Group 3

Independent T-test

Member : 張立勳　陳沛群　鄭柏武　黃筱云

Table of contents

1.

Assumptions

Assumptions

Independence of the observations：

Observations are independent within each group and between groups.

Normality：

The data for each group should be approximately normally distributed.

Homogeneity of variances：

The variability of the data in each group is similar.

2.

Purpose of test & type of data

Purpose of test & type of data

Compare two sample means from unrelated groups

To determine if the samples are different from each other

Data type : numerical

3.

Null hypothesis
&
Alternative hypothesis

Null hypothesis & Alternative hypothesis

Left-tailed test

$H_0: \mu_1 - \mu_2 \leq c$
$H_1: \mu_1 - \mu_2 \nleq c$
where $c \in \mathcal{R}$

Two-tailed test

$H_0: \mu_1 - \mu_2 = c$
$H_1: \mu_1 - \mu_2 \neq c$
where $c \in \mathcal{R}$

Right-tailed test

$H_0: \mu_1 - \mu_2 \geq c$
$H_1: \mu_1 - \mu_2 \ngeq c$
where $c \in \mathcal{R}$

Test statistic

$$T = \frac{\bar{X}_1 - \bar{X}_2 - c}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t\,(n_1 + n_2 - 2) \quad \text{under } H_0$$

Where

n_1, n_2 : sample size of observations in first/second group

$\bar{x}_1, \bar{x}_2$: sample mean of observations in first/second group

S_1^2, S_2^2 : sample variance of observations in first/second group

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \xrightarrow{estimate} \sigma_1^2 = \sigma_2^2 = \sigma^2$$

4.

Real Case

Real Case

Data source :

"**Water Quality and Potability**" from Kaggle

Data description :

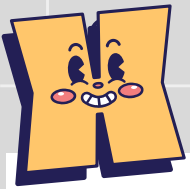
Turbidity level : a measure of water clarity.

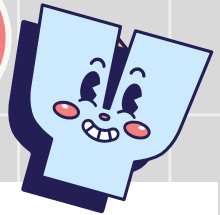
Potability : water potability with values 1 (potable) and 0 (not potable).

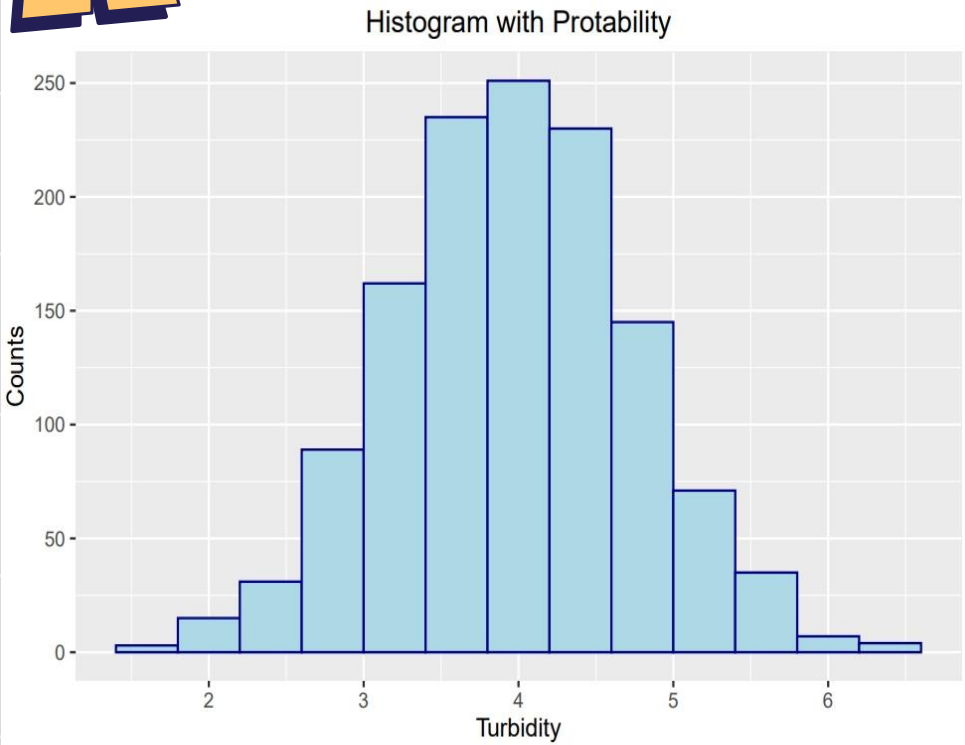
Purpose :

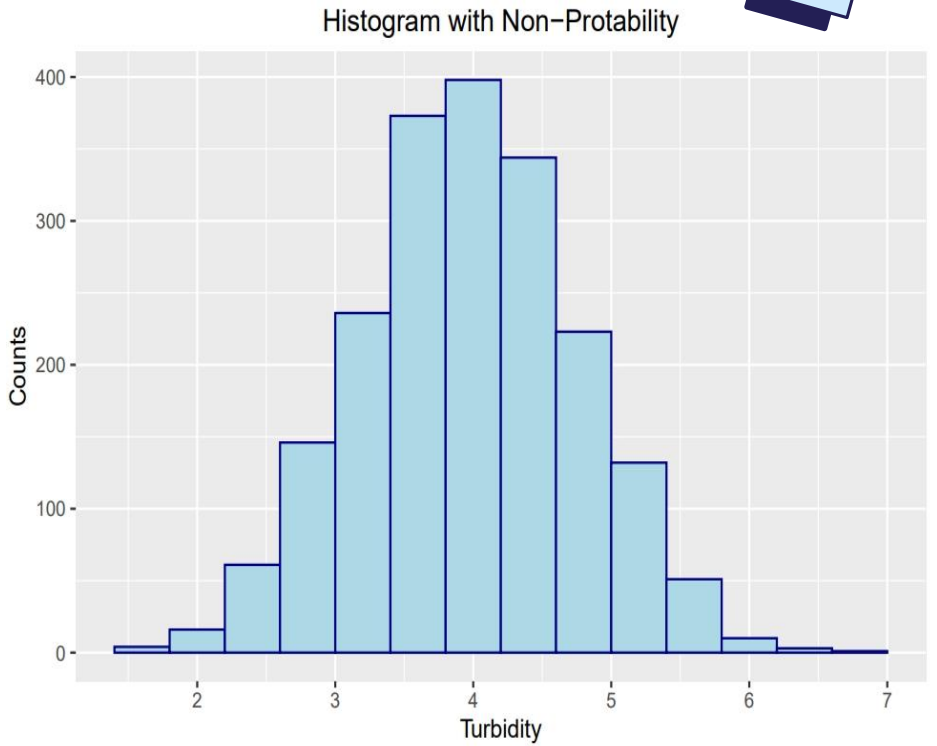
Determine whether the turbidity of potable water and non-potable water is significantly different.

Real Case






Histogram with Protability
Counts
250
200
150
100
50
0
2
3
4
5
6
Turbidity


Histogram with Non−Protability
Counts
400
300
200
100
0
2
3
4
5
6
7
Turbidity

Real Case

$\mathbf{H_0}$: The turbidity of potable water and non-potable water is not different.

$\mathbf{H_1}$: The turbidity of potable water non-potable water is different.

We set $\alpha = 0.05$ for following tests

Shapiro-Wilk tests

test normality for two samples.

F test

test if two variances are the same.

P-values of each test are summarized to the table (rounded to four decimal places).

Note-taking and study skills

	Sample 1	Sample 2
Shapiro-Wilk test	0.9296	0.9822
F test	0.9986	

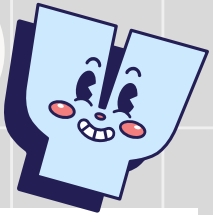
常態假設

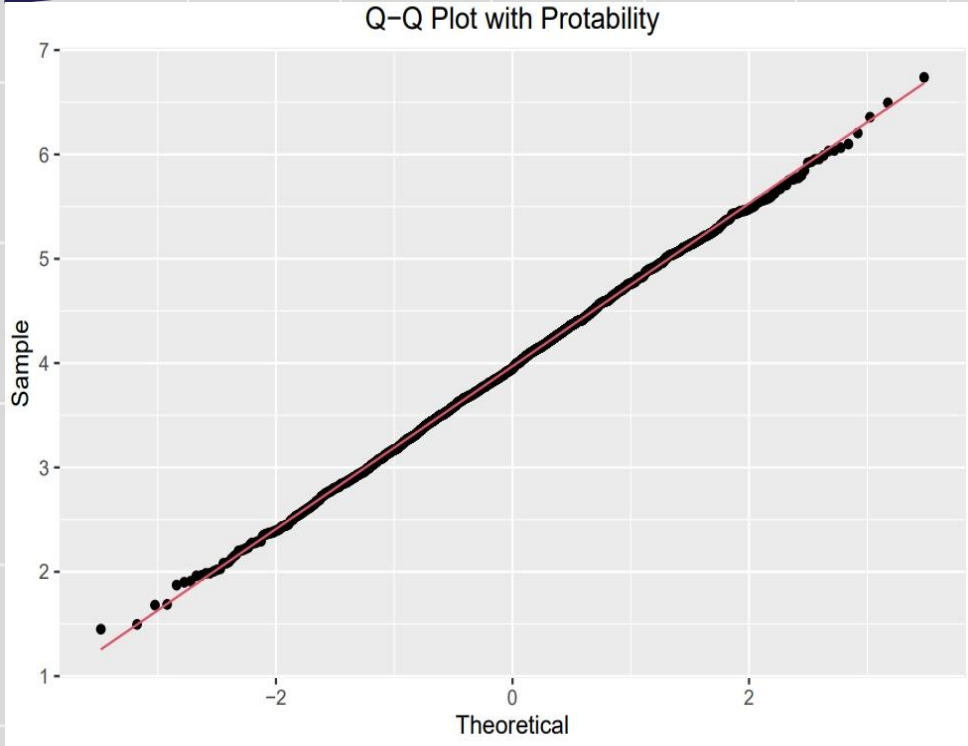
shapiro.test(data)

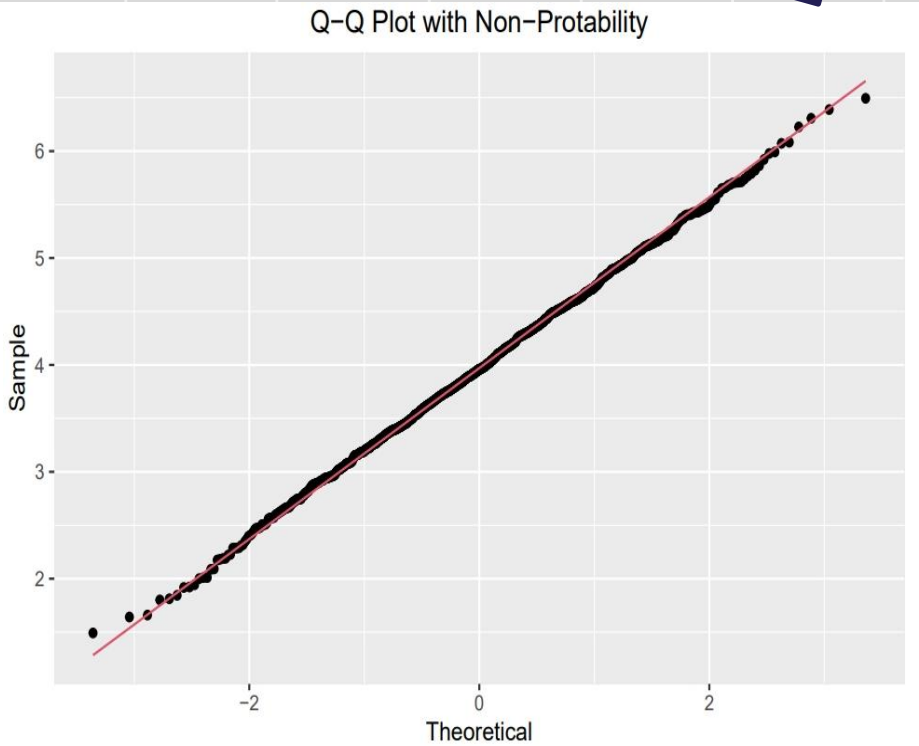
同質變異數

var.test(data1, data2, alternative = "two.sided")

Real Case




Q−Q Plot with Protability
Sample
Theoretical
7
6
5
4
3
2
1
−2
0
2


Q−Q Plot with Non−Protability
Sample
Theoretical
6
5
4
3
2
−2
0
2

Real Case

T-test

t.test(data1, data2, alternative ="two.sided",
var.equal = TRUE)

P-values :

0.9279

95 percent confidence interval :

[-0.0573 , 0.0523]

Conclusion :

The turbidity of potable water and non-potable water is not significantly different.

Job Assignment / Reference

張立勳	程式撰寫、資料整理
陳沛群	程式撰寫、資料整理
鄭柏武	資料找尋、資料整理
黃筱云	簡報製作、資料整理

https://www.kaggle.com/datasets/uom190346a/water-quality-and-potability

Thanks!