

Lecture 57: Mann–Whitney test with ties; normal approximation; Hodges–Lehmann shift estimator

Statistical Inference–II ■ Units IV–V: Nonparametric Inference

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

M.Sc. Statistics, Semester III
Academic Session 2026–2027

Previous lecture

Lecture 56: Mann–Whitney–Wilcoxon test: null distribution.

Today

- Mann–Whitney and Wilcoxon Rank-Sum Tests
 - Rank-Sum Statistic
 - Interpretation under a Location-Shift Model
 - Two-Sample Hodges–Lehmann Estimator

For tie groups of pooled sizes $t_1, t_2, \dots$,

$$\text{Var}(U_X) = \frac{mn}{12} \left[N + 1 - \frac{\sum_j (t_j^3 - t_j)}{N(N-1)} \right].$$

The standardised statistic is

$$Z = \frac{U_X - mn/2}{\sqrt{\text{Var}(U_X)}},$$

with continuity correction.

Exact critical values for $m, n \leq 20$: Tables in Appendix B of the lecture notes. For a one-sided test use the statistic (U_X or U_Y) that is small under H_1 ; for a two-sided test use $\min(U_X, U_Y)$.

Mann–Whitney–Wilcoxon test

1. Check that the two samples are independent.
2. Pool and rank all observations, using average ranks for ties.
3. Compute a rank sum and convert it to U_X or U_Y .
4. Use the exact permutation distribution for small samples; otherwise use the tie-corrected normal approximation.
5. Interpret a location-shift conclusion only when the population shapes are sufficiently similar.

Example 57.1

Let

$$X = (8, 9, 11, 12), \quad Y = (14, 15, 16, 18).$$

Test $H_0 : F = G$ against a two-sided alternative.

Solution

Every X observation is below every Y observation. The X ranks are 1, 2, 3, 4, so

$$R_X = 10, \quad U_X = 10 - \frac{4(5)}{2} = 0.$$

Under the null, each of the $\binom{8}{4} = 70$ allocations is equally likely. Only the completely separated allocation occurs in each extreme direction; therefore the exact two-sided probability is $2/70 = 0.0286$, and H_0 is rejected at the 5% level.

Example 57.2 (Mann–Whitney test with ties: normal approximation)

Marks obtained by students taught by a lecture method (X) and by a problem-based method (Y) were

X ($m = 10$) : 52, 61, 58, 47, 66, 55, 61, 49, 70, 58

Y ($n = 12$) : 68, 72, 61, 75, 64, 80, 70, 66, 77, 59, 73, 70

Test at the 5% level whether the two methods are equally effective, against the alternative that the problem-based method gives higher marks.

Example: Mann–Whitney test with ties (cont.)

Solution

Pooled ranks. Tied values receive the average of the ranks they occupy:

Value	47	49	52	55	58	58	59	61	61	61	64
Sample	X	X	X	X	X	X	Y	X	X	Y	Y
Rank	1	2	3	4	5.5	5.5	7	9	9	9	11
Value	66	66	68	70	70	70	72	73	75	77	80
Sample	X	Y	Y	X	Y	Y	Y	Y	Y	Y	Y
Rank	12.5	12.5	14	16	16	16	18	19	20	21	22

Example: Mann–Whitney test with ties (cont.)

Solution (cont.)

Statistics. $R_X = 1 + 2 + 3 + 4 + 5.5 + 5.5 + 9 + 9 + 12.5 + 16 = 67.5$, so

$$U_X = R_X - \frac{m(m+1)}{2} = 67.5 - 55 = 12.5,$$

$$U_Y = mn - U_X = 120 - 12.5 = 107.5.$$

(Check: $R_Y = 185.5$ and $185.5 - 78 = 107.5$.) Small values of U_X support H_1 .

Using the table. From the Mann–Whitney table in Appendix B of the lecture notes, the one-sided 5% critical value for $(m, n) = (10, 12)$ is 34; since $U_X = 12.5 \leq 34$, H_0 is rejected. (The table is exact for untied data; with a few ties it remains a good guide.)

Example: Mann–Whitney test with ties (cont.)

Solution (cont.)

Normal approximation with tie correction. The tie groups are 58 ($t = 2$), 61 ($t = 3$), 66 ($t = 2$) and 70 ($t = 3$), so
 $\sum(t^3 - t) = 6 + 24 + 6 + 24 = 60$. With $N = 22$,

$$\mathbb{E}(U_X) = \frac{mn}{2} = 60,$$

$$\text{Var}(U_X) = \frac{mn}{12} \left[N + 1 - \frac{\sum(t^3 - t)}{N(N - 1)} \right] = 10 \left[23 - \frac{60}{462} \right] = 228.70.$$

Example: Mann–Whitney test with ties (cont.)

Solution (cont.)

With continuity correction,

$$z = \frac{12.5 + 0.5 - 60}{\sqrt{228.70}} = \frac{-47}{15.123} = -3.108, \quad \mathbb{P} \approx \Phi(-3.108) = 0.0009.$$

Conclusion. H_0 is rejected at the 5% (indeed the 1%) level: the problem-based method gives higher marks. The tie correction changes the variance only from 230 to 228.70; it matters only when ties are extensive.

If

$$G(x) = F(x - \Delta)$$

and the two distributions have the same shape, the test is naturally interpreted as a test of $\Delta = 0$.

Without a common-shape assumption, rejection means that the two distributions differ in a manner detected by their ranks; it need not be only a median difference.

Remark 57.3 (Location versus spread)

For the bolt diameters of the Wald–Wolfowitz runs test example, the X ranks are 1, 2, 3, 10, 11, 12, so $R_X = 39$ and $U_X = 39 - 21 = 18$, close to its null mean $mn/2 = 21$; the exact two-sided p -value is 0.73.

The Mann–Whitney test therefore finds no difference, although the Wald–Wolfowitz runs test rejected H_0 . Machine 1 differs in *spread*, not in location, and a rank-sum test is designed to detect location shifts.

The location shift may be estimated by the two-sample Hodges–Lehmann estimator (Hodges and Lehmann, 1963)

$$\hat{\Delta}_{HL} = \text{median}\{Y_j - X_i : 1 \leq i \leq m, 1 \leq j \leq n\}.$$

Since U_Y computed from $X_1, \dots, X_m$ and the shifted sample $Y_1 - \Delta, \dots, Y_n - \Delta$ equals $\#\{(i, j) : Y_j - X_i > \Delta\}$, a count that decreases from mn to 0 as Δ increases, $\hat{\Delta}_{HL}$ is the point at which this count crosses its null mean $mn/2$ (the midpoint of the interval on which it equals $mn/2$, when there is one).

A confidence interval is obtained by inverting the rank-sum test. The estimator is robust and equivariant under changes of origin and scale.

Example: two-sample Hodges–Lehmann estimate

Example 57.4 (Two-sample Hodges–Lehmann estimate)

Lifetimes (hundreds of hours) of components made by an old process X and a new process Y were

$$X = (12, 15, 17, 21), \quad Y = (16, 19, 23, 24, 31).$$

Assuming a location-shift model $G(x) = F(x - \Delta)$, estimate Δ .

Solution

The $mn = 20$ differences $Y_j - X_i$ (rows X_i , columns Y_j) are

	16	19	23	24	31
12	4	7	11	12	19
15	1	4	8	9	16
17	-1	2	6	7	14
21	-5	-2	2	3	10

Example: two-sample Hodges–Lehmann estimate (cont.)

Solution (cont.)

In increasing order:

$-5, -2, -1, 1, 2, 2, 3, 4, 4, \mathbf{6}, \mathbf{7}, 7, 8, 9, 10, 11, 12, 14, 16, 19.$

With an even number of differences the median is the average of the 10th and 11th values, so

$$\hat{\Delta}_{HL} = \frac{6 + 7}{2} = 6.5.$$

For comparison, the difference of the sample medians is $23 - 16 = 7$ and the difference of the sample means is $22.6 - 16.25 = 6.35$. The new process is estimated to add about 650 hours to the typical lifetime.

Practice (Lecture Notes, Chapter 3)

- Exercises (Section 3.15): 3.22, 3.23.

Next lecture

Lecture 58: One-sample Kolmogorov–Smirnov test.