STATISTICAL METHODS

Topic : Paired T-Test

GROUP 7

會計所　黃敬涵　R16114016

數據所　曾文海　RE6124027

統計所　洪瑞廷　R26121015

統計所　李宗祐　R26124047

Content

01 Assumption / Limitation

02 Purpose of the test

03 What are type of data?

04 Null Hypothesis Alternative Hypothesis

05 Test statistic

06 Coding in R

Purpose of the test

- Each subject is measured twice, resulting in pairs of observations.

↓

- Determine whether the mean difference between two sets of observations is zero.
- If it equals to zero, it means no effect.

Pre-test Value VS Post-test Value

Assumption / Limitation

- The observations are defined as the **differences** between two sets of values, and each assumption refers to these differences.

Assumption :

- Observations must be **independent**.
- The dependent variable must be **continuous**.
- The dependent variable should be approximately **normally** distributed.
- The dependent variable should **not** contain any **outliers**.

What are type of data?

The difference of the data must be **continuous/numeric**.

學生	測驗 1 分數	測驗 2 分數	分數差
Bob	63	69	6
Nina	65	65	0
Tim	56	62	6
Kate	100	91	-9

Null Hypothesis
Alternative Hypothesis

Null Hypothesis: The difference between two groups is zero.

Alternative Hypothesis: The difference between two groups is not zero.

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

$$H_1 : \mu_d > 0$$

$$H_1 : \mu_d < 0$$

Test statistic

- D = Differences between two paired samples
- d_i = The i^{th} observation in D
- N = The sample size
- $\bar{d}$ = The sample mean of the differences
- $\hat{\sigma}$ = The sample standard deviation of the differences
- T = The critical value of a t-distribution with (n − 1) degrees of freedom
- t = The t-statistic (t-test statistic) for a paired sample t-test
- p = The p-value (probability value) for the t-statistic.

1. Calculate the sample mean.

$$\bar{d} = \frac{d_1 + d_2 + \cdots + d_n}{n}$$

2. Calculate the sample standard deviation.

$$\hat{\sigma} = \sqrt{\frac{(d_1 - \bar{d})^2 + (d_2 - \bar{d})^2 + \cdots + (d_n - \bar{d})^2}{n-1}}$$

3. Calculate the test statistic.

$$t = \frac{\bar{d} - 0}{\hat{\sigma}/\sqrt{n}}$$

4. Calculate the probability of observing the test statistic under the null hypothesis.

$$p = 2 \bullet \Pr(T > |t|) \quad (\text{two} - \text{tailed})$$

$$p = \Pr(T > t) \quad (\text{upper} - \text{tailed})$$

$$p = \Pr(T < t) \quad (\text{lower} - \text{tailed})$$

Coding in R – data description

- This dataset is composed of 3 variables: ID, X, Y.
- X is the score of pre-test.
- Y is the score of post-test.

ID	X : pre-test score	Y : post-test score
1	-2.6	-2.1
2	-2.2	-2.1
3	-2.8	-1.9
4	-2.1	-2.3
⋮	⋮	⋮
26	-2.7	-2.8
(μ, σ)	$(-2.038, 1.45)$	$(-1.896, 1.54)$

Coding in R

```
library(stats)
# x is before ; y is after
x <- data1017$x
y <- data1017$y
# discription of data
summary(data1017)
```

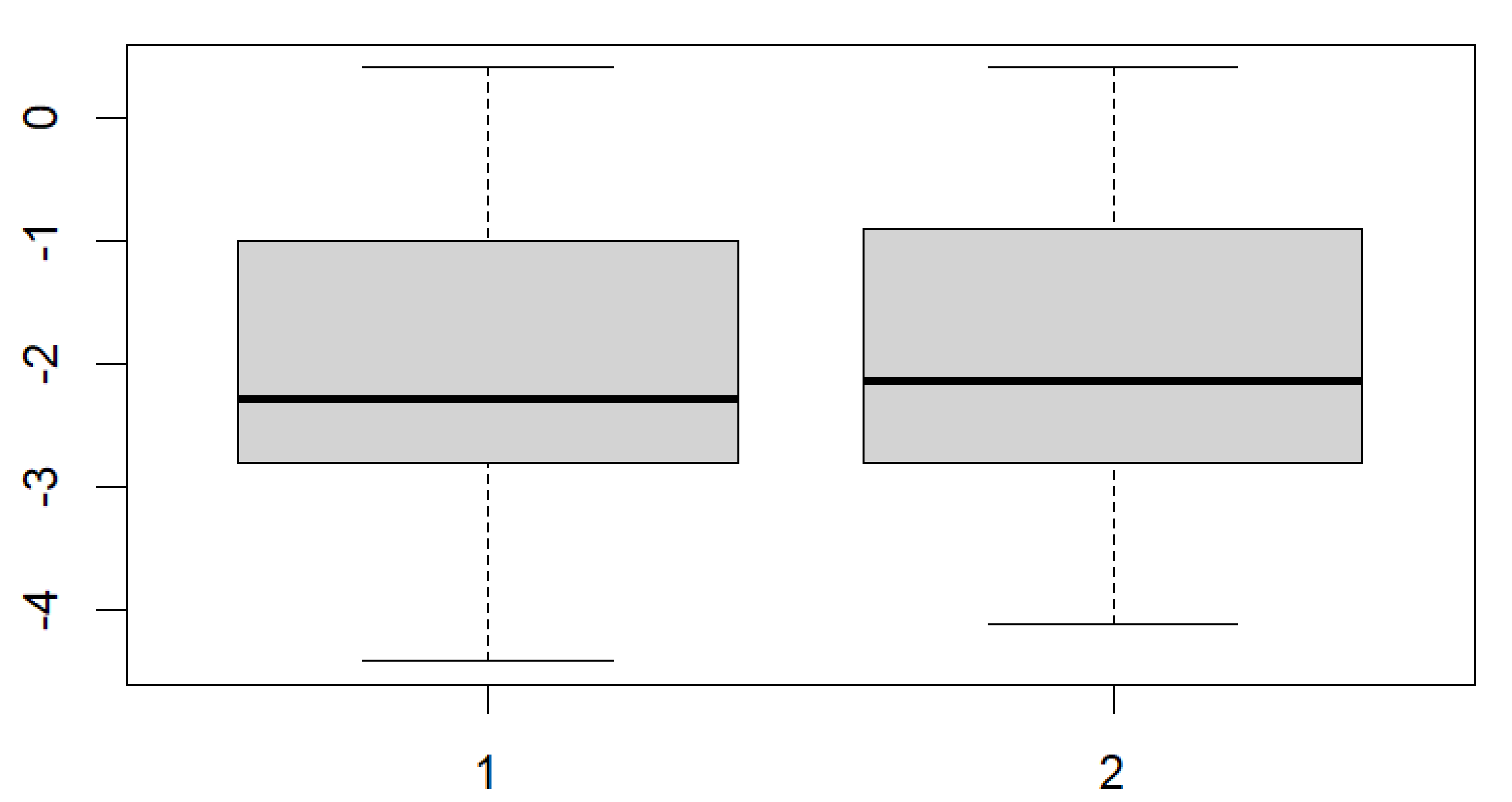
mean of X is not equal to mean of Y,
so we know that H_1 is $\mu_x - \mu_y \neq 0$.

```
> summary(data1017)
       ID              x                 y         
 Min.   : 1.00   Min.   :-4.400   Min.   :-4.100  
 1st Qu.: 7.25   1st Qu.:-2.775   1st Qu.:-2.750  
 Median :13.50   Median :-2.300   Median :-2.150  
 Mean   :13.50   Mean   :-2.038   Mean   :-1.896  
 3rd Qu.:19.75   3rd Qu.:-1.025   3rd Qu.:-0.925  
 Max.   :26.00   Max.   : 0.400   Max.   : 0.400  
```

Coding in R

```
# check outlier
boxplot(x, y)
```

From the boxplot,
we find that there is not outlier in x and y



Coding in R

```r
# chck if x and y follow normal distribution
shapiro.test(x)
shapiro.test(y)
```

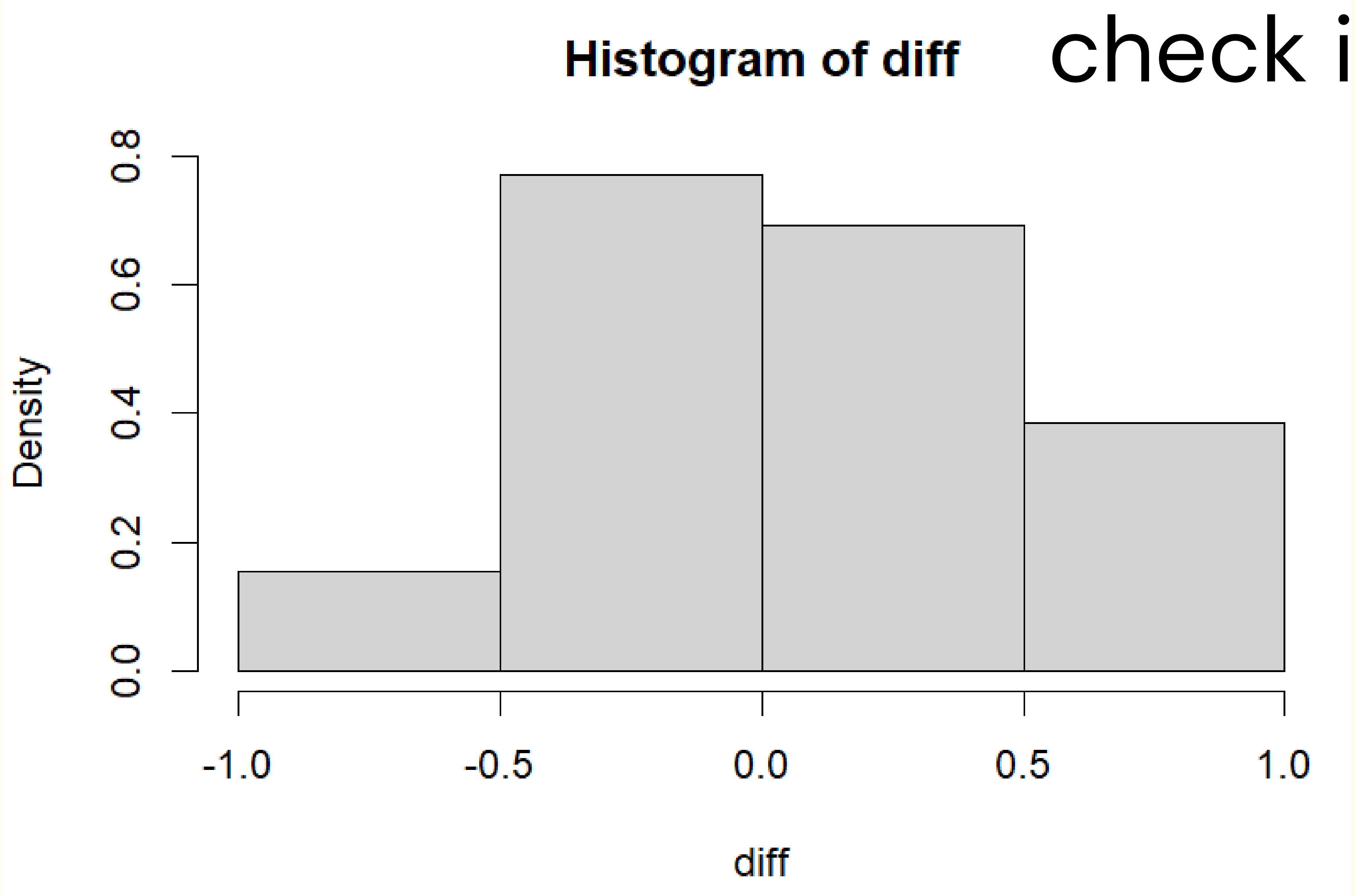
```
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.96417, p-value = 0.4355

> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.96485, p-value = 0.4512
```

We check whether x and y follow normal distribution

(H_0: follow normal distribution)

Coding in R

```r
# define minus of before and after
diff <- y - x
hist(diff, probability = TRUE)
```



check if the diff is (nearly) symmetric.

Coding in R

```
# paired t test
t.test(x, y, paired = TRUE, alternative = 'two.sided')
```

```
> # paired t test
> t.test(x, y, paired = TRUE, alternative = 'two.sided')

        Paired t-test

data:  x and y
t = -0.17532, df = 27, p-value = 0.8621
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -0.3175855  0.2675855
sample estimates:
mean difference
         -0.025
```

paired t test：

p-valued >0.05

so we do not reject H_0,

i.e, μ_x and μ_y has no significant difference.

Reference

- https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/paired-sample-t-test/
- https://www.jmp.com/zh_tw/statistics-knowledge-portal/t-test/paired-t-test.html
- https://www.yongxi-stat.com/paired-sample-t-test/

THANK YOU