Statistical Methodology for Software Engineering

Hadas Lapid, PhD

Contents

Non-Parametric Comparison of Two Independent Group Means:

The Wilcoxon Rank Sum Test / Mann–Whitney–Wilcoxon/ Mann–Whitney U test

Reference:
https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test#Calculations

Inapplicable cases for t-test

- Small samples (<10 per group)
- Skewed/unsymmetrical mutual distribution
- Ordinal data with unequal bin size

→ t-test will result in incorrect P_v and <u>may wrongly reject H_0</u>

Nonparametric Methods

- Do not assume any distributional form
- Refer to the variable as ordinal/discrete rather than continuous
- Indicate a shift between distributions (equivalent to comparison of means)
- Size dependent

Wilcoxon Rank Sum Test
(Mann–Whitney–Wilcoxon /Mann-Whitney Test)

- Y_1 and Y_2 are two numerical, independent, samples.
- D_1 and D_2 distributions of respective parent populations
- Test Hypothesis:

$$H_0: D_1\ expected\ value = D_2\ expected\ value$$
$$H_1: D_1\ expected\ value \neq D_2\ expected\ value$$

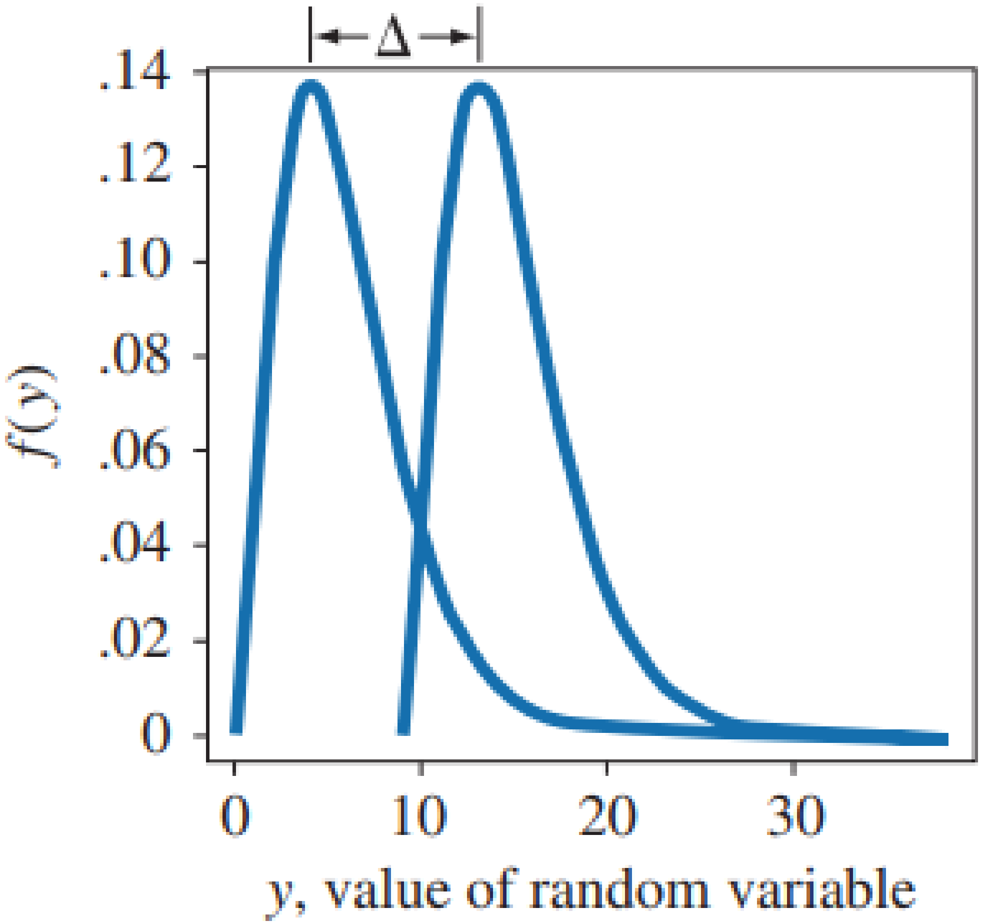
In other words:

D_1 and D_2 have similar center of mass, or D_1 is shifted with respect to D_2

Wilcoxon Rank-Sum Test

$$y \overset{d}{=} x + \Delta$$

FIGURE 6.4
Skewed population distributions identical in shape but shifted



Wilcoxon Rank Sum Test notations

- N = n_1+n_2
- $n_{1,}$ n_2 = sample size of 1st and 2nd group respectively.
- Data samples are ranked from 1 to N
- If two or more values are equal, mean ranks are assigned
- T_1, T_2 = sum of ranks in 1st and 2st groups respectively

$$H_0 : D_1 = D_2 \Rightarrow \frac{T_1}{n_1} \approx \frac{T_2}{n_2}$$

$$H_1 : D_1 \neq D_2 \Rightarrow \frac{T_1}{n_1} \neq \frac{T_2}{n_2}$$

Wilcoxon Rank Sum Test
Numerical Example

- Hypothesis: rats exposed to caffeine complete an IQ assignment faster than placebo rats.
- 5 rats were dosed with caffeine, 8 rats were given placebo.
- Time to complete the assignment was measured

$$H_0: D_{caffeine} = D_{placebo} \rightarrow \frac{T_{caffeine}}{n_{caffeine}} \approx \frac{T_{placebo}}{n_{placebo}}$$

$$H_1: D_{caffeine} < D_{placebo} \rightarrow \frac{T_{caffeine}}{n_{caffeine}} < \frac{T_{placebo}}{n_{placebo}}$$

Wilcoxon Rank Sum Test

Numerical Example

$Y_{caffeine}$	14	9	6	7	11
Rank (placebo)	9	4	1	2	6

$T_{caffeine}=22$, $n_{caffeine} = 5$, $\frac{T_{caffeine}}{n_{caffeine}} = \frac{22}{5} = 4.4$

$Y_{placebo}$	10	16	12	24	8	18	45	13
Rank (placebo)	5	10	7	12	3	11	13	8

$T_{placebo}= 69$, $n_{placebo} = 8$, $\frac{T_{placebo}}{n_{placebo}} = \frac{69}{8} = 8.625$

- For technical reasons, test distributions based on $T_{caffeine}$
- Reject H_0 if the $T_{caffeine}$ is unusually small

Wilcoxon Rank Sum Test

How small is $T_{caffeine}$?

Calculate all possibilities for choosing 5 ranks with sum ≤ 22

Allocation #	1	2	3	4	5	6	7	8	9	10	11	12	13	$T_{caffeine}$
1	X	X	X	X	X									15
2	X	X	X	X		X								16
3	X	X	X	X			X							17
4	X	X	X		X	X								17
…														
41	X	X	X			X				X				22
42	X	X		X		X			X					22

42 possibilities to allocate 5 ranks with $T_{caffeine} \leq 22$

Wilcoxon Rank Sum Test
Calculating Wilcoxon's 1-sided P_{value}

- # of combinations to choose 5 (k) ranks from 13 (n) optional ranks:

$$\binom{n}{k} = \binom{13}{5} = \frac{13!}{5! \cdot 8!} = 1287$$

- Under $H_{0,}$ the probability for each combination is: $\frac{1}{1287}$
- Probability for rank sum 22 is $\frac{42}{1287}$ = P_v = 0.0333
- 0.033<0.05 → $P_v<\alpha$
- Reject H_0 and decide H_1

Wilcoxon Rank Sum Test

Calculating Wilcoxon's 2-sided P_{value}

- If $\mathrm{H_1}: \mathrm{D_{caffeine}} \neq \mathrm{D_{placebo}}$

we would need to account for all possibilities of large values
(all possibilities from 48 to 55)

$\rightarrow P_V(2-sided) = \frac{42+42}{1287}$=0.0666

0.06>0.05 → $P_V > \alpha$

Wilcoxon Rank-Sum Test
Large Sample Normal Approximation

- If T_1 or T_2 are large (large sample size), then by central limit theorem, this sum will have approximately normal distribution.
- Use normal approximation to determine if the sum of ranks is significantly large/small.
- Standardization will be done on the mean and s.d. of T:

$$\mu(T_1) = \frac{n_1}{2}(N+1)$$

$$\mu(T_2) = \frac{n_2}{2}(N+1)$$

$$\sigma(T_1) = \sigma(T_2) = \sqrt{\frac{n_1 \cdot n_2(N+1)}{12}}$$

$$Z = \frac{T - \mu(T)}{\sigma(T)}$$

Wilcoxon Rank-Sum Test

Large Sample Normal Approximation - Example

- Used only if n_1, $n_2 > 10$
- For purpose of illustration, use example shown above:

$$\mu\left(T_{caffeine}\right) = \frac{n_{caffeine} \cdot (N+1)}{2} = \frac{5 \cdot 14}{2} = 35$$

$$\sigma = \sqrt{\frac{n_{caffeine} \cdot n_{\text{placebo}}(N+1)}{12}} = \sqrt{\frac{5 \cdot 8 \cdot 14}{12}} = 6.83$$

$$Z = \frac{T - \mu(T)}{\sigma(T)} = \frac{22 - 35}{6.83} = -1.903$$

- Under standardized normal approximation: P(z=-1.903) = 0.0285 (1-sided)
- Comparable to the exact test (0.033), approximated P_v is more permissive towards H_1
- Most computer programs use the large sample approximation