StatR-05

Advanced

台灣人工智慧學校

# 無母數統計

**吳漢銘**

國立臺北大學 統計學系

http://www.hmwu.idv.tw

# 無母數統計 - 大綱

- 主題1
  - **Non-parametric Models**
  - **Non-parametric Tests**
    - Sign Test，Wilcoxon Signed-Rank Test (paired), Mann-Whitney Test, Kruskal-Wallis Test
  - **事後比較檢定 (Post Hoc Tests):** Tukey's HSD Test

# 無母數統計
# Non-parametric Statistics

- Nonparametric statistics is based on either
  - being distribution-free or having a specified distribution but with the distribution's parameters unspecified.
  - includes both descriptive statistics and statistical inference.
- **Non-parametric inferential statistical methods**: Sign test, Wilcoxon signed-rank test, Mann–Whitney U test, Kolmogorov–Smirnov test, Kruskal–Wallis one-way ANOVA,...
- **Non-parametric models**: kernel density estimation, non-parametric regression, ...

kernel regression

$$\hat{f}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)}$$

nonparametric regression

$$y_i = f_0(x_i) + \epsilon_i, \quad i = 1, \ldots n,$$

$\epsilon_1, \ldots \epsilon$ are still i.i.d. random errors with $\mathbb{E}(\epsilon_i) = 0$

*k-nearest-neighbors* regression.

$$\hat{f}(x) = \frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} y_i$$

https://en.wikipedia.org/wiki/Nonparametric_statistics

# 平均數檢定 in R

| Hypothesis Testing | One Sample | Two Samples | | > two Groups |
|---|---|---|---|---|
| | - | Paired data | Unpaired data | Complex data |
| Parametric (variance equal) | t-test<br>`t.test(x, mu = 0)` | t-test<br>`t.test(x-y, var.equal = TRUE)`<br>`t.test(x, y, paired = TRUE, var.equal = TRUE)` | t-test<br>`t.test(x, y, var.equal = TRUE)` | One-Way Analysis of Variance (ANOVA)<br>`aov(x~g, data)`<br>`oneway.test(x~g, data, var.equal = TRUE)` |
| Parametric (variance not equal) | | Welch t-test<br>`t.test(x-y)`<br>`t.test(x, y, paired = TRUE)` | Welch t-test<br>`t.test(x, y)` | Welch ANOVA<br>`oneway.test(x~g, data)` |
| Non-Parametric (無母數檢定) | Wilcoxon Signed-Rank Test<br>`wilcox.test(x, mu = 0)` | Wilcoxon Signed-Rank Test<br>`wilcox.test(x-y)`<br>`wilcox.test(x, y, paired = TRUE)` | Wilcoxon Rank-Sum Test (Mann-Whitney U Test)<br>`wilcox.test(x, y)` | Kruskal-Wallis Test<br>`kruskal.test(x, g)` |

`pairwise.t.test {stats}`: Calculate pairwise comparisons between group levels with corrections for multiple testing
`TukeyHSD {stats}`: Compute Tukey Honest Significant Differences

# Sign Test

- Given *n* pairs of data, the sign test tests the hypothesis that the **median** of the differences in the pairs is zero.
- The test statistic is the number of positive differences.
- If the null hypothesis is true, then the numbers of positive and negative differences should be approximately the same.

| Pair | Before | After | Sign |
|---|---|---|---|
| 1 | 89 | 73 | + |
| 2 | 83 | 77 | + |
| 3 | 80 | 58 | + |
| 4 | 72 | 77 | − |
| 5 | 77 | 70 | + |
| 6 | 74 | 62 | + |
| 7 | 69 | 67 | + |
| 8 | 65 | 68 | − |
| 9 | 60 | 44 | + |
| 10 | 55 | 50 | + |
| 11 | 54 | 46 | + |
| 12 | 50 | 38 | + |
| 13 | 42 | 47 | − |
| 14 | 48 | 40 | + |
| 15 | 44 | 43 | + |
| 16 | 38 | 29 | + |
| 17 | 36 | 25 | + |

**The Sign Test:**

when $n_1 = n_2 \leq 50$

$H_0 : P = Q = \frac{1}{2}$

$H_1 : P \neq Q \neq \frac{1}{2}$

$T = \#"+"$

At $\alpha = 0.01$, two-tailed test,

reject $H_0$ if $T \geq 14$ when $N = 17$.

(Binomial Probability)

# Wilcoxon Signed-Rank Test (paired)

- Null hypothesis: the **population median** from which both samples were drawn is the same.
- The sum of the ranks for the "positive" values is calculated and compared against a precomputed table to a p-value.
- If the null hypothesis is true, the sum of the ranks of the positive differences should be about the same as the sum of the ranks of the negative differences.

| Pair | Before | After | Diff. | Rank |
|---|---|---|---|---|
| 1 | 89 | 73 | 16 | 15.5 |
| 2 | 83 | 77 | 6 | 7 |
| 3 | 80 | 58 | 22 | 17 |
| 4 | 72 | 77 | −5 | 5 |
| 5 | 77 | 70 | 7 | 8 |
| 6 | 74 | 62 | 12 | 13.5 |
| 7 | 69 | 67 | 2 | 2 |
| 8 | 65 | 68 | −3 | 3 |
| 9 | 60 | 44 | 16 | 15.5 |
| 10 | 55 | 50 | 5 | 5 |
| 11 | 54 | 46 | 8 | 9.5 |
| 12 | 50 | 38 | 12 | 13.5 |
| 13 | 42 | 47 | −5 | 5 |
| 14 | 48 | 40 | 8 | 9.5 |
| 15 | 44 | 43 | 1 | 1 |
| 16 | 38 | 29 | 9 | 11 |
| 17 | 36 | 25 | 11 | 12 |

**The Wilcoxon signed-rank Test:**

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

$T = \min\{\sum_+ \text{Rank}, \sum_- \text{Rank}\}$

At $\alpha = 0.01$, two-tailed test,
reject $H_0$ if $T \neq 23$ when $N = 17$.
(Table)

(The zero difference is ignored when assigning ranks. $N_{new} = N_{old} - \#\{ties\}$ )

$$T = \min\{\textstyle\sum_+ \text{Rank} = 140, \sum_- \text{Rank} = 13\} = 13$$

The obtained T=13 is less than the critical value 23, so we reject $H_0$.

# Mann-Whitney Test

## (Wilcoxon Rank-Sum Test, unpaired)

- The data from the two groups are combined and given ranks. (1 for the largest, 2 for the second largest,... )

- The ranks for the larger group are summed and that number is compared against a precomputed table to a p-value.

| Group $G_1$ | Group $G_2$ | Rank $G_1$ | Rank $G_2$ |
|---|---|---|---|
| 26 | 16 | 3 | 11 |
| 22 | 10 | 4 | 17 |
| 19 | 8 | 7.5 | 19 |
| 21 | 13 | 5.5 | 13.5 |
| 14 | 19 | 12 | 7.5 |
| 18 | 11 | 9 | 15.5 |
| 29 | 7 | 2 | 20 |
| 17 | 13 | 10 | 13.5 |
| 11 | 9 | 15.5 | 18 |
| 34 | 21 | 1 | 5.5 |
| $n_1 = 10$ | $n_2 = 10$ | $R_1 = 69.5$ | $R_2 = 140.5$ |

**The Mann-Whitney $U$ Test:**

$H_0 : F_1 = F_2$

$H_1 : F_1 \neq F_2$

$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

or

$U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$

$R_i = \sum_i \text{Rank}$

At $\alpha = 0.05$, two-tailed test for $n_1 = 10, n_2 = 10$, reject $H_0$ if $U \leq 23$ or $U' \geq 77$ (Table)

$U$: the number of times that a score from Group 1 is lower in rank than a score from Group 2.

$U = 85.5$, $U' = 14.5$

The obtained $U = 85.5$ is less than the critical value 77, so we reject $H_0$.

# Kruskal-Wallis Test

- The Kruskal Wallis test can be applied in the one factor ANOVA case. It is a non-parametric test for the situation where the ANOVA normality assumptions may not apply.

- Each of the $n_j$ should be at least 5 for the approximation to be valid.

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_1: \mu_i \neq \mu_j \quad \text{for at least one set of } i \text{ and } j$$

$$W = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$$

$W \sim \chi^2_{k-1}$ under $H_0$

Reject $H_0$ if $W > CHIPPF(\alpha, k-1)$, the chi-square percent point function

**Groups**

| 1 | 2 | ... j ... | k |
|---|---|---|---|
| $X_{11}$ | $X_{12}$ | $\cdots\ X_{1j}\ \cdots$ | $X_{1k}$ |
| $X_{21}$ | $X_{22}$ | $\cdots\ X_{2j}\ \cdots$ | $X_{2k}$ |
| | | $\cdots$ | |
| $X_{i1}$ | $X_{i2}$ | $\cdots\ X_{ij}\ \cdots$ | $X_{ik}$ |
| $\vdots$ | $X_{n_2 2}$ | $\cdots\ \vdots\ \cdots$ | $X_{n_k k}$ |
| $X_{n_1 1}$ | | $X_{n_j j}$ | |

**Rank Data**

| 1 | 2 | ... j ... | k |
|---|---|---|---|
| $R_{11}$ | $R_{12}$ | $\cdots\ R_{1j}\ \cdots$ | $R_{1k}$ |
| $R_{21}$ | $R_{22}$ | $\cdots\ R_{2j}\ \cdots$ | $R_{2k}$ |
| | | $\cdots$ | |
| $R_{i1}$ | $R_{i2}$ | $\cdots\ R_{ij}\ \cdots$ | $R_{ik}$ |
| $\vdots$ | $R_{n_2 2}$ | $\cdots\ \vdots\ \cdots$ | $R_{n_k k}$ |
| $R_{n_1 1}$ | | $R_{n_j j}$ | |

$$F(x) = P(X \leq x) = P(X \leq G(\alpha)) = \alpha$$

$$x = G(\alpha) = G(F(x))$$

The percent point function (ppf) is the inverse of the cumulative distribution function.

# Parametric vs. Non-Parametric Test

## Parametric Tests

- Assume that the data follows a certain distribution (normal distribution).
- Assuming equal **variances** and unequal variances.
- **More powerful.**
- Widely Implemented.
- Not appropriate for data with outliers.

## Non-Parametric Tests

- When certain assumptions about the underlying population are questionable (e.g. normality).
- Does not assume normal distribution
- No variance assumption
- Less powerful.
- Widely Implemented.
- Decrease effects of outliers (Robust)
- Not recommended if there is less than 5 replicates per group.

# 事後檢定 (Post Hoc Tests)
## Tukey's Honestly Significant Difference (HSD) Test

- **Null hypothesis**: all means being compared are from the same population (i.e. $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$)

$$q_s = \frac{Y_A - Y_B}{SE},$$

$Y_A$ is the larger of the two means being compared,
$Y_B$ is the smaller of the two means being compared, and
$SE$ is the standard error of the sum of the means.

- This $q_s$ value can then be compared to a *q value* from the studentized range distribution.
- If the $q_s$ value is larger than the critical value $q_\alpha$ obtained from the distribution, the two means are said to be significantly different at level $\alpha$ , $0 \le \alpha \le 1$.

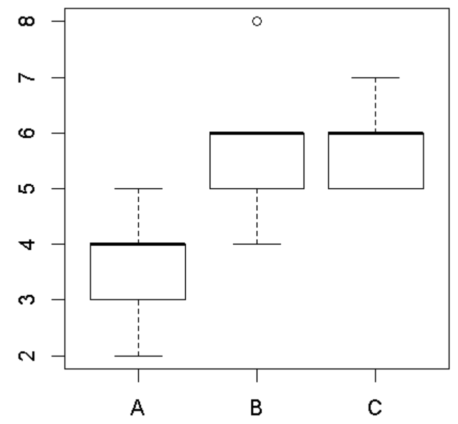
- **Assumptions for the test**
  - Observations are **independent** within and among groups.
  - The groups for each mean in the test are **normally distributed**.
  - **equal within-group variance** across the groups.
  - equal **sample sizes**.

# 範例: ANOVA + Post Hoc Test

- A drug company tested three formulations of a pain relief medicine for migraine headache sufferers. For the experiment 27 volunteers were selected and 9 were randomly assigned to one of three drug formulations. The subjects were instructed to take the drug during their next migraine headache episode and to report their pain on a scale of 1 to 10 (10 being most pain).

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Drug A: | 4 | 5 | 4 | 3 | 2 | 4 | 3 | 4 | 4 |
| Drug B: | 6 | 8 | 4 | 5 | 4 | 6 | 5 | 8 | 6 |
| Drug C: | 6 | 7 | 6 | 6 | 7 | 5 | 6 | 5 | 5 |

```
> pain <- c(4, 5, 4, 3, 2, 4, 3, 4, 4, 6, 8, 4, 5,
+  4, 6, 5, 8, 6, 6, 7, 6, 6, 7, 5, 6, 5, 5)
> drug <- c(rep("A", 9), rep("B", 9), rep("C", 9))
> migraine <- data.frame(pain, drug)
> plot(pain ~ drug, data=migraine)
> migraine.aov <- aov(pain ~ drug, data=migraine)
> summary(migraine.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
drug         2  28.22  14.111   11.91 0.000256 ***
Residuals   24  28.44   1.185
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1
> # reject the null hypothesis of equal means for all three drug grou
```




```
> kruskal.test(pain ~ drug, data=migraine)
        Kruskal-Wallis rank sum test
data:  pain by drug
Kruskal-Wallis chi-squared = 14.395, df = 2, p-value = 0.0007483
```

# Pairwise Comparisons

```
> pairwise.t.test(pain, drug, p.adjust="bonferroni")

        Pairwise comparisons using t tests with pooled SD

data:  pain and drug

  A       B
B 0.00119 -
C 0.00068 1.00000

P value adjustment method: bonferroni
>
> TukeyHSD(migraine.aov)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = pain ~ drug, data = migraine)

$drug
         diff        lwr      upr     p adj
B-A 2.1111111  0.8295028 3.392719 0.0011107
C-A 2.2222222  0.9406139 3.503831 0.0006453
C-B 0.1111111 -1.1704972 1.392719 0.9745173
>
> # conclude that the mean pain is significantly different for drug A
```