# Nonparametric Statistics

## Ch.3 Two-sample problem

# Motivation

- In two-sample problems, similar difficulties as in the one-sample problem may happen. (non-normal parent distribution, existence of outliers)
- A two-sample nonparametric method is an alternative to the two-sample t-test.
- Our interest is to see if the location parameters of two different distributions are the same or not.

# Review : Two-sample t-test

Let $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ be a random sample from $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$. A test for the hypothesis

$$H_0: \mu_X = \mu_Y \qquad vs \qquad H_1: \mu_X \neq \mu_Y$$

can be performed with the test statistic

$$T = \frac{\bar{X} - \bar{Y}}{S_p / \sqrt{\frac{1}{m} + \frac{1}{n}}},$$

if $\sigma_X^2 = \sigma_Y^2$. Here, $S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$ is called the "pooled sample variance".

- T has the t-distribution with m+n-2 degrees of freedom under $H_0$.
- Rejection region for the test at the significance level $\alpha$ is $\{t: |t| > t_{\alpha/2}(m+n-2)\}$ where $t_{\alpha/2}(m+n-2)$ satisfies $P(T > t_{\alpha/2}(m+n-2))= \alpha/2$ with $T \sim T(m+n-2)$.
- p-value is computed by $P(|T| > |t_0|)$ where $t_0$ is the observed value of the test statistic.

Note] In the case that the alternative is one-sided like $H_1: \mu_X > \mu_Y$ or $H_1: \mu_X < \mu_Y$, the rejection region and the p-value will be given as:

$H_1: \mu_X > \mu_Y$: $\{t: t > t_\alpha(m+n-2)\}$ and $P(T > t_0)$

$H_1: \mu_X < \mu_Y$: $\{t: t < t_\alpha(m+n-2)\}$ and $P(T < t_0)$

# Review : Two-sample t-test

- This test is based on the assumption that two random samples are independent. If not, we need to consider a different method such as paired t-test.
- Two-sample t-test is the test for the difference of means.
- In the case of $\sigma_X^2 \neq \sigma_Y^2$, the test statistic is given as

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}}.$$

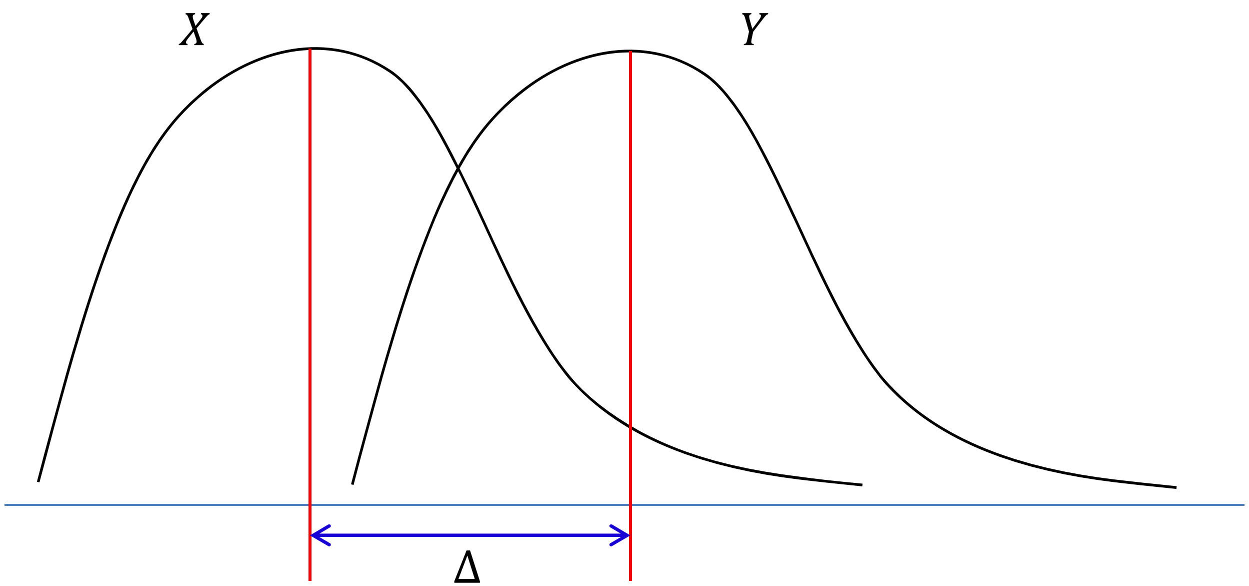
  And, this follows t-distribution with the degrees of freedom $t_{df}$ which has a very complicated form.
- If m and n are large, the test statistic approximately follows the standard normal distribution.

# Wilcoxon rank sum test

Let $X_1, \dots, X_m$ be a random sample from $F(\cdot)$, and $Y_1, \dots, Y_n$ be a random sample from $F(\cdot - \Delta)$.

- We assume that $F$ is continuous.
- The parameter $\Delta$ is called the **location shift**.
- This can be rewritten as

$$X + \Delta \equiv^d Y \quad , \qquad X \sim F$$



# Wilcoxon rank sum test

A test for the hypothesis

$$H_0 : \Delta = 0 \qquad vs \qquad H_1 : \Delta \neq 0$$

can be performed with the test statistic

$$W_n = \sum_{i=1}^{n} R_i \quad ,$$

where $R_i$ is the rank of $Y_i$ among m+n pooled observations.

- Note that $E(W_n) = n(m+n+1)/2$ under the null hypothesis. Therefore, farther values of $W_n$ from $n(m+n+1)/2$ support the alternative hypothesis.
- p-value can be computed by statistical packages.
- If the alternative hypothesis is $H_1 : \Delta > 0$, then large values of $W_n$ support $H_1$.
- Here, $\Delta$ need not be the difference of medians.
- Note that we assume that two distributions is the same in shape.

# Wilcoxon rank sum test : Example

Ex] The weight reduction of 10 patients were measured. They were separated into two groups, and assigned to two different dietary treatments. We want to see if there is a clear difference between two methods.

| | | | | | | |
|---|---|---|---|---|---|---|
| A | 5.7 | 7.3 | 7.6 | 6.0 | 6.5 | 5.9 |
| B | 4.9 | 7.4 | 5.3 | 4.6 | | |

(1) t-test

$$\bar{x}_A = 6.5\,, \bar{y}_B = 5.525\,, s_A^2 = 0.62\,, s_B^2 = 1.4825$$

Therefore, the pooled sample variance is

$$s_p^2 = \frac{5 \times 0.62 + 3 \times 1.4825}{8} = 0.9434.$$

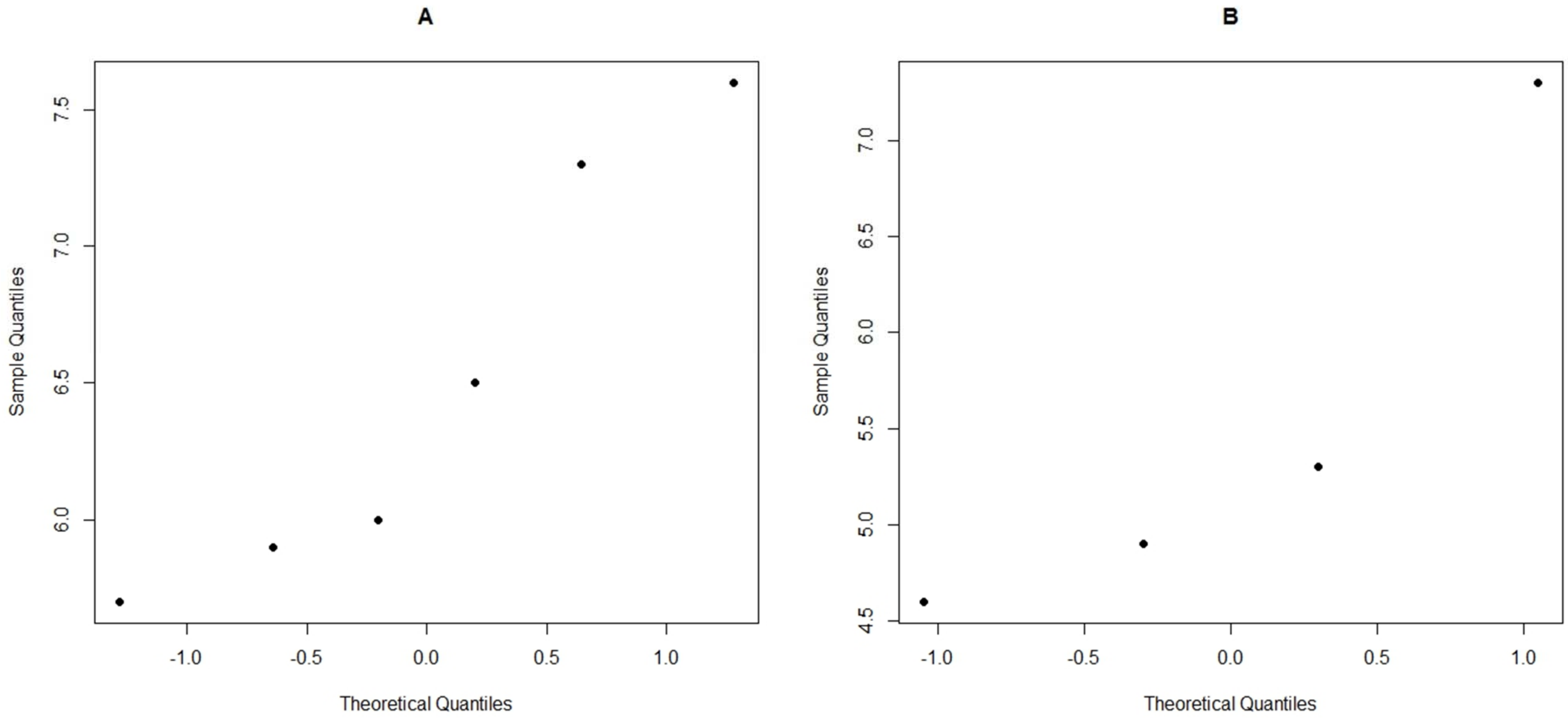
The observed test statistic is

$$\frac{6.5 - 5.525}{\sqrt{0.9434}\sqrt{\frac{1}{6} + \frac{1}{4}}} = 1.56.$$

The p-value is 0.1585.

# Wilcoxon rank sum test : Example

Normality assumptions???




=> They may come from non-normal distributions, though the samples are too small so that we cannot say too much about it.

# Wilcoxon rank sum test : Example

(2) Wilcoxon rank sum test

| | | | | | | |
|---|---|---|---|---|---|---|
| A | 5.7 | 7.3 | 7.6 | 6.0 | 6.5 | 5.9 |
| B | 4.9 | 7.4 | 5.3 | 4.6 | | |

– Sorting

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4.6 | 4.9 | 5.3 | 5.7 | 5.9 | 6.0 | 6.5 | 7.3 | 7.4 | 7.6 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

– The observed test statistic:

$$w_n = 1 + 2 + 3 + 9 = 15$$

– The p-value (by **R**) is 0.172. Therefore, we cannot reject the null hypothesis at the significance level 0.05.

# Wilcoxon rank sum test : large sample approximation

The test statistic for Wilcoxon rank sum test is

$$W_n = \sum_{i=1}^{n} R_i \,.$$

(1) $E(W_n) = n(m+n+1)/2$

(2) $\mathrm{Var}(W_n) = mn(m+n+1)/12.$

under the null hypothesis $H_0\colon \Delta = 0$.

It can be shown that

$$\frac{W_n - n(m+n+1)/2}{\sqrt{mn(m+n+1)/12}} \sim N(0,1) \text{ for sufficiently large } m, n.$$

under $H_0\colon \Delta = 0$. Therefore, we can test $H_0\colon \Delta = 0$ using the standard normal distribution.

# Wilcoxon rank sum test : large sample approximation (Example)

Ex1 : revisited]

| | | | | | | |
|---|---|---|---|---|---|---|
| A | 5.7 | 7.3 | 7.6 | 6.0 | 6.5 | 5.9 |
| B | 4.9 | 7.4 | 5.3 | 4.6 | | |

$$w_n = 15.$$

Then, the observed test statistic by the large sample approximation is

$$\frac{15 - 4 \times 11/2}{\sqrt{6 \times 4 \times 11\ /12}} = -1.49,$$

and therefore the p-value is

$$P(|Z| \geq 1.49) = 0.1362$$

where $Z$ stands for the standard normal distribution.

Note that the exact p-value is 0.172.

Note] Actually, the samples are too small to apply large sample theory in this example.

# Mann-Whitney test

A test for the hypothesis

$$H_0 : \Delta = 0 \qquad vs \qquad H_1 : \Delta \neq 0$$

can be performed with the test statistic

$$U_{m,n} = \sum_{j=1}^{m} \sum_{i=1}^{n} I(X_j < Y_i) \quad ,$$

- This test is called the "Mann-Whitney test". Actually, this test is equivalent to the Wilcoxon rank sum test. It can be easily verified from the definition of $R_i$.

$$R_i = \sum_{k=1}^{n} I(Y_k \le Y_i) + \sum_{j=1}^{m} I(X_j < Y_i) .$$

$$W_n = \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} \sum_{k=1}^{n} I(Y_k \le Y_i) + \sum_{i=1}^{n} \sum_{j=1}^{m} I(X_j < Y_i)$$

$$= \frac{n(n+1)}{2} + \sum_{i=1}^{n} \sum_{j=1}^{m} I(X_j < Y_i) = U_{m,n} + \frac{n(n+1)}{2}$$

# Mann-Whitney test : large sample approximation

The test statistic for Mann-Whitney test is

$$U_{m,n} = \sum_{j=1}^{m} \sum_{i=1}^{n} I(X_j < Y_i) .$$

(1) $E(U_{m,n}) = mn/2$

(2) $\mathrm{Var}(U_{m,n}) = mn(m+n+1)/12.$

under the null hypothesis $H_0 : \Delta = 0.$

It can be shown that

$$\frac{U_{m,n} - mn/2}{\sqrt{mn(m+n+1)/12}} \sim \ N(0,1) \ \text{ for sufficiently large } m, n.$$

under $H_0 : \Delta = 0$. Therefore, we can test $H_0 : \Delta = 0$ using the standard normal distribution.

# Mann-Whitney test : Example

Ex1 : revisited]

| | | | | | | |
|---|---|---|---|---|---|---|
| A | 5.7 | 7.3 | 7.6 | 6.0 | 6.5 | 5.9 |
| B | 4.9 | 7.4 | 5.3 | 4.6 | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4.6 | 4.9 | 5.3 | 5.7 | 5.9 | 6.0 | 6.5 | 7.3 | 7.4 | 7.6 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 0 | 0 | 0 | | | | | | 5 | |

The observed test statistic :

$$u_{m,n} = 5.$$

Note that, $15 = w_n = u_{m,n} + \frac{n(n+1)}{2} = 5 + 10$. Therefore, it gives the same result as the Wilcoxon rank sum test.

# Estimation associated with the Mann-Whitney statistic

- Define

$$D_{ij} = Y_i - X_j, \qquad 1 \le j \le m\,, 1 \le i \le n$$

- Point estimation

$$\hat{\Delta} = median\{D_{ij} : 1 \le j \le m\,, 1 \le i \le n\}$$

- Confidence interval : Let $k_\alpha$ be the integer satisfying $P\left(U_{m,n} \ge k_\alpha\right) = \alpha/2$, then $100(1-\alpha)\%$ confidence interval for $m$ is given as

$$\left(D_{(mn-k_\alpha+1)}, D_{(k_\alpha)}\right)$$

where $D_{(r)}$ denotes $r^{th}$ order statistic.

- When $n$ is large enough,

$$k_\alpha \approx \frac{mn}{2} + z_{\alpha/2}\left(\frac{mn(m+n+1)}{12}\right)^{1/2}.$$

# Asymptotic relative efficiency

- In this chapter, ARE shows how efficient the Mann-Whitney test is compared to the two sample t-test.
- If ARE is larger than 1, it means that the nonparametric method is preferable.
- The following table represents AREs of the Mann-Whitney test to the two sample t-test for some selected probability distributions.

| Dist. | Normal | Uniform | Logistic | Double exponential | Cauchy | t(3) | t(5) |
|---|---|---|---|---|---|---|---|
| ARE | 0.955 | 1.000 | 1.097 | 1.500 | $\infty$ | 1.900 | 1.240 |

- Note that this table is exactly the same as the table for the Wilcoxon signed rank test in the one-sample problem.