mean, instead of two means from two samples).
  - In this case, the two samples are **paired** (or **dependent**).

11.2 Independent Samples

- In many cases, however, the two groups of measurements of interests are **not** paired. 【cross-sectional study…】
- For example, we are interested in the serum iron level of two groups of children:
  - One group healthy (with mean μ_1)
  - The other group suffering from cystic fibrosis (囊腫性纖維化) (with mean μ_2)
  - Two types of populations are **independent** and normally distributed.

The Null Hypothesis

- We wonder **if the two population means are equal**. (That is, if the difference of serum iron level would **attribute** to this hereditary disease or not).
- So the null hypothesis would be

$$\boxed{H_0 : \mu_1 - \mu_2 = 0}$$

or

$$\boxed{H_0 : \mu_1 = \mu_2}$$

		Group 1	Group 2
Population	Mean	μ_1	μ_2
	Standard Deviation	σ_1	σ_2
Sample	Mean	$\bar{x}_1$ =18.9	$\bar{x}_2$ =11.9
	Standard Deviation	s_1 =5.9	s_2 =6.3
	Sample Size	n_1 =9	n_2 =13

- Two samples each was drawn from these two normally distributed populations, each with its own mean, STD and sample size.
- We'd like to perform a *t*-test on the proposed null hypothesis

$$\boxed{H_0 : \mu_1 - \mu_2 = 0}$$

11.2.1 Equal Variances

- We first **assume that the two population variations are identical** (i.e., $\sigma_1 = \sigma_2 = \sigma$)
- By **central limit theorem**, we can perform the following conversion where the normally distributed $\overline{x}$ would become a standard normal distribution z.

$$z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$$

Cont'd

- Since we are dealing with the samples from two independent normal distribution, we utilize **an extension of the central limit theorem** that the difference in means $\bar{x}_1 - \bar{x}_2$ is approximately normal with mean $\mu_1 - \mu_2$ and standard error $\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$
- Since $\sigma_1 = \sigma_2 = \sigma$, we now have the z conversion as

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

Before…

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2/n_1 + \sigma^2/n_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2[(1/n_1) + (1/n_2)]}}$$

Cont'd

- **If the population variance σ^2 is known**, we may use this *z*-statistics to test the proposed null hypothesis, as we have done before.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2/n_1 + \sigma^2/n_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\boxed{\sqrt{\sigma^2[(1/n_1) + (1/n_2)]}}}$$

$$\frac{\sigma}{\sqrt{n}}$$

$$\cancel{\frac{\sigma}{\sqrt{n_1 + n_2}}}$$

Note the correct formula for the denominator (分母)

Using t-statistics

- As noted earlier, **it is much more common that the true value of σ^2 is not known**. In this case, **we use t-statistics** instead of z-statistics:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 [(1/n_1) + (1/n_2)]}}$$

where the **new quantity s_p is a "pooled" estimate of the variance**, which replaces σ^2 used in the z-statistics.

Computing s_p^2

- The pooled estimate of the variance combines information from both samples to produce a more reliable estimate for the variance σ^2. (Recall we still **assume that the two population variations are identical**)
- If we know all measurements in the samples, we may compute it by

$$s_p^2 = \frac{\sum_{i=1}^{n_1} (x_{i1} - \overline{x}_1)^2 + \sum_{j=1}^{n_2} (x_{j2} - \overline{x}_2)^2}{n_1 + n_2 - 2} \quad ❶$$

Similar to the formula of variance given before.

Cont'd

- If the two standard variations (s_1 and s_2) of two samples are known, the previously formula can be easily replaced by

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \qquad ❷$$

One can see that this is actually a **weighted average** of the two sample variances s_1^2 and s_2^2.

Example 4

- Consider the two-group children example mentioned earlier:

		Group 1	Group 2
Population	Mean	μ_1	μ_2
	Standard Deviation	σ_1	σ_2
Sample	Mean	$\bar{x}_1$ =18.9	$\bar{x}_2$ =11.9
	Standard Deviation	s_1 =5.9	s_2 =6.3
	Sample Size	n_1 =9	n_2 =13

We'd like to test whether children with cystic fibrosis (group 2) would have a level of iron in their blood on average as healthy children (group 1). There we test the null hypothesis that the two population means are identical.

Solution

- The null hypothesis states that there is no difference in the underlying population mean iron levels for the two groups of children.
- **Two-sided t-test** will be performed, with a preset level of significance α=0.05.

$$\boxed{H_0 : \mu_1 - \mu_2 = 0}$$

- We first compute the **pooled estimate of the variance** using formula 2 we have seen earlier.

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

$$= \frac{(9 - 1)(5.9)^2 + (13 - 1)(6.3)^2}{9 + 13 - 2}$$

$$= \frac{(8)(34.81) + (12)(39.69)}{20}$$

Effective degree of freedom when using t-correction

$$= 37.74.$$

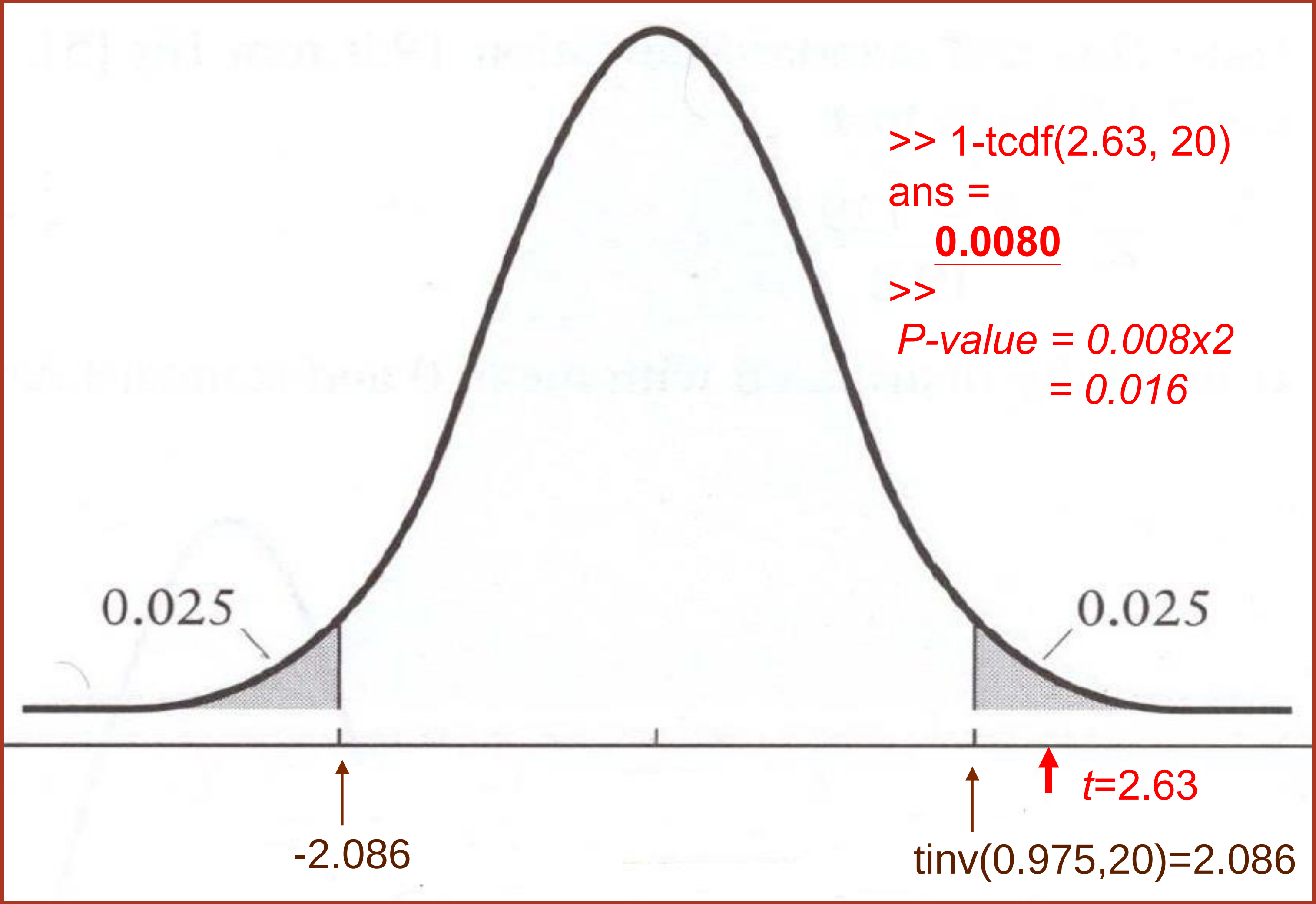
- With this, we can now compute for the t-statistics:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \boxed{(\mu_1 - \mu_2)}}{\sqrt{s_p^2\left[(1/n_1) + (1/n_2)\right]}}$$

$$= \frac{(18.9 - 11.9) - \boxed{0}}{\sqrt{(37.74)\left[(1/9) + (1/13)\right]}}$$

$$= 2.63.$$

Two-sided distribution (df=20)



Conclusion

- We computed for a p-value of 0.016. **This is smaller than α=0.05 for us to reject the null hypothesis.**
- That is, the **difference** between the mean serum iron level of healthy children (group 1) and ones with cystic fibrosis (group 2) is **statistically significant**.
- Based on these samples, it appears that children with cystic fibrosis **suffer** from an iron deficiency.

Estimating 95% CI for this difference of mean

$$P\left(-2.086 \le \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2[(1/n_1) + (1/n_2)]}} \le 2.086\right) = 0.95.$$

So the 95% CI for $\mu_1 - \mu_2$ can be computed as

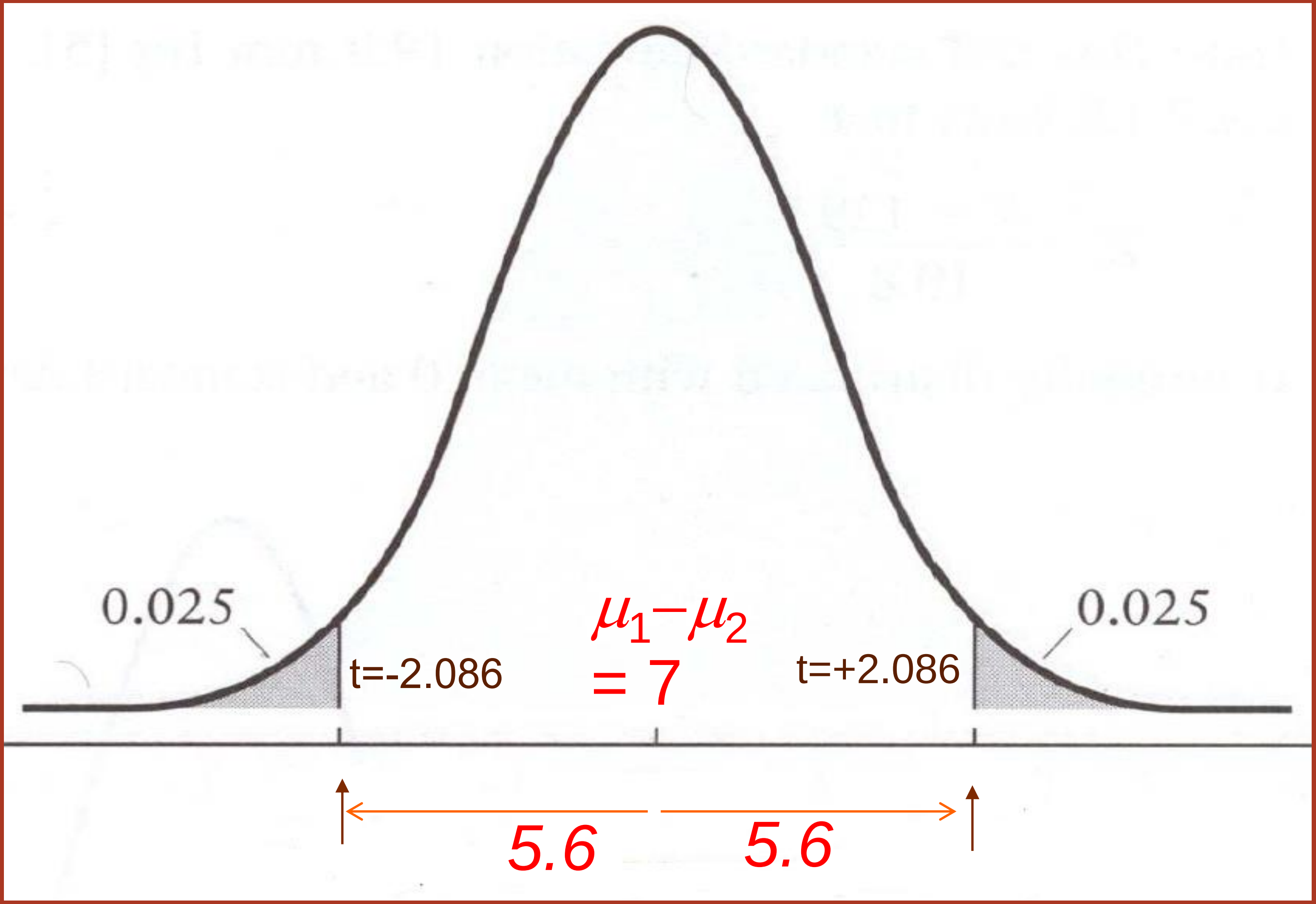
$$(\bar{x}_1 - \bar{x}_2) \pm (2.086)\sqrt{s_p^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]},$$

, or

$$(18.9 - 11.9) \pm (2.086)\sqrt{(37.74)\left[\frac{1}{9} + \frac{1}{13}\right]}.$$

Cont'd

- That is, we are 95% confident that **the interval (1.4, 12.6)** covers $\mu_1 - \mu_2$, the true difference in mean serum iron levels for the two populations of children.
- Note that **the interval does not contain the value 0** (meaning that $\mu_1 = \mu_2$), which is consistent with the result from the hypothesis testing we performed earlier.


0.025
0.025
$\mu_1 - \mu_2$
$= 7$
t=-2.086
t=+2.086
5.6
5.6

Summary

- The analysis we have seen in this lecture will pave way into Chapter 12, where we will do **analysis of variance comparing two or more populations** of interval data.
- This is an extension of the two-sample *t*-test we have seen here to **three or more** samples.

11.2.2 Unequal Variances?

- What if the two datasets do not have equal variance?
- In other words, earlier mentioned using s_p^2 as an estimate of he common variance σ^2 won't be valid.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

$$= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2[(1/n_1) + (1/n_2)]}}$$

- A *t*-test must be used, and the standard error term now uses s_1 and s_2 rather than σ_1 and σ_2.
- The exact t-distribution is difficult to derive, and an approximation must be used to compute the "***effective***" degree of freedom to use the conventional *t*-distribution..

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}.$$

- The value of *v* below is rounded to the nearest integer, which will serve as the degree of freedom when we approximate this distribution to a t-distribution.

$$v = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{[(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)]};$$

Example 5

- Effect of an anti-hypertensive drug against persons >= 60 yr who suffered from systolic blood pressure over 160 mmHg.
- Group 1
  - subjects receiving **active drug** for one year.
  - Mean systolic BP = μ_1, unknown variation.
- Group 2
  - subjects receiving **placebo** for one year
  - Mean systolic BP = μ_2, unknown variation and is different from group 1.

- H_0: $\mu_1 = \mu_2$
- H_A: $\mu_1 \neq \mu_2$

		Group 1	Group 2
Population	Mean	μ_1	μ_2
	Standard Deviation	σ_1	σ_2
Sample	Mean	$\bar{x}_1$ =142.5	$\bar{x}_2$=156.5
	Standard Deviation	s_1 =15.7	s_2=17.3
	Sample Size	n_1 =2308	n_2=2293

Compute the test statistics:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$$

$$= \frac{(142.5 - 156.5) - 0}{\sqrt{[(15.7)^2/2308] + [(17.3)^2/2293]}}$$

$$= -28.74.$$

Compute the “effective” degree of freedom:

$$v = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{[(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)]}$$

$$= \frac{[(246.49/2308) + (299.29/2293)]^2}{[(246.49/2308)^2/(2308 - 1) + (299.29/2293)^2/(2293 - 1)]}$$

$$= 4550.5.$$ ⇨ 4550 (rounding **down** to nearest integer)

- We know t-distribution of df=4550 is practically the same as normal distribution.
- t = -28.74 is far left to z = -1.96 which cuts the left tail at 2.5%. This concludes the difference is significant.

- One may now compute twice the value of tcdf(-28.74, 4550), which is the p-value for this test. The result is zero.
- This concludes the patients receiving the drug do have a lower mean systolic blood pressure than those receiving placebo.
- We may compute the 95% confidence interval of the difference of BP between these two samples as (-15.0, -13.0). [Verify this by yourself!!!]
- This too helps in rejecting the hypothesis that two means are equal.

		Group 1	Group 2
Population	Mean	μ_1	μ_2
	Standard Deviation	σ_1	σ_2
Sample	Mean	$\bar{x}_1 = 142.5$	$\bar{x}_2 = 156.5$
	Standard Deviation	$s_1 = 15.7$	$s_2 = 17.3$
	Sample Size	$n_1 = 2308$	$n_2 = 2293$

- Note that the sample sizes are big. This effectively raises DF and the absolute value of t (ideal conditions to reject the null hypothesis).
- ***What if the sample sizes are small, say, 10 each?***

```
>> x1=142.5; x2=156.5; s1=15.7; s2=17.3;
>> t=((x1-x2)-0)/sqrt(s1^2/10+s2^2/10)
t =  -1.8950

>> v=((s1^2/10+s2^2/10)^2)/((s1^2/10)^2/9+(s2^2/10)^2/9)
v =   17.8331

>> 2*tcdf(t,v)
ans =    0.0744
```

With this p-value, we cannot ascertain the two samples are having different mean values.