Psychol. Statistics using R

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Nonparametric Testing

Changwei W. Wu, Ph.D.

Graduate Institute of Mind, Brain and Consciousness
Research Center of Brain and Consciousness
Taipei Medical University

Statistics

1. **Recap of Distribution & Assumptions**
   - Violation of assumptions, Nonparametric: Ranks & Sign Test
2. **Nonparametric Tests**
   - Wilcoxon signed-rank & rank-sum tests
3. **[R] Hands-on Practices**

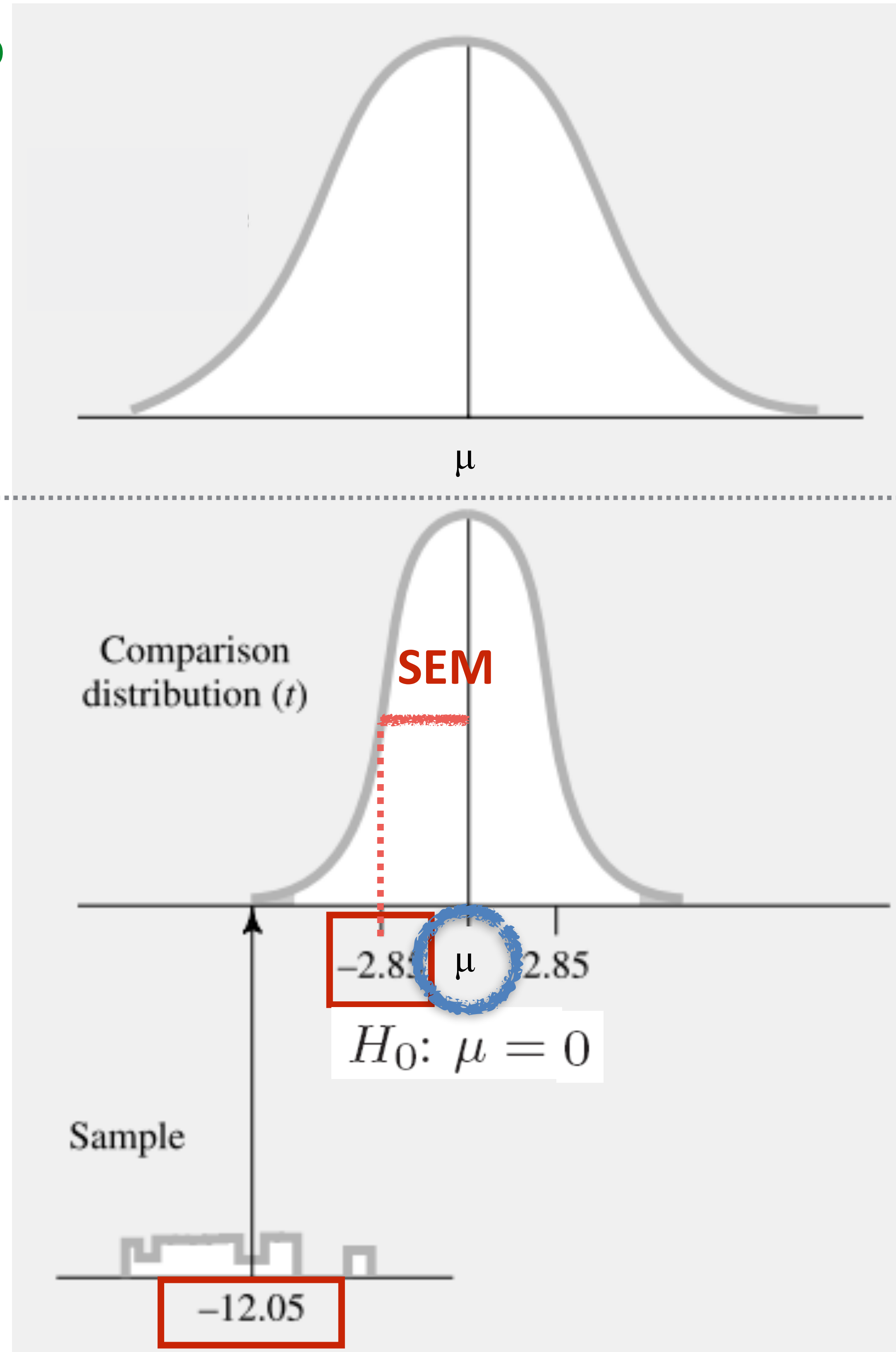
Theories

Practice

Assignment

Parametric Statistics

➡ Use the sample mean to compare with the population mean in **H_0 distribution**



Population

null hypothesis (H_0)

from current data for inference

Estimation

unreliability for estimation

from the presumed sampling distribution (e.g., t-distribution)

Sample

descriptive stat.

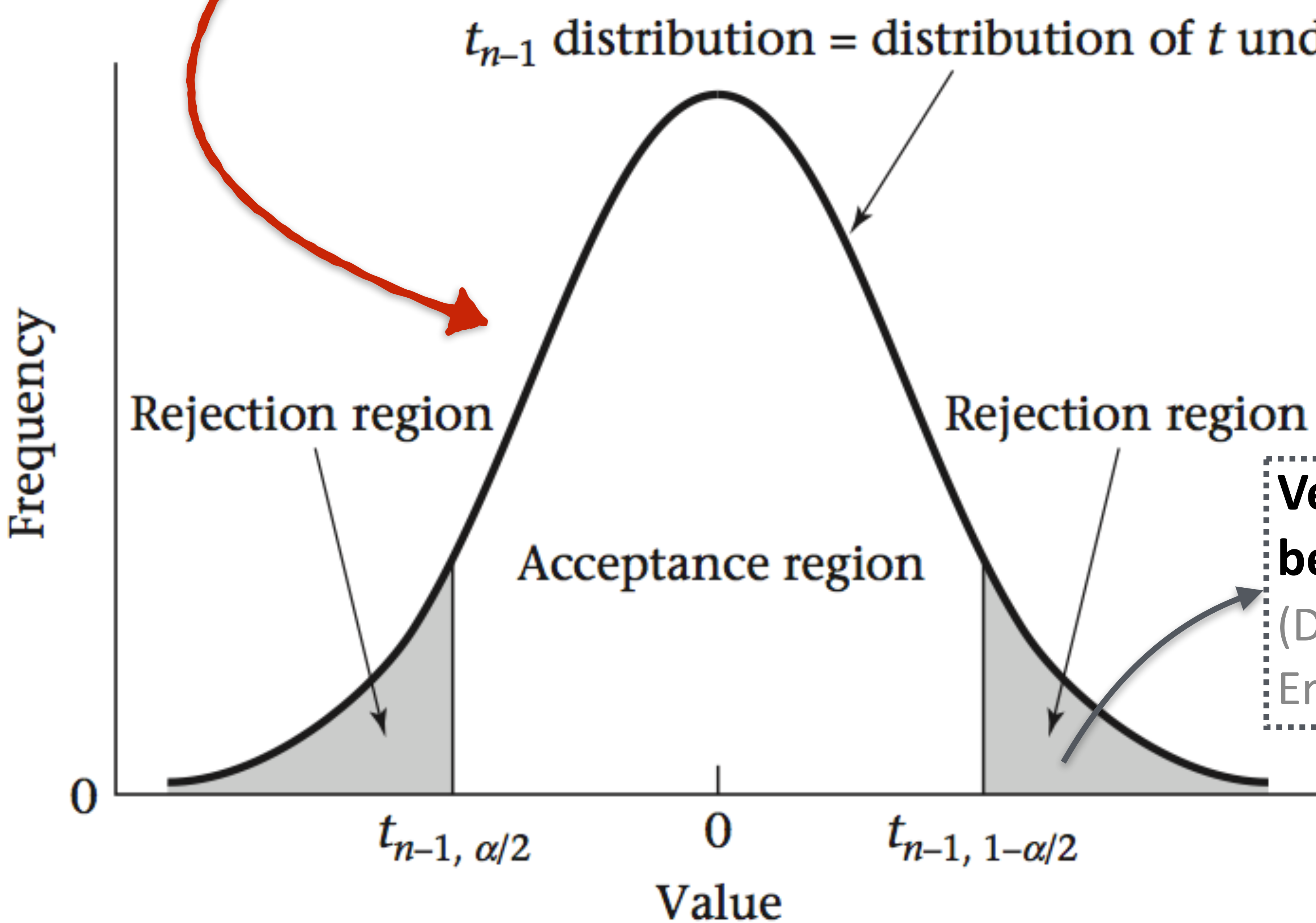
- **Compare the sample mean to 0 according to the SEM**

$$\frac{-12.05}{2.85} = -4.2$$

Parametric Testing

$$H_0\colon \mu = c \quad E(t) = 0$$
$$H_a\colon \mu \neq c \quad E(t) \neq 0$$

This prob. distribution corresponds to H_0



Decision

Very unlikely to be H_0

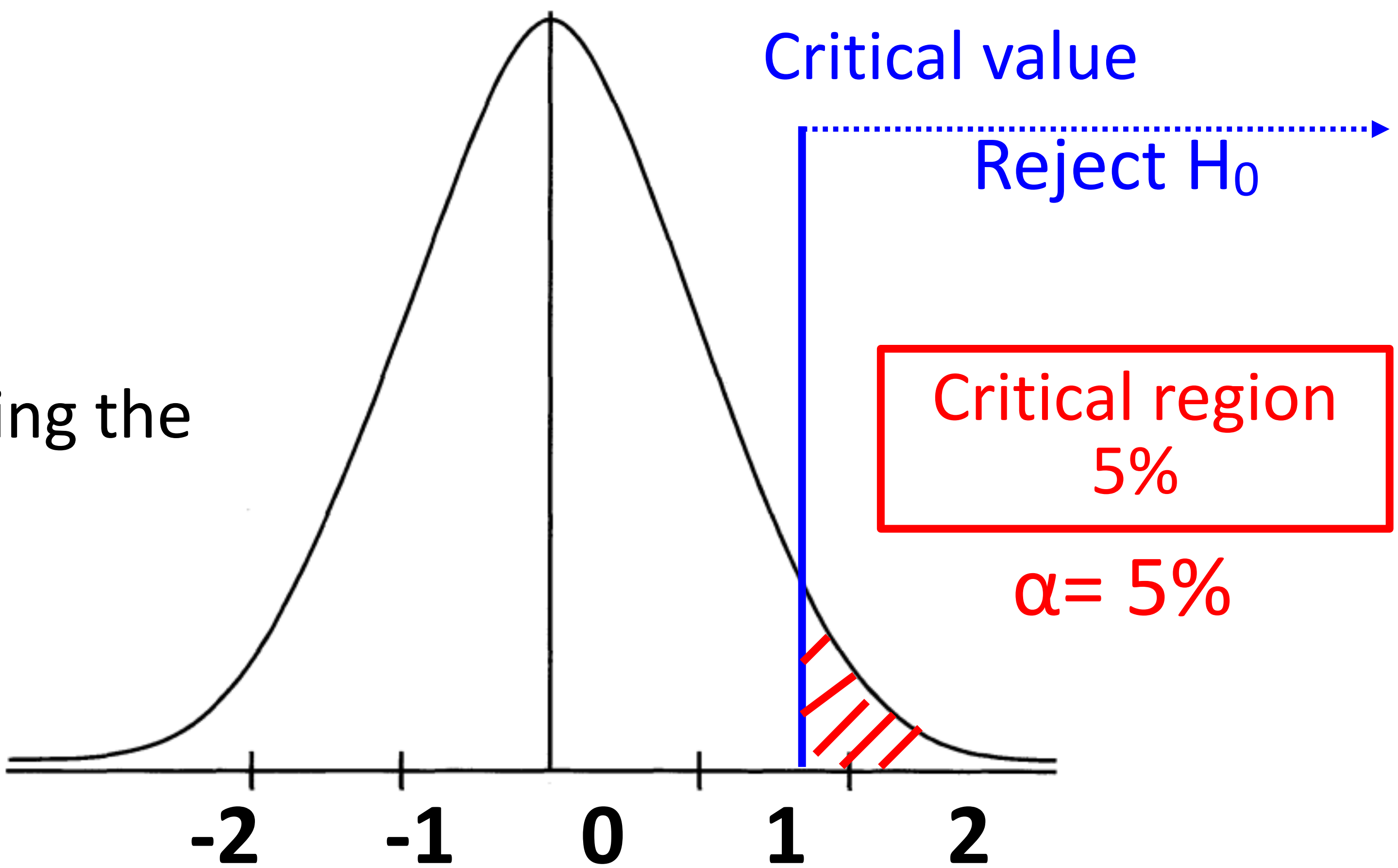
(Decision may be Erroneous)

Types of decision errors

The probability of a **type I error** is the probability of rejecting the null hypothesis when H_0 is true.

- **Type I error (α)**
  - ➡ Falsely support the research hypothesis
  - ➡ The chance of finding significant effects when no effects are in reality.
  - ➡ Control the probability by establishing the cutoff (critical value) or alpha level (p=.05, α=.05, critical region = 5%)

- We are willing to risk falsely supporting the research hypothesis 5% of the time.



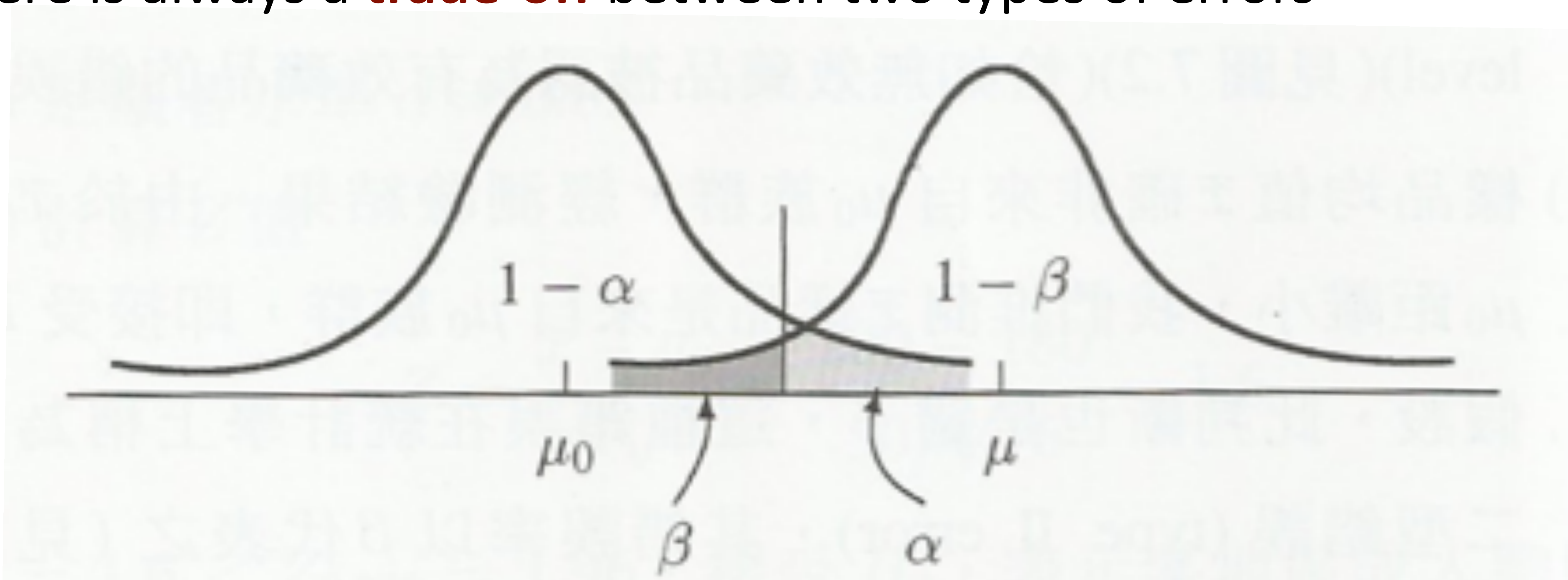
Types of decision errors

The probability of a **type II error** is the probability of accepting the null hypothesis when H_1 is true. This probability is a function of μ as well as other factors.

- **Type II error (β)**
  - ➡ Missing an effect that actually exists.
  - ➡ The chance of accepting no effect when difference exists in reality.
  - ➡ We know that α and **β** are inversely related:

    When α increases, **β** decreases.

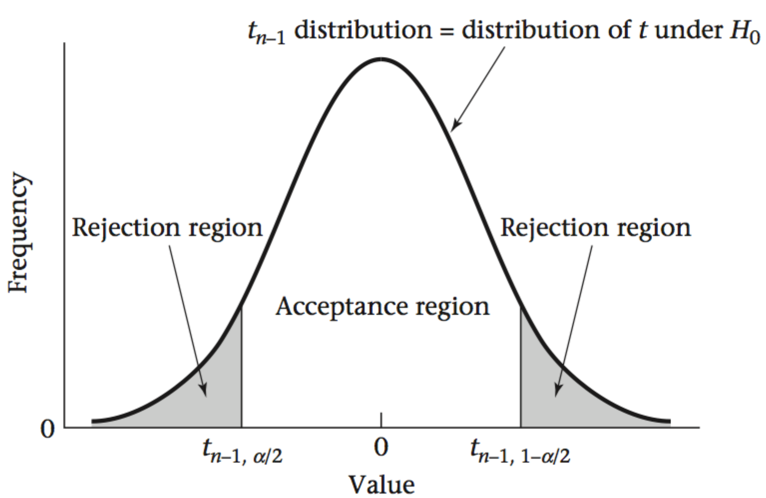
    When α decreases, **β** increases.
  - ➡ There is always a **trade-off** between two types of errors



Decision in Hypothesis Testing

- ***Type I error***
  - Falsely support the alternative hypothesis when there is no effect in reality.

- ***Power** (1 - Type II error)*
  - Testing can detect differences when there is actually an effect.



		Real Situation	
		Research Hypothesis is True	Null Hypothesis is True
Decision	Reject the Null Support Research	**Right**	**Wrong** Type I error
	Accept the Null Inconclusive	**Wrong** Type II error	**Right**

Assumptions of *t*-test

◉ ***Parametric tests***

- Based on Normal distribution (from original population);
- Data are measured at the **interval scales**.

➡(*independent t*-test) Scores in different conditions are independent.

◉ ***Normality:***

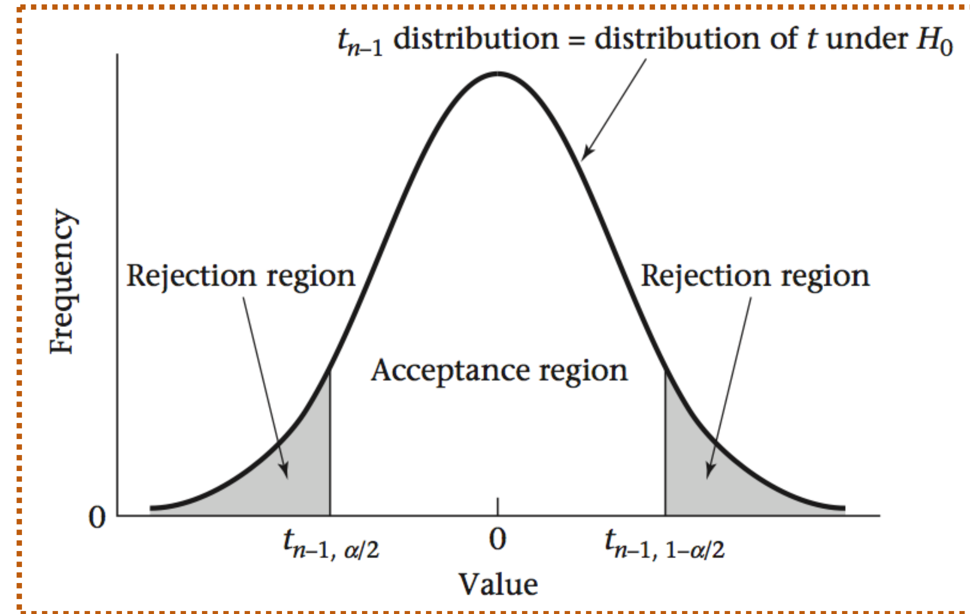
➡(*independent t*-test) Sampling distribution should be normally distributed.

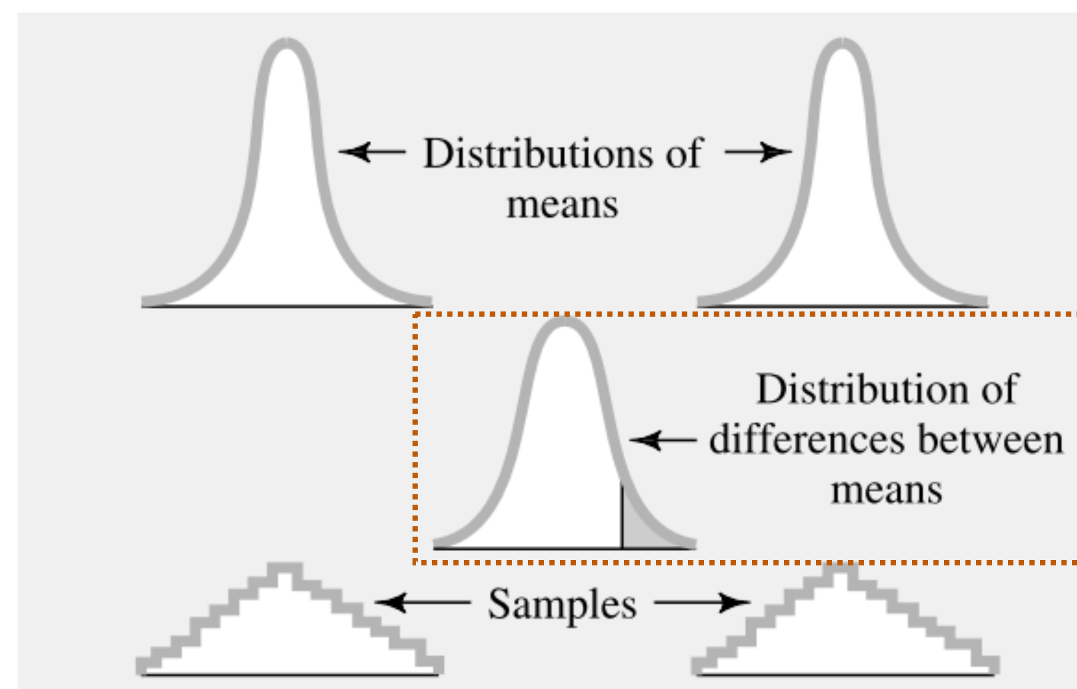
➡(*paired t*-test) Sampling distribution of the differences should be normal.

◉ ***Homogeneity of variance:***

➡(*independent t*-test) Theoretically, variance shall be equal.

How large the unreliability is…

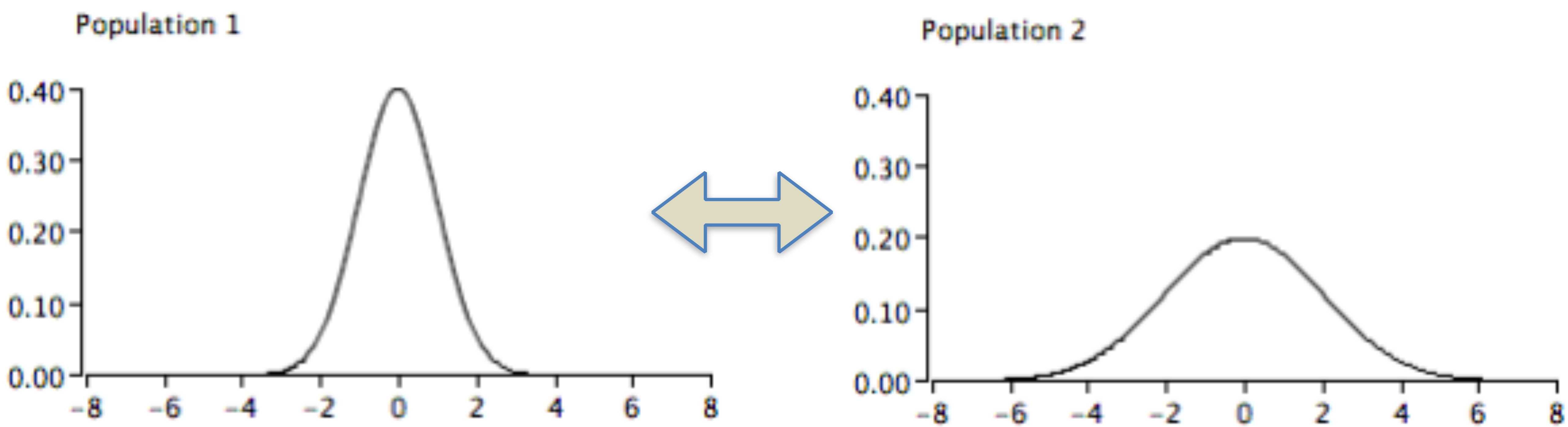




臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

Violation of Assumptions

- Violation of **Variance Homogeniety**
  - (t-test) **Welch’s approximation** to adjust degree of freedom
  - (f-test) **Welch’s F test** for adjustments **(*oneway.test* in R)**
  - (f-test) → **Non-parametric counterparts**



Violation of Assumptions

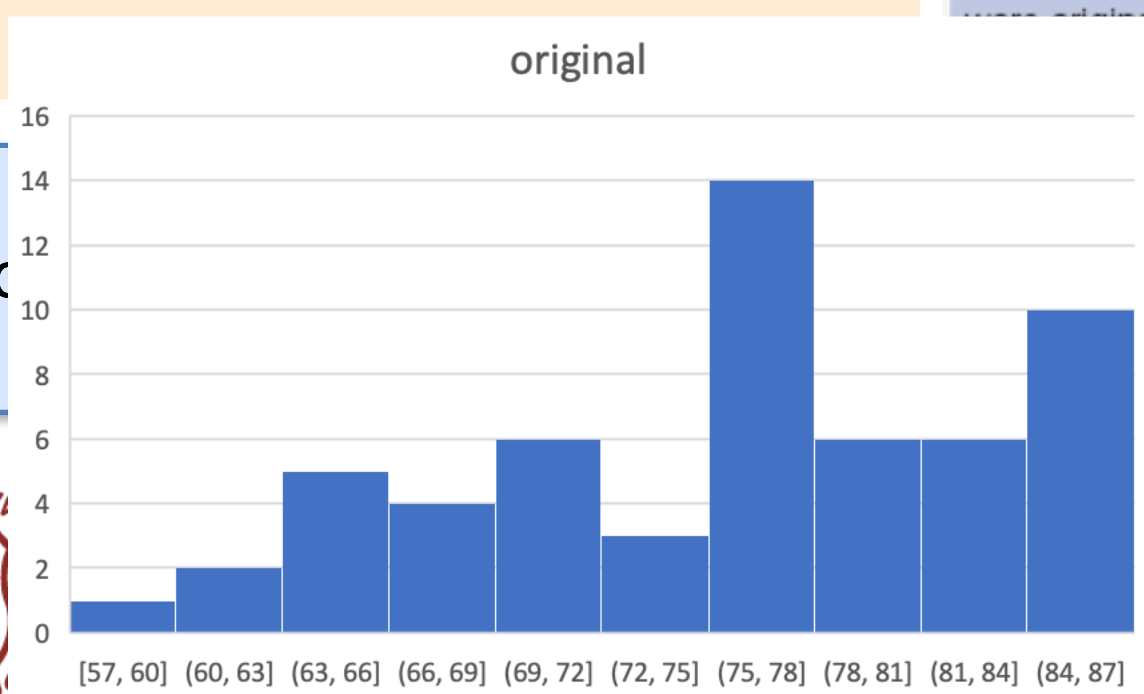
- Violation of **Normality Assumption**
  - **Data Transformation**

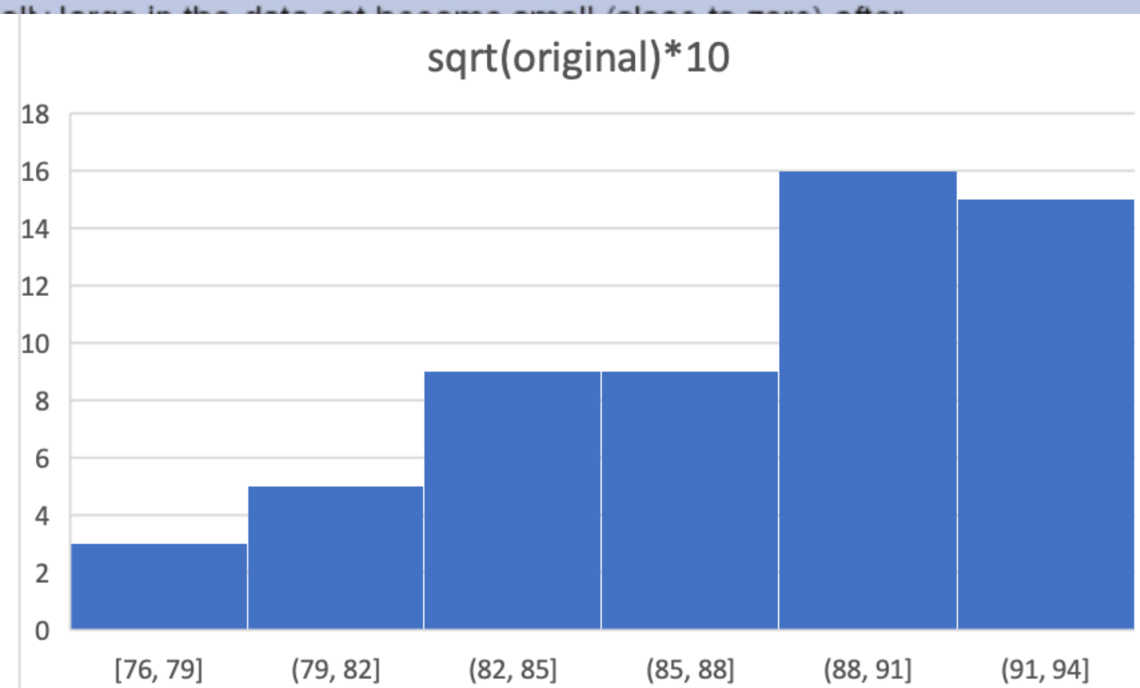
    (changes distribution, such as *log* or *sqrt*)

  - (t-test or f-test) → **Non-parametric tests**

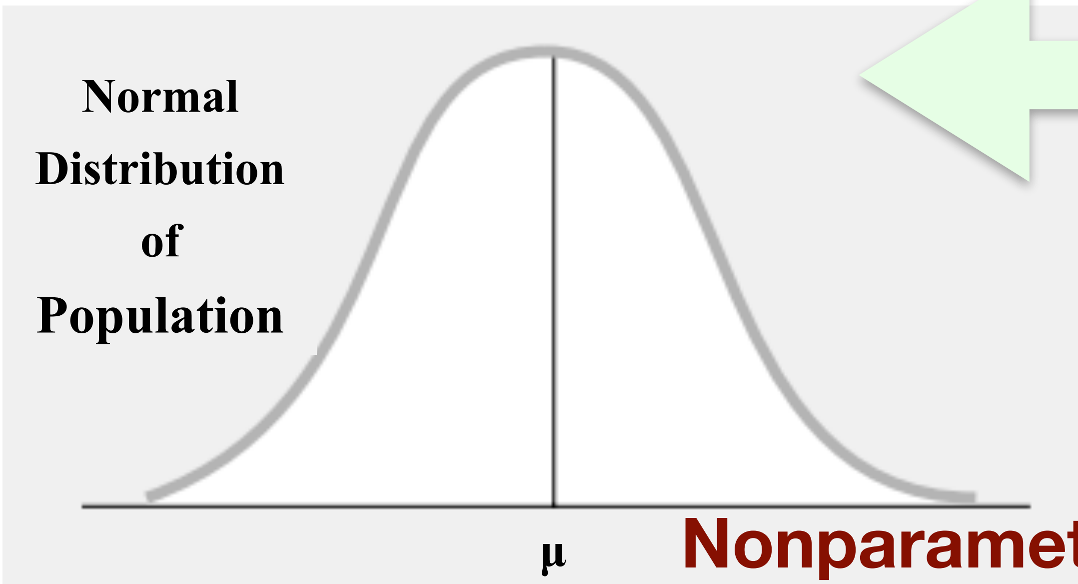
Lo

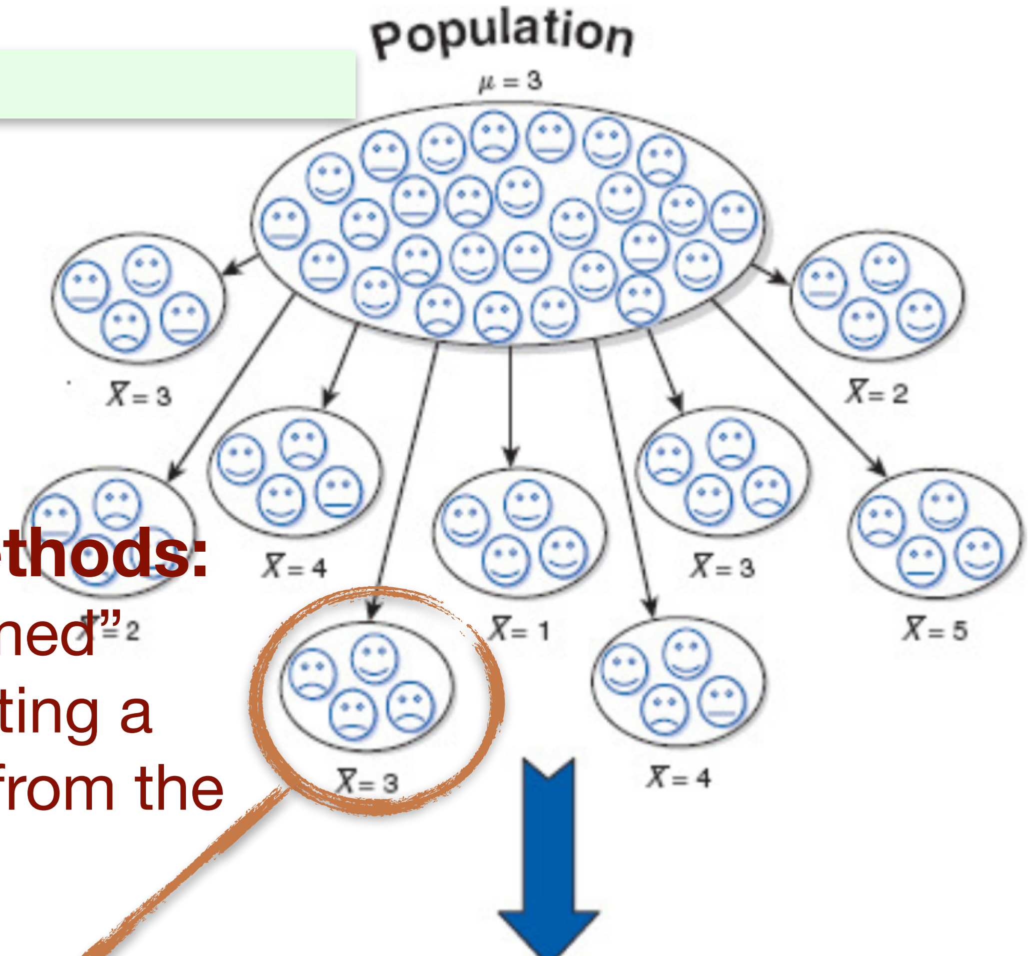
Data Transformation	Can Correct For
Log transformation (log(X_i)): Taking the logarithm of a set of numbers squashes the right tail of the distribution. As such it's a good way to reduce positive skew. However, you can't take the log of zero or negative numbers, so if your data tend to zero or produce negative numbers you need to add a constant to all of the data before you do the transformation. For example, if you have zeros in the data then do log(X_i + 1), or if you have negative numbers add whatever value makes the smallest number in the data set positive.	Positive skew, unequal variances
Square root transformation ($\sqrt{X_i}$): Taking the square root of large values has more of an effect than taking the square root of small values. Consequently, taking the square root of each of your scores will bring any large scores closer to the centre – rather like the log transformation. As such, this can be a useful way to reduce positive skew; however, you still have the same problem with negative numbers (negative numbers don't have a square root).	Positive skew, unequal variances
Reciprocal transformation (1/X_i): Dividing 1 by each score also reduces the impact of large scores. The transformed variable will have a lower limit of 0 (very large numbers will become close to 0). One thing to bear in mind with this transformation is that it reverses the scores: scores that	Positive skew, unequal variances
big scores have become small and small scores have become big!	ve skew





Parametric vs. Nonparametric





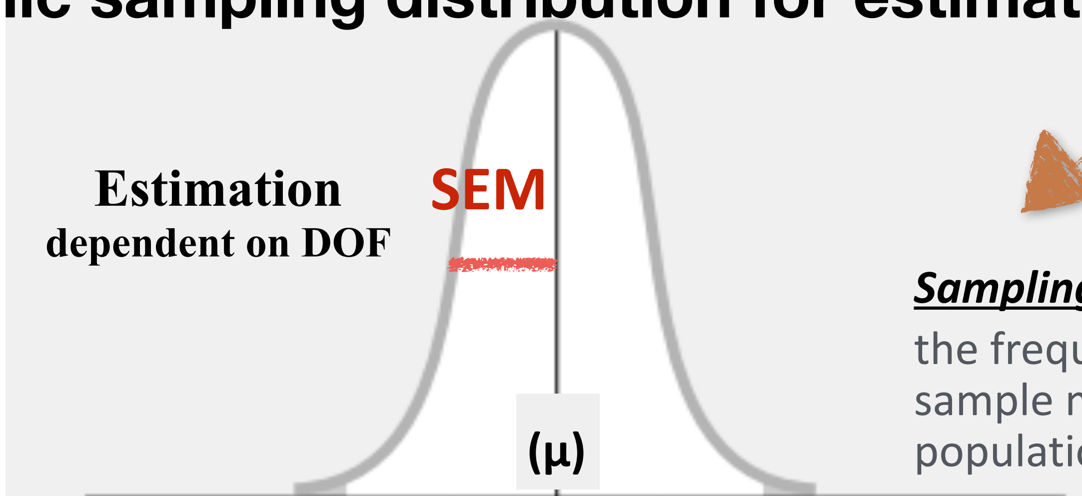
Nonparametric methods:

No use of " presumed" distribution, generating a sampling distribution from the single data

H_0: $\mu = 0$

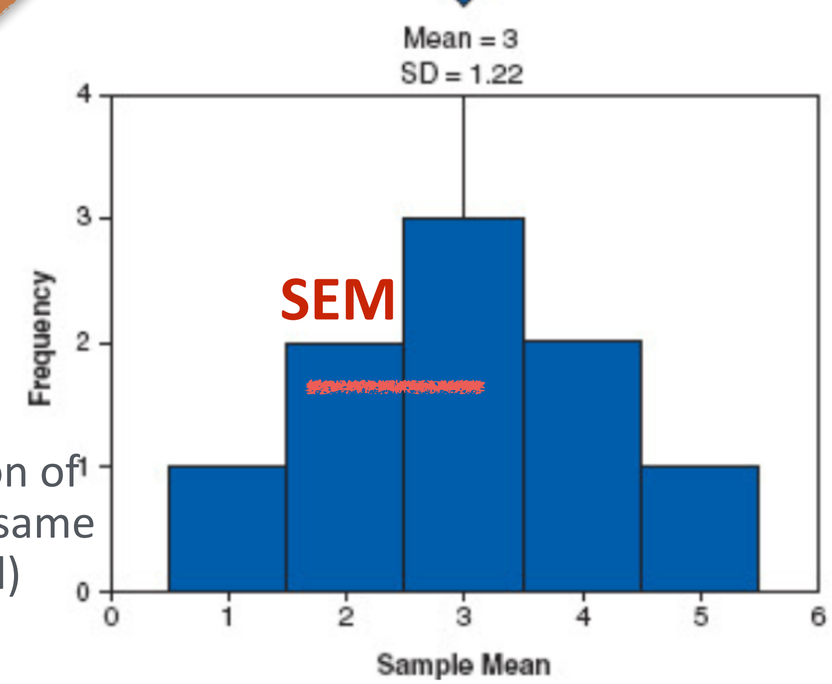
Parametric methods:

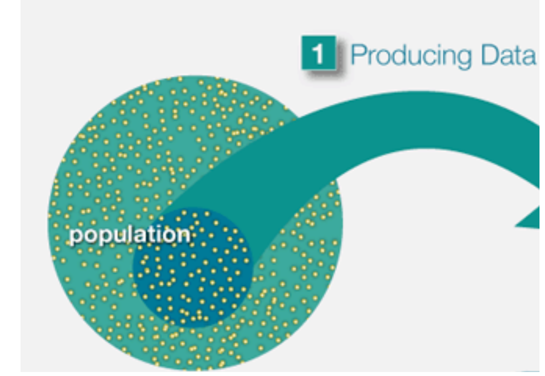
With single data, we use a" presumed" distribution to mimic sampling distribution for estimation



Sampling distribution:

the frequency distribution of sample means from the same population. (symmetrical)





Types of Variables

- **Numeric variables**
  - <u>Equal-interval variables</u> (interval scales)
    - Continuous variable: e.g., GPA; height
    - Discrete counts: e.g., the number of time visiting dentist
    - Proportions: e.g., percentage, rate
  - <u>**Rank-order** variables</u> (ordinal/discrete scales)
    - e.g., order of finishing a race; birth order of children
    - Physical activity level (low, moderate and high)

- **Nominal variables**
  - Gender (male, female)
  - Ethnicity (Caucasian, African American, Asian and Hispanic)
  - Profession (surgeon, doctor, nurse, dentist)

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Non-Parametric Statistics

- Does not require the assumption of **Normality**
  - **Distribution-free tests,** fewer assumptions, broader applicability
  - Useful when the sample size < 30 (non-Gaussian dist.)
  - Uses **ranks** (1 as the lowest value) instead of the values
  - Less susceptible to the outliers
  - Remove 0 difference (remove it from DOF as well)
  - Average generated ranks for equal difference values

- While parametric stat. focuses on the **mean** of a population, nonparametric stat. focuses on the **median**.
- Even though it is ***robust***, nonparametric tests suffer from **Less power** than the parametric statistics. (Because we lost some information because of ranking)

Robustness = the tests are reliable even when the normal assumptions are not met.

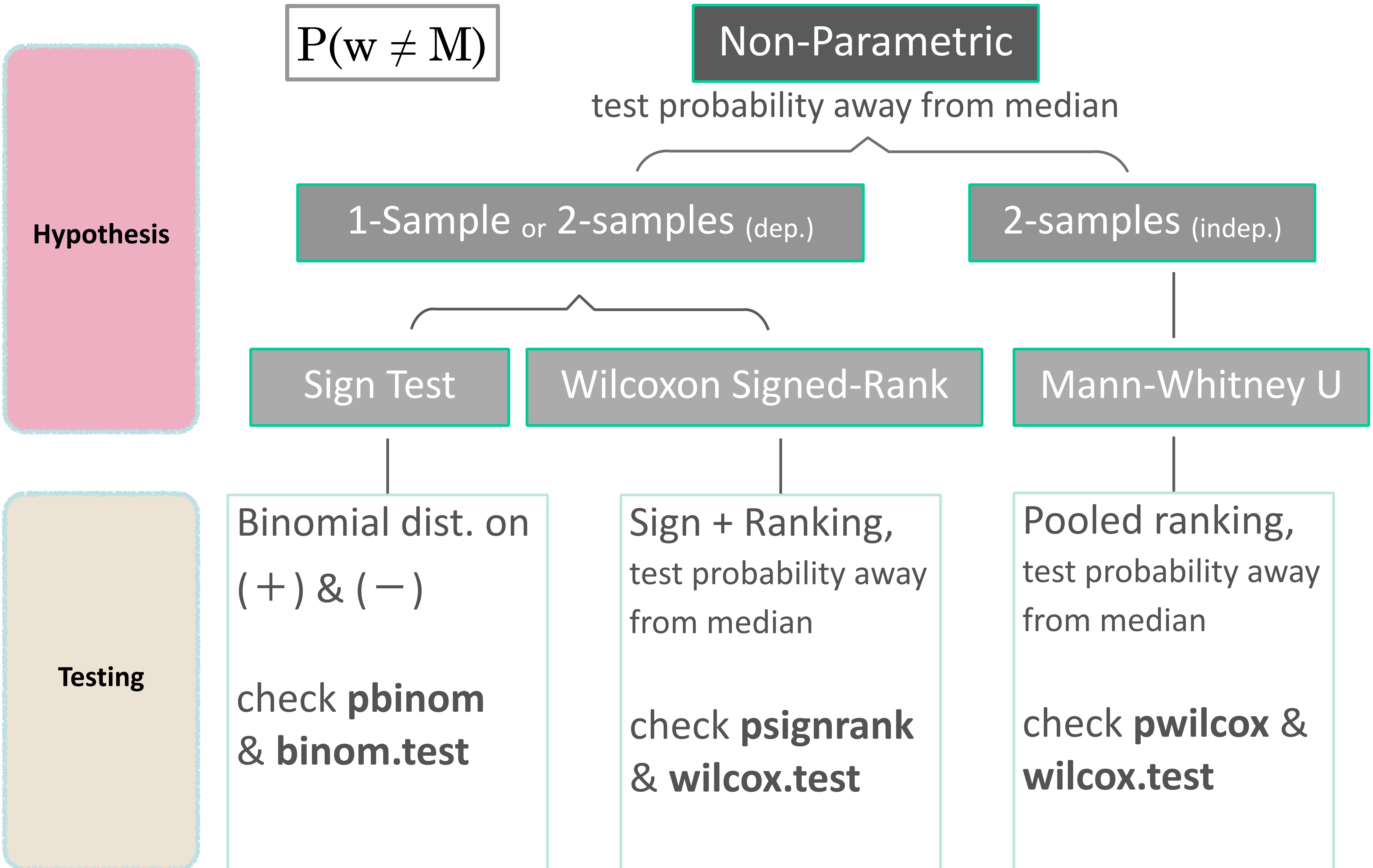
Non-Parametric Tests

	Parametric	Nonparametric
Assumed distribution	normal	any
Assumed variance	homogenous	any
Type of variables	interval or ratio	nominal or ordinal
Data set relationships	independent	any
Central measure	mean	median
Advantage	more powerful	simple, less affected by outliers

https://maxstat.de/images/Dokumentation/
parametric_vs__nonparametric_tests.htm#:~:text=Nonparametric%20tests%20are%20less%20powerful,position%20of%20pairs%20of%20scores.

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Non-Parametric Statistics & Hypothesis Testing



Non-Parametric Tests

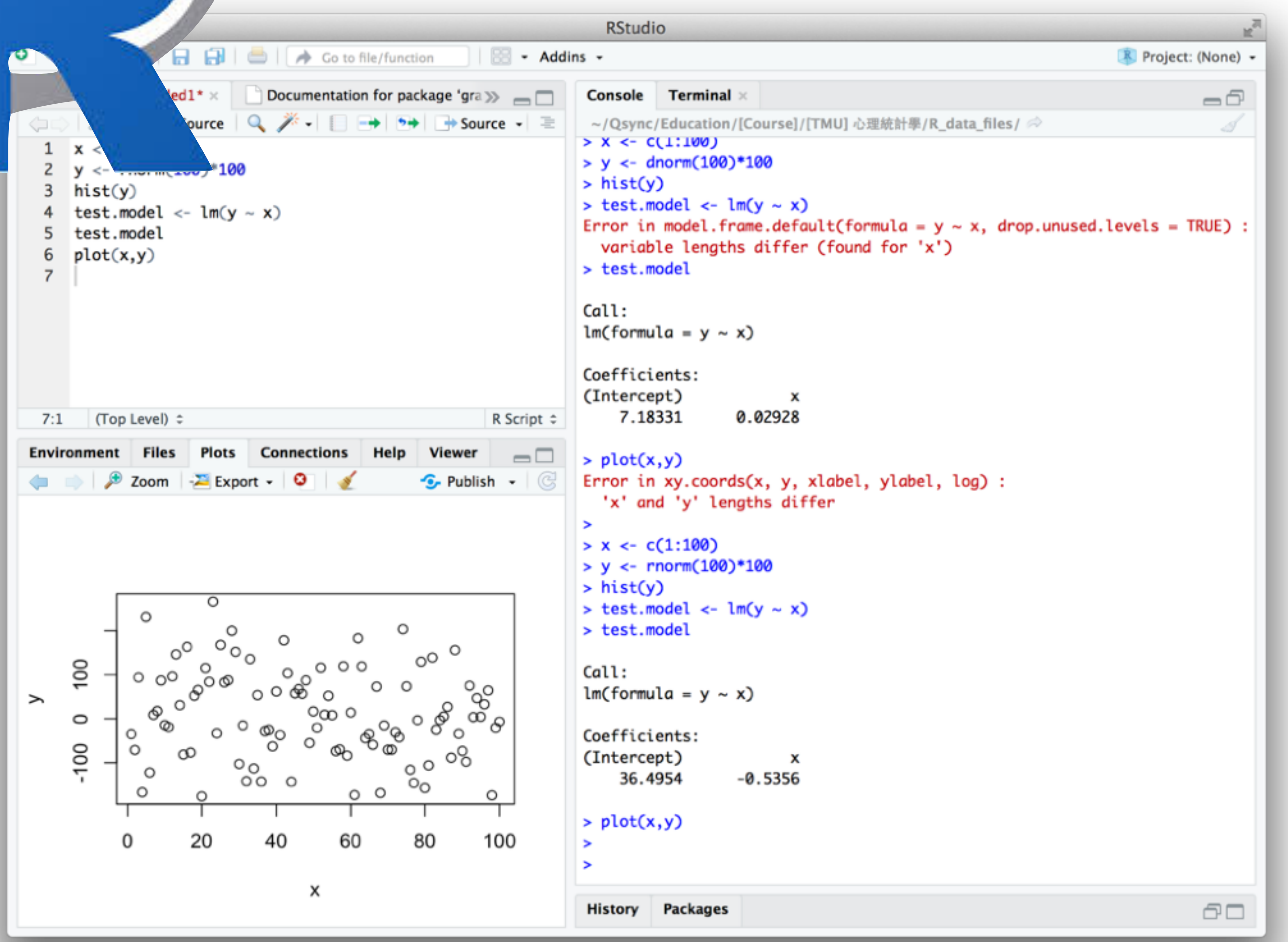
- ***Sign test*** requires no assumption on the datasets.

- ***Wilcoxon signed-rank test*** compares two repeated-measure conditions and the data violate any assumption of the paired t-test.
  - Report the W-statistic, the significance value and the effect size.
  - Draw a box plot to present the medians and corresponding ranges.

- ***Wilcoxon rank-sum (Mann-Whitney U) test*** compares two independent-measure conditions and the data violate any assumption of the independent t-test.
  - Report the U-statistic, the significance value, and the effect size.
  - Draw a box plot to present the medians and corresponding ranges.

Counterpart of One-sample t-test

① SIGN TEST



Signed Test

① Hypothesis

The hypotheses are the same as for the sign test:

- *Two-tailed.* H_0: $M = M_0$ versus H_a: $M \neq M_0$.
- *Left-tailed.* H_0: $M \geq M_0$ versus H_a: $M < M_0$.
- *Right-tailed.* H_0: $M \leq M_0$ versus H_a: $M > M_0$.

② Assumption

The assumptions of this test are more restrictive than those of the sign test.

1. The continuous random variable X is *symmetric* about a median M.
2. $X_1, \ldots, X_{n'}$ denotes a random sample of size n' from the distribution of X.
3. M_0 denotes a hypothesized median for X.

(1) Exact Method

Monte Carlo approach:

'Generate' a distribution by the sample itself ("Sign" compared with Median)

IF sample size is large or ties exist, we need to use normal approximation.

(2) Normal Approximation

TAIPEI MEDICAL UNIVERSITY 臺北醫學大學

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

(1) Exact Method

Monte Carlo approach:

'Generate' a distribution by the sample itself ("Sign" only)

Original data: [50 25 30 24 18]

Binarize by 20: [+ + + + −]

1. [− − − − −]
2. [+ − − − −]
3. [− + − − −]
4. [− − + − −]
5. [− − − + −]
6. [− − − − +]
7. [− − − + +]
8. [− − + + −]

…

How many possibilities do we have?

→ all signed conditions = $2^5 = 32$

What is the chance with 5 + ?

Example. A bank manager claims that the median number of customers per day is no more than 750. A teller doubts the accuracy of this claim. The numbers of bank customers per day for 16 randomly selected days are listed below. At α = 0.05, can the teller reject the bank manager's claim.

775	765	801	742	754	753	739	751
745	750	777	769	756	760	782	789

Ho: median ≤ 750. This is the claim.

Ha: median > 750. Right tailed test

Compare the data to the hypothesized median 750.

+	+	+	−	+	+	−	+
−	0	+	+	+	+	+	+

https://www.youtube.com/watch?v=ztmua4TrLLM

(2) Normal Approximation

If the sample size is large (>40) enough:

Use normal (z) distribution for approximation

To find a 95% confidence interval with a larger sample ($n > 20$), again list the elementary estimates in order. To find the endpoint depth d for the confidence interval limits, use the normal approximation discussed earlier. In particular, for a 95% confidence interval d is chosen so that

Continuity correction

$$P(S \le d) = P\left(Z \le \frac{d + 0.5 - \mu}{\sigma}\right) = P\left(Z \le \frac{d + 0.5 - \frac{n}{2}}{\frac{\sqrt{n}}{2}}\right) = 0.025.$$

From Table C.3, for the standard normal curve, $P(Z \le -1.96) = 0.025$, so we can solve for d in the previous equation:

$$\frac{d + 0.5 - \frac{n}{2}}{\frac{\sqrt{n}}{2}} = -1.96,$$

which is equivalent to

$$d = \frac{n - 1 - 1.96\sqrt{n}}{2}.$$

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Signed Test

EXAMPLE 6.8. Southeastern Queensland is home to more than 40 species of frogs. Among them are several species of *Crinia*, which are small, highly variable brownish or grayish froglets usually found in soaks, swamps, or temporary ponds. One way to distinguish these froglets is by their lengths. A population of froglets from a single pond appear to be of the same species. You tentatively identify them as clicking froglets *C. signifera*, a species known to have a median length of 30 mm. To be honest, you are not really sure that's what they are. A random sample of 18 froglets is captured and their lengths (in mm) are recorded in the table below. Is it reasonable to assume that these are clicking froglets?

24.1	22.9	23.0	26.1	25.0	30.8	27.1	23.2	22.8
23.7	24.6	30.3	23.9	21.8	28.1	25.4	31.2	30.9

→ all signed conditions = 2^{18}

difference = [-5.9 -7.1 -7.0 -3.9 -5.0 0.8 -2.9 -6.8 -7.2 -6.3 -5.4 0.3 -6.1 -8.2 -1.9 -4.6 1.2 0.9]

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Signed Test

SOLUTION TO EXAMPLE 6.8. Since we decided on a two-tailed test, we have H_0: $M = 30$ mm versus H_a: $M \neq 30$ mm. Perform the test with $\alpha = 0.05$. Of the 18 observations only 4 exceeded 30 mm so $S_+ = 4$ and 14 were smaller so $S_- = 14$. The test statistic is the smaller of these two, $S_+ = 4$. From Table C.1, the P value is

$$F(4) = P(S_+ \leq 4) = 0.0154.$$

How to calculate the CDF?

→ Use binomial distribution

Since 0.0154 is smaller than $\frac{\alpha}{2} = 0.025$, we reject H_0. If H_0 were true, $E(S_+) = E(S_-) = \frac{18}{2} = 9$. But $S_+ = 4$ and $S_- = 14$ and are sufficiently far from this expectation to reject H_0. Based on the evidence, it is reasonable to conclude that these are *not* clicking froglets.

→ **Use SIGN.test in R**

```
> SIGN.test(a, md=30)
```

E68<-c(24.1, 22.9, 23.0, 26.1, 25, 30.8, 27.1, 23.2, 22.8, 23.7, 24.6, 30.3, 23.9, 21.8, 28.1, 25.4, 31.2, 30.9)

Practice

Signed Test

EXAMPLE 6.9. In June of 1996 in the Finger Lakes region of New York, a series of warm days in the middle of the month caused people to comment that it was "unusually" warm. The median normal maximum temperature for Ithaca in June is 75°F. Do the data (the daily maximum temperatures for all 30 days in June) support the notion that June was unusually warm? (Based on data reported in The Ithaca Climate Page, `http://snow.cit.cornell.edu/climate/ithaca/moncrt_06-96.html`, July 2000.)

72	78	79	75	74	70	77	83	85	85
84	85	80	78	82	83	83	77	68	69
76	76	80	72	73	70	70	78	78	78

```
E69<-c(72, 78, 79, 75, 74, 70, 77, 83, 85, 85, 84, 85, 80, 78, 82, 83, 83, 77, 68, 69, 76, 76, 80, 72, 73, 70, 70, 78, 78, 78)
```


An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Practice

Signed Test

EXAMPLE 6.9. In June of 1996 in the Finger Lakes region of New York, a series of warm days in the middle of the month caused people to comment that it was "unusually" warm. The median normal maximum temperature for Ithaca in June is 75°F. Do the data (the daily maximum temperatures for all 30 days in June) support the notion that June was unusually warm? (Based on data reported in The Ithaca Climate Page, `http://snow.cit.cornell.edu/climate/ithaca/moncrt_06-96.html`, July 2000.)

72	78	79	75	74	70	77	83	85	85
84	85	80	78	82	83	83	77	68	69
76	76	80	72	73	70	70	78	78	78

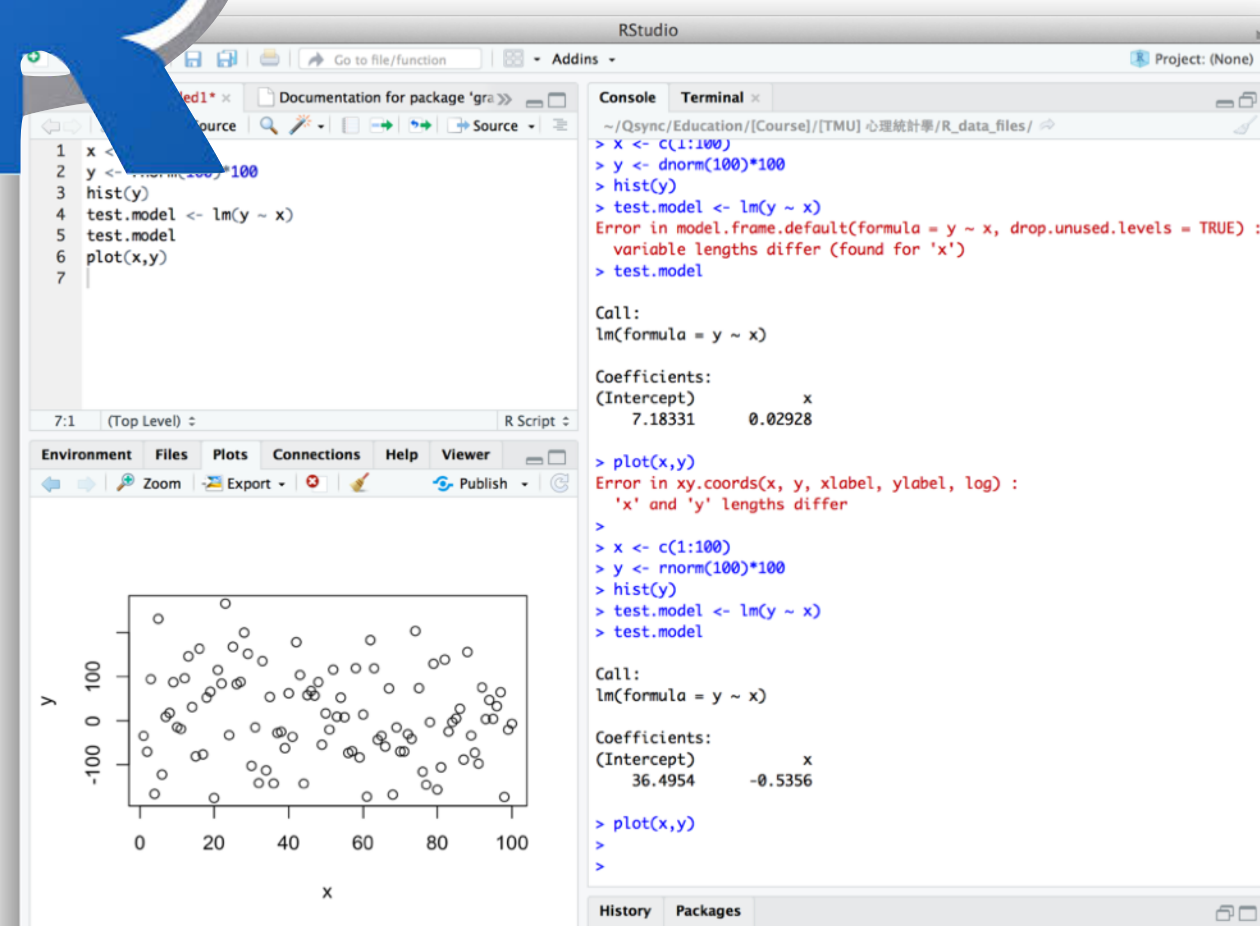
SOLUTION. There are 30 observations, but one of them equals $M_0 = 75°$. So the reduced sample size is $n = 29$. Since there is some reason to believe that temperatures were higher than normal, we perform a right-tailed test with H_0: $M \leq 75°$ versus H_a: $M > 75°$, with $\alpha = 0.05$. The test statistic is $S_- = 9$, the number of observations below the median. Then the P value using the normal approximation with $\mu = \frac{29}{2}$ and $\sigma = \frac{\sqrt{29}}{2}$ is

$$F(9) = P(S \leq 9) = P\left(Z \leq \frac{9.5 - \frac{29}{2}}{\frac{\sqrt{29}}{2}}\right) = P(Z \leq -1.86) = 0.0314 < 0.05.$$

There is sufficient evidence to reject H_0. June of 1996 was unusually warm.

Counterpart of paired t-test

② WILCOXON SIGNED-RANK TEST



TAIPEI MEDICAL UNIVERSITY
臺北醫學大學

Wilcoxon Signed-Rank Test

Again the hypothesis test can take one of three forms:

① Hypothesis

- *Two-tailed.* H_0: $M_{X-Y} = 0$ and H_a: $M_{X-Y} \neq 0$. The alternative hypothesis means that X_i tends to be different from Y_i.
- *Left-tailed.* H_0: $M_{X-Y} \geq 0$ and H_a: $M_{X-Y} < 0$. The alternative hypothesis means that X_i tends to be smaller than Y_i.
- *Right-tailed.* H_0: $M_{X-Y} \leq 0$ and H_a: $M_{X-Y} > 0$. The alternative hypothesis means that X_i tends to be larger than Y_i.

② Assumption

1. X and Y are continuous random variables.
2. $(X_1, Y_1), \ldots, (X_{n'}, Y_{n'})$ denotes a random sample of size n' from the distribution of (X, Y).
3. Consider the n' continuous differences $X_1 - Y_1, \ldots, X_{n'} - Y_{n'}$. The underlying assumption is that there is no difference between X and Y. More precisely, the assumption is that the median difference $M_{X-Y} = 0$.

The elementary estimates for the paired sign test are just the observed differences $X_i - Y_i$ in pairs.

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Exact Method

Monte Carlo approach:

'Generate' a distribution by the sample itself ("Sign" or "Rank")

five observations and the corresponding values of (+) & (-)

$+1, -2, -3, +4, +5$

$W_+ = 10$

$W_- = 5$

→ **How to test on this?**

→ all signed conditions = $2^5 = 32$

Ranking	W_+	Ranking	W_+
$-1, -2, -3, -4, -5$	0	$+1, +2, +3, -4, -5$	6
$+1, -2, -3, -4, -5$	1	$+1, +2, -3, +4, -5$	7
$-1, +2, -3, -4, -5$	2	$+1, +2, -3, -4, +5$	8
$-1, -2, +3, -4, -5$	3	$+1, -2, +3, +4, -5$	8
$-1, -2, -3, +4, -5$	4	$+1, -2, +3, -4, +5$	9
$-1, -2, -3, -4, +5$	5	$+1, -2, -3, +4, +5$	10
$+1, +2, -3, -4, -5$	3	$-1, +2, +3, +4, -5$	9
$+1, -2, +3, -4, -5$	4	$-1, +2, +3, -4, +5$	10
$+1, -2, -3, +4, -5$	5	$-1, +2, -3, +4, +5$	11
$+1, -2, -3, -4, +5$	6	$-1, -2, +3, +4, +5$	12
$-1, +2, +3, -4, -5$	5	$+1, +2, +3, +4, -5$	10
$-1, +2, -3, +4, -5$	6	$+1, +2, +3, -4, +5$	11
$-1, +2, -3, -4, +5$	7	$+1, +2, -3, +4, +5$	12
$-1, -2, +3, +4, -5$	7	$+1, -2, +3, +4, +5$	13
$-1, -2, +3, -4, +5$	8	$-1, +2, +3, +4, +5$	14
$-1, -2, -3, +4, +5$	9	$+1, +2, +3, +4, +5$	15

Probability density function (PDF):

w	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$P(W = w)$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$
$P(W \leq w)$	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{5}{32}$	$\frac{7}{32}$	$\frac{10}{32}$	$\frac{13}{32}$	$\frac{16}{32}$	$\frac{19}{32}$	$\frac{22}{32}$	$\frac{25}{32}$	$\frac{27}{32}$	$\frac{29}{32}$	$\frac{30}{32}$	$\frac{31}{32}$	$\frac{32}{32}$

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Wilcoxon Signed-Rank Test

- **Wilcoxon signed-rank test** compares two conditions when the same participants take part in each condition and the resulting data violate any assumption of the dependent t-test.
- Look at the p-value. If the value < .05 then two groups are different.
- Report the W-statistic and the significance value. Also report the medians and their corresponding ranges (with boxplot).
- Calculate the effect size if possible.

Wilcoxon Signed-Rank Test

EXAMPLE 6.13. Adult heights are known to be symmetrically distributed. We measured the heights of 10 male faculty in the science division at Hobart and William Smith Colleges. The median male adult height in the U.S. is supposedly 178 cm. Do the data collected support this assertion? Set up the appropriate two-tailed hypothesis and determine whether H_0 can be rejected at the $\alpha = 0.05$ level.

171 175 177 178 180 182 190 192 195 202

SOLUTION. The null hypothesis is H_0: $M = 178$ versus the two-sided alternative H_a: $M \neq 178$. Note that one of the observations is the same as the suspected median so it must be removed. The differences of the remaining nine sample values are calculated and their absolute values are ranked.

X_i	171	175	177	180	182	190	192	195	202
$X_i - 178$	−7	−3	−1	2	4	12	14	17	24
$\lvert X_i - 178\rvert$	7	3	1	2	4	12	14	17	24
Signed rank	−5	−3	−1	+2	+4	+6	+7	+8	+9

Under symmetry assumption, we expect $W_- \sim W_+$

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

③ Testing

⑤ Decision

Wilcoxon Signed-Rank Test

SOLUTION. The null hypothesis is H_0: $M = 178$ versus the two-sided alternative H_a: $M \neq 178$. Note that one of the observations is the same as the suspected median so it must be removed. The differences of the remaining nine sample values are calculated and their absolute values are ranked.

X_i	171	175	177	180	182	190	192	195	202
$X_i - 178$	-7	-3	-1	2	4	12	14	17	24
$\lvert X_i - 178\rvert$	7	3	1	2	4	12	14	17	24
Signed rank	-5	-3	-1	$+2$	$+4$	$+6$	$+7$	$+8$	$+9$

From the ranks we see that

$$W_+ = 9 + 8 + 7 + 6 + 4 + 2 = 36 \qquad W_- = \lvert -5 - 3 - 1\rvert = 9.$$

As a check, $W_+ + W_- = 45 = \frac{9(9+1)}{2} = \frac{n(n+1)}{2}$.

➡ **What is the p-value and the decision?**

For a two-sided test, the test statistic is the minimum of W_+ and W_-; so $W = W_- = 9$. From Table C.6, the P value for the test statistic $W = 9$ with $n = 9$ is 0.0645. Since $0.0645 > 0.025 = \frac{\alpha}{2}$, we cannot reject H_0. There is no evidence that the median male height is different from 178 cm.

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Practice

Wilcoxon Signed-Rank Test

EXAMPLE 7.13. Body height and arm span are related quantities; taller people generally have longer arms. One rule of thumb says that adult height is nearly equal to adult span. The arm spans of the 10 science faculty in Example 6.13 were also measured. Do the data support the contention? (Height and span data are symmetric.)

Height	171	175	177	178	180	182	190	192	195	202
Span	173	178	182	182	188	185	186	198	193	202

- **Use R to test on the dataset.**

  (PASWR package)

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Practice

Wilcoxon Signed-Rank Test

SOLUTION. Test the null hypothesis H_0: $M_{H-S} = 0$ versus the two-sided alternative hypothesis, H_a: $M_{H-S} \neq 0$ at the $\alpha = 0.05$ level using the Wilcoxon signed-rank test because of the symmetry assumption. Calculate the 10 differences and rank them using signed ranks. Note the use of midranks and that the observation with a 0 difference has not been ranked:

$H_i - S_i$	-2	-3	-5	-4	-8	-3	4	-6	2	0
$\lvert H_i - S_i \rvert$	2	3	5	4	8	3	4	6	2	0
Signed rank	-1.5	-3.5	-7	-5.5	-9	-3.5	$+5.5$	-8	$+1.5$	$*$

The test statistic is $W_+ = 7$, which is the smaller of W_- and W_+. Using Table C.6 with $n = 9$, the P value is 0.0371. Since $0.0371 > 0.025 = \frac{\alpha}{2}$, we cannot reject H_0 in favor of the alternative hypothesis in this two-sided test. The data support the claim that adult height and arm span are nearly equal.

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

DEMO

Ecstasy vs. Alcohol

Load: **Drug.dat**

- **Background**: We want to test the depression levels between ecstasy and alcohol users.

- Here we hypothesize that ecstasy users' depression level changes between Wednesday and Sunday. Let's test whether there will be no difference between days.

Participant	Drug	BDI (Sunday)	BDI (Wednesday)
1	Ecstasy	15	28
2	Ecstasy	35	35
3	Ecstasy	16	35
4	Ecstasy	18	24
5	Ecstasy	19	39
6	Ecstasy	17	32
7	Ecstasy	27	27
8	Ecstasy	16	29
9	Ecstasy	13	36
10	Ecstasy	20	35
11	Alcohol	16	5
12	Alcohol	15	6
13	Alcohol	20	30
14	Alcohol	15	8
15	Alcohol	16	9
16	Alcohol	13	7
17	Alcohol	14	6
18	Alcohol	19	17
19	Alcohol	18	3
20	Alcohol	18	10

DEMO

Ecstasy vs. Alcohol

- **Background**: We want to test the change of depression levels between Wednesday and Sunday for ecstasy users. Let's hypothesize that there will be no difference between days.

Dependent measure of ecstasy users over Wedn/Sun

Measure: Beck's Depression Inventory (BDI)

- **Set up Hypothesis**:

① Hypothesis

$$\begin{cases} H_0: BDI_{(Wednesday)} = BDI_{(Sunday)} \\ H_a: BDI_{(Wednesday)} \neq BDI_{(Sunday)} \end{cases}$$

臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

DEMO

Ecstasy vs. Alcohol

- **Hypothesis**:
  - $H_0: BDI_{(Wednesday)} = BDI_{(Sunday)}$
  - $H_a: BDI_{(Wednesday)} \neq BDI_{(Sunday)}$
- **Data import**:
  - *drugData <- read.delim("Drug.dat", header = TRUE)*

② Assumption

- **Assumption check:**
  - *by(drugData$BDIchange, drugData$drug, stat.desc, basic = FALSE, norm = TRUE)*

```
drugData$drug: Alcohol
    kurt.2SE   normtest.W   normtest.p
  0.36973617   0.82795980   0.03161929
---------------------------------------
drugData$drug: Ecstasy
    kurt.2SE   normtest.W   normtest.p
  -0.5128991    0.9087803    0.2727175
```

- Let's focus on Ecstasy data →

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

DEMO

Ecstasy vs. Alcohol

Exclude zero diff.

Split ranks for equal values

BDI Sunday	BDI Wednesday	Difference	Sign	Rank	Positive ranks	Negative ranks
Ecstasy						
15	28	13	+	2.5	2.5	
~~35~~	~~35~~	~~0~~	~~Exclude~~			
16	35	19	+	6	6	
18	24	6	+	1	1	
19	39	20	+	7	7	
17	32	15	+	4.5	4.5	
~~27~~	~~27~~	~~0~~	~~Exclude~~			
16	29	13	+	2.5	2.5	
13	36	23	+	8	8	
20	35	15	+	4.5	4.5	
Total =					36	0
Alcohol						
16	5	−11	−	9		9
15	6	−9	−	7		7
20	30	10	+	8	8	
15	8	−7	−	3.5		3.5
16	9	−7	−	3.5		3.5
13	7	−6	−	2		2
14	6	−8	−	5.5		5.5
19	17	−2	−	1		1
18	3	−15	−	10		10
18	10	−8	−	5.5		5.5
Total =					8	47

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

DEMO

Ecstasy vs. Alcohol

③ Testing

General form:

- ***wilcox.test(outcome ~ predictor, data=dataframe, paired=FALSE/TRUE)***
- ***wilcox.test(data1, data2, paired=FALSE/TRUE)***

- **Parameters**:
  - alternative=("two.sided" / "less" / "greater"): choose one.
  - mu=0: difference between means = 0 is default null hypothesis
  - exact=TRUE: normal approximation is used if n>50 or if ties exist.
  - correct=TRUE: by default it performs a continuity correction
  - conf.level=0.95: by default setting confidence interval = 95%
  - na.action: exclude the missing data by default (=na.exclude)

- **Usage:**
  - ***wilcox.test(wedsBDI ~ drug, data = drugData, exact = FALSE, correct= FALSE, conf.int = T)***

DEMO

Ecstasy vs. Alcohol

③ Testing

- **Usage:**
  - *wilcox.test(ecstasyData$wedsBDI, ecstasyData$sundayBDI, paired = TRUE, correct= FALSE, conf.int = T)*

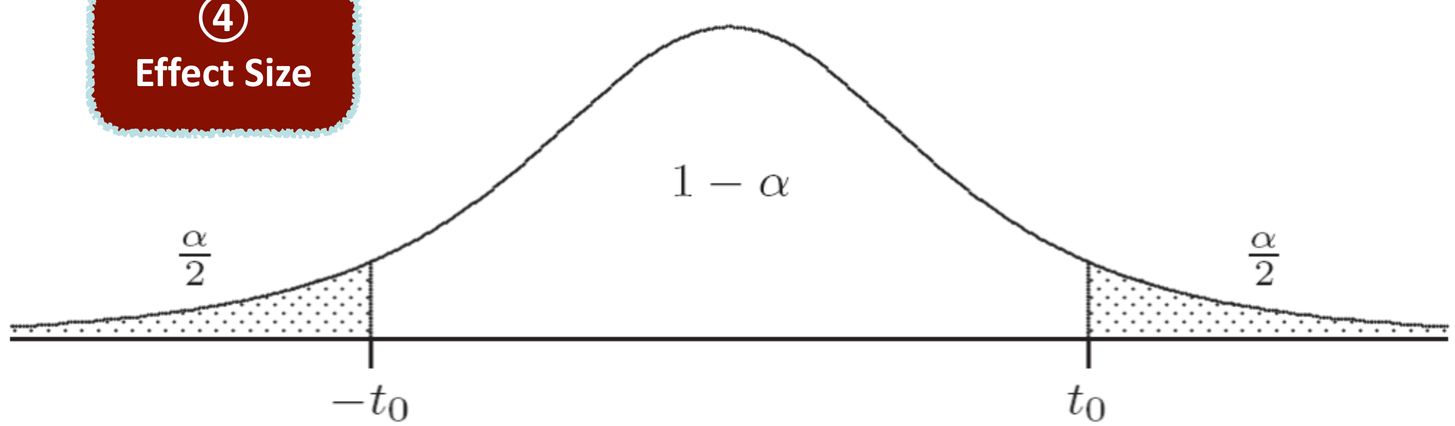
- **Results:**

```
        Wilcoxon signed rank test

data:  ecstasyData$wedsBDI and ecstasyData$sundayBDI
V = 36, p-value = 0.01151
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 10.50007 19.50002
sample estimates:
(pseudo)median
      15.95848
```

z<- qnorm(wilcoxModel$p.value/2)
r<- z/ sqrt(N)

④ Effect Size



DEMO

Ecstasy vs. Alcohol

④ Effect Size

- **Effect size**: assume the distribution as ideal, using **z → r** transform: (value from the p-value, Rosenthal, 1991, p.19)

```
ecstasyData$wedsBDI and ecstasyData$sundayBDI Effect Size, r =  -0.5649
```

⑤ Decision

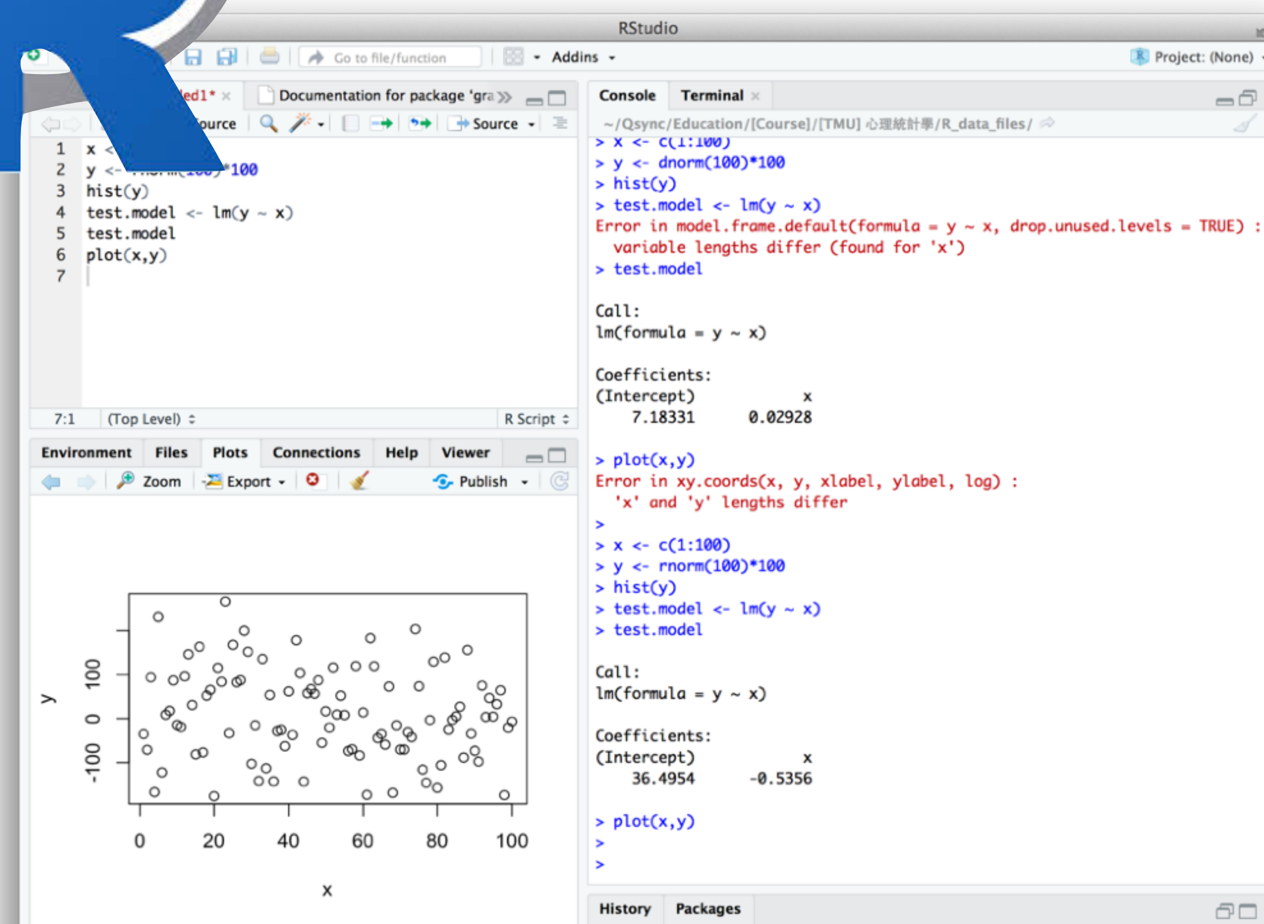
- **Reporting decision:**
  For ecstasy users, depression levels were significantly higher on Wednesday (median=33.5) than on Sunday (median=17.5), *p* = .012, effect size (*r*) = -0.56.

臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

- **How about Alcohol users?**

Counterpart of independent t-test

③ WILCOXON RANK-SUM TEST (MANN-WHITNEY U TEST)



TAIPEI MEDICAL UNIVERSITY

Wilcoxon Rank-Sum Test
Mann-Whitney U Test

Let M_X and M_Y denote the medians of X and Y, respectively. The hypothesis test can take one of three forms:

① Hypothesis

- *Two-tailed.* H_0: $M_X = M_Y$ versus H_a: $M_X \neq M_Y$. This tests whether the X's tend to be different from the Y's.
- *Left-tailed.* H_0: $M_X \geq M_Y$ versus H_a: $M_X < M_Y$. This tests whether the X's tend to be smaller than the Y's.
- *Right-tailed.* H_0: $M_X \leq M_Y$ versus H_a: $M_X > M_Y$. This tests whether the X's tend to be larger than the Y's.

② Assumption

The assumptions for the Wilcoxon rank-sum test are the following:

1. X and Y are continuous random variables.
2. The data consist of two *independent*, random samples. $X_1, \ldots, X_m$ denotes a random sample of size m from the distribution of X, and $Y_1, \ldots, Y_n$ denotes a random sample of size n from the distribution of Y. Note that m and n are often, but not always, different.
3. The null hypothesis is that the X and Y populations are identical.

For convenience, we will assume that the X's represent the smaller sample, that is, $m \leq n$. This makes listing critical values of the test statistic easier. The Wilcoxon rank-sum test examines whether the populations differ in location (median).

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Wilcoxon Rank-Sum Test (Mann-Whitney U Test)

- **Wilcoxon rank-sum test** compares two conditions when different participants take part in each condition and the resulting data violate any assumption of the independent t-test.
- Look at the p-value. If the value < .05 then two groups are different.
- Calculate the mean rank.
- Report the **U-statistic** and the significance value. Also report the medians and their corresponding ranges (with boxplot).
- Calculate the effect size and report it.

Mann-Whitney U Test

EXAMPLE 7.10. Two color morphs (green and red) of the same species of sea star were found at Polka Point, North Stradbroke Island. One might suppose that the red color morph is more liable to predation and, hence, those found might be smaller than the green color morph. The following data are the radial lengths of two random samples of these sea stars (in mm). Are red color morphs significantly smaller than green color morphs?

Red	108	64	80	92	40	
Green	102	116	98	132	104	124

SOLUTION. Let $X = R$ represent the red sea star morphs since this sample has the smaller number of observations and $Y = G$ represent the green morphs. To test whether red color morphs tend to be smaller than the green, the appropriate test is left-tailed with $\boxed{H_0\colon M_R \geq M_G \text{ versus } H_a\colon M_R < M_G.}$ The test is done at the $\alpha = 0.05$ level. The first step is to rank the combined data set. We denote the red sea star observations and ranks in **bold** for convenience:

TAIPEI MEDICAL UNIVERSITY 臺北醫學大學

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Mann-Whitney U Test

SOLUTION. Let $X = R$ represent the red sea star morphs since this sample has the smaller number of observations and $Y = G$ represent the green morphs. To test whether red color morphs tend to be smaller than the green, the appropriate test is left-tailed with H_0: $M_R \geq M_G$ versus H_a: $M_R < M_G$. The test is done at the $\alpha = 0.05$ level. The first step is to rank the combined data set. We denote the red sea star observations and ranks in **bold** for convenience:

Size	**40**	**64**	**80**	**92**	98	102	104	**108**	116	124	132
Rank	**1**	**2**	**3**	**4**	5	6	7	**8**	9	10	11

The test statistic is the sum of the red sea star ranks,

$$W_X = W_R = 1 + 2 + 3 + 4 + 8 = 18.$$

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2}$$

$$U_1 + U_2 = n_1 n_2.$$

Sum of ranks assigned to the Red group: $U_1 = R_1 - \frac{n_1(n_1+1)}{2}$ = 18-15 = 3

➡ **What is the p-value and the decision? (report U instead of W)**

```
data:  r and g
W = 3, p-value = 0.01515
alternative hypothesis: true location shift is less than 0
```

臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

DEMO

Ecstasy vs. Alcohol

- **Background**: We want to test the depression levels between ecstasy and alcohol users. Let's imagine that the **ecstasy group is more depressed than the alcohol group**. What is the working hypothesis to be tested?

Independent measure over 2 groups

Measure: Beck's Depression Inventory (BDI)

- **Set up Hypothesis**:

① Hypothesis

$$H_0: BDI_{(ecstasy)} = BDI_{(alcohol)}$$

$$H_a: BDI_{(ecstasy)} \neq BDI_{(alcohol)}$$

臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

DEMO

Ecstasy vs. Alcohol

- **Hypothesis**: H_0: $BDI_{(ecstasy)} = BDI_{(alcohol)}$
  H_a: $BDI_{(ecstasy)} \neq BDI_{(alcohol)}$

Load: **Drug.dat**

- **Data import**:
  - *drugData <- read.delim("Drug.dat", header = TRUE)*

② Assumption

- **Assumption check:**
  - *by(drugData[,c(2:3)], drugData$drug, stat.desc, basic=FALSE, norm=TRUE)*

```
drugData$drug: Alcohol
             sundayBDI     wedsBDI
normtest.W  0.95946584  0.753466511
normtest.p  0.77976459  0.003933024
------------------------------------
drugData$drug: Ecstasy
             sundayBDI    wedsBDI
normtest.W  0.81063991  0.9411413
normtest.p  0.01952060  0.5657814
```

- Let's focus on Wednesday data ➔

TAIPEI MEDICAL UNIVERSITY 1960 臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

DEMO

Ecstasy vs. Alcohol

Wednesday Data

Score	3	5	6	6	7	8	9	10	17	24	27	28	29	30	32	35	35	35	36	39
Potential Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual Rank	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15	17	17	17	19	20
Group	A	A	A	A	A	A	A	A	A	E	E	E	E	A	E	E	E	E	E	E

Sum of Ranks for Alcohol (A) = 59 Sum of Ranks for Ecstasy (E) = 151

Sunday Data

Score	13	13	14	15	15	15	16	16	16	16	17	18	18	18	19	19	20	20	27	35
Potential Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual Rank	1.5	1.5	3	5	5	5	8.5	8.5	8.5	8.5	11	13	13	13	15.5	15.5	17.5	17.5	19	20
Group	A	E	A	A	A	E	A	A	E	E	E	E	A	A	E	A	E	A	E	E

Sum of Ranks for Alcohol (A) = 90.5 Sum of Ranks for Ecstasy (E) = 119.5

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

https://en.wikipedia.org/wiki/Mann-Whitney_U_test

DEMO

Ecstasy vs. Alcohol

```
Wilcoxon rank sum test
```

- **Results**:

```
data:  wedsBDI by drug
W = 4, p-value = 0.0004943
alternative hypothesis: true location shift is not equal to 0
```

④ Effect Size

- **Effect size**: assume the distribution as ideal, using **z → r** transform: (Rosenthal, 1991)

$$r = 1 - \frac{2U}{n_1 n_2}$$

U is the smaller of the two

```
wedsBDI by drug Effect Size, r =  -0.779
```

⑤ Decision

- **Reporting decision:**

  Depression levels in ecstasy users on Wednesday (median=33.5) were significantly more depressed than alcohol users (median=7.5), U = 4, p < .001, effect size (r) = -0.78.

臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

- **How about using wedsDay in t-test?**

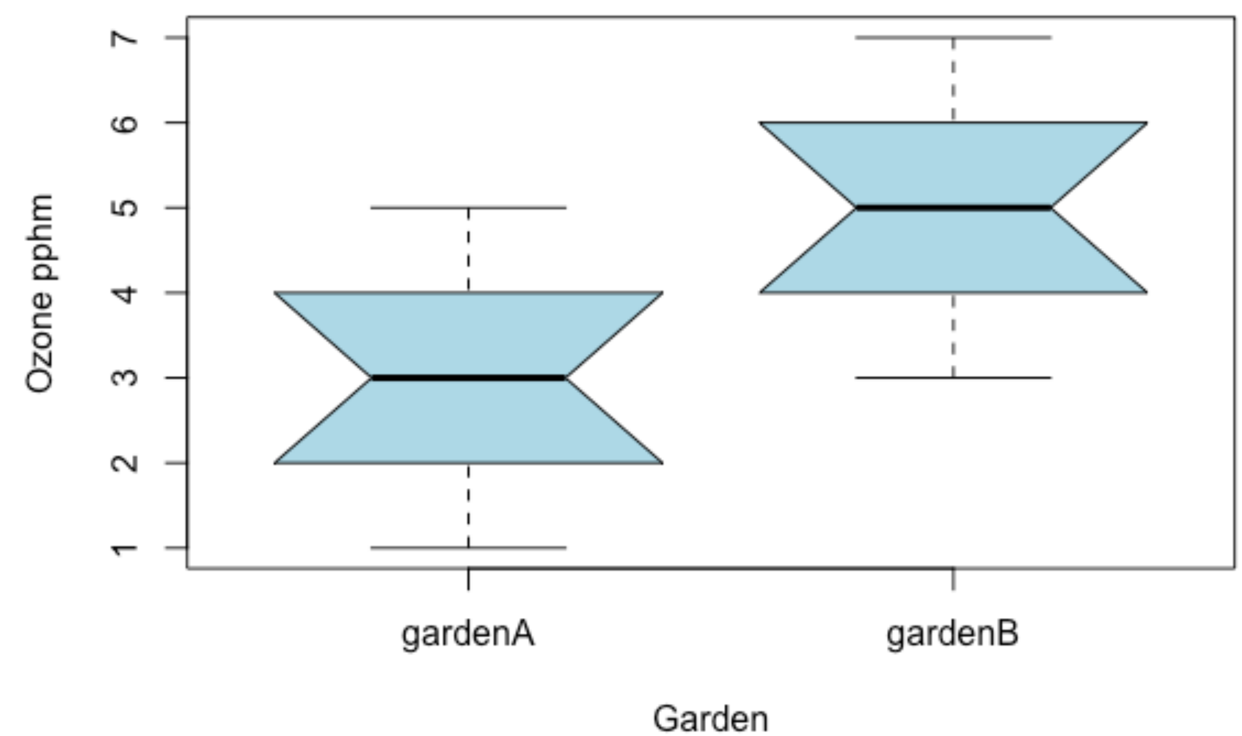
Practice

Ozone Levels

- **Background**: We have atmospheric ozone concentrations measured in parts per hundred million (pphm) in two commercial lettuce-growing gardens (we shall call the gardens A and B for simplicity).
- Based on the collected data, do the two gardens have different ozone concentrations?

Load: **ozone.csv**

- **Check the assumptions**
- **Use both t-test & nonparametric method to test it.**
- **What is the difference in results?**



Discussion

1. **Violation of Assumptions**
   - Nonparametric, Ranks & Signed Tests

1. **Nonparametric Tests: Two Samples**
   - Wilcoxon signed-rank & rank-sum tests

2. **[R] Hands-on Practices**

One-Sample Tests

One-sample tests of hypothesis with continuous variables		
Hypothesis	**Sampling assumptions**	**Test statistic**
H_0: $\sigma^2 = c$ H_0: $\sigma = c$; H_0: $\sigma^2 = c^2$	None	χ^2 test
H_0: $\mu = c$	Normal population or approximately normal under Central Limit Theorem	t test
H_0: μ or $M = c$	Symmetrical distribution	One-sample Wilcoxon signed-rank test
H_0: $M = c$	None	One-sample sign test

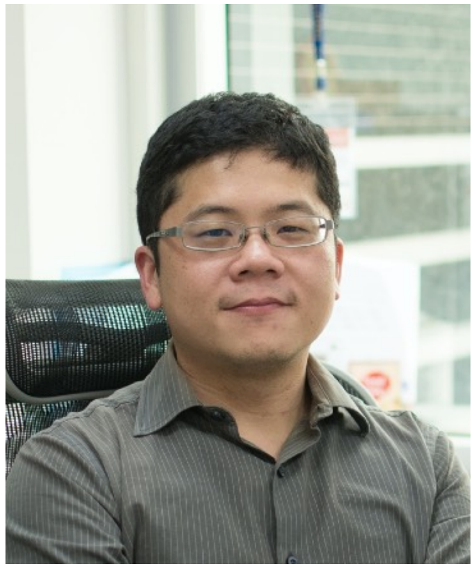
臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Two-Sample Tests

Two-sample tests of hypothesis with continuous variables		
Hypothesis	**Sampling assumptions**	**Test statistic**
H_0: $\sigma_1^2 = \sigma_2^2$	Independent samples Normal populations	F test
H_0: $\mu_1 = \mu_2$	Independent samples Normal populations Equal variances	Unpaired t test with pooled variance
H_0: $\mu_1 = \mu_2$	Independent samples Normal populations Unequal variances	Welch's approximate t test
H_0: $\mu_d = c$	Paired data Normal populations	Paired t test
H_0: $M_X = M_Y$	Independent samples	Wilcoxon rank-sum test
H_0: $M_{X-Y} = 0$	Paired data	Sign test
H_0: $M_{X-Y} = 0$	Paired data Symmetric populations	Wilcoxon signed-rank test for paired data

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.



THANK YOU FOR YOUR ATTENTION

E-mail: sleepbrain@tmu.edu.tw