Nonparametric Equivalence Tests: A Comparison of Three Approaches

This note compares three nonparametric methods for testing the equivalence of two distributions: Meier's median-difference approach, Wellek's Mann–Whitney-based test, and the Wilcoxon TOST under a shift model. Each method targets a different notion of "no meaningful difference," leading to distinct hypothesis formulations and test statistics.

1 Meier's Fully Nonparametric Median-Difference Approach

Meier[1] considers the distribution of differences $Z = X - Y$. Because the median of Z may not be unique, define

$$Z_L \;=\; \min\{\text{all medians of } Z\} \quad \text{and} \quad Z_U \;=\; \max\{\text{all medians of } Z\}.$$

Given equivalence bounds Δ_L and Δ_U, Meier's hypothesis pair is

$$\boxed{H_0 :\; Z_L < \Delta_L \quad \text{or} \quad Z_U > \Delta_U \quad \text{versus} \quad H_1 :\; \Delta_L \;\le\; Z_L \;\le\; Z_U \;\le\; \Delta_U.}$$

That is, the null states that the entire set of possible medians does *not* lie inside $[\Delta_L, \Delta_U]$, while the alternative requires the entire median interval to be contained within the specified bounds.

2 Wellek's Mann–Whitney-based Equivalence Test

Wellek[2] focuses on

$$y \;=\; P(X > Y),$$

and tests whether y is sufficiently close to 0.5. Given small margins $\varepsilon_1, \varepsilon_2 > 0$, Wellek's hypotheses become

$$\boxed{H_0 :\; y \;\le\; \tfrac{1}{2} \;-\; \varepsilon_1 \quad \text{or} \quad y \;\ge\; \tfrac{1}{2} \;+\; \varepsilon_2 \quad \text{versus} \quad H_1 :\; \tfrac{1}{2} - \varepsilon_1 \;<\; y \;<\; \tfrac{1}{2} + \varepsilon_2.}$$

So under H_1, the probability that X exceeds Y stays in the *equivalence band* around 0.5, whereas H_0 says y is *outside* that band (either too far below 0.5 or too far above 0.5).

3 Wilcoxon TOST (Shift-Model Version)

The usual Wilcoxon TOST[3] assumes a *shift* model — i.e., that the only difference between the two distributions is a constant location shift Δ. Let

$$\Delta \;=\; (\text{population median of X}) \;-\; (\text{population median of Y}).$$

Equivalence limits Δ_L and Δ_U define

$$\boxed{H_0 :\; \Delta \;\le\; \Delta_L \quad \text{or} \quad \Delta \;\ge\; \Delta_U \quad \text{versus} \quad H_1 :\; \Delta_L \;<\; \Delta \;<\; \Delta_U.}$$

Equivalently, one can say "the difference in medians (under a shift-only assumption) is *outside* $[\Delta_L, \Delta_U]$ vs. inside that interval."

4 Summary of Each "Difference" Being Tested

- **Meier**: Looks directly at the distribution of $Z = X - Y$, testing whether *all medians* of that distribution lie inside $[\Delta_L, \Delta_U]$.
- **Wellek**: Tests whether $P(X > Y)$ is near 0.5 (within $\pm\varepsilon$) — so effectively checking that the *Mann–Whitney "probability of superiority"* is neither too large nor too small.
- **Wilcoxon TOST**: Assumes a *location shift* Δ between the two distributions and tests if $\Delta \in [\Delta_L, \Delta_U]$ by running two one-sided rank-based tests for the shift.

References

[1] Meier, U. "Nonparametric Equivalence Testing with Respect to the Median Difference". In: *Pharmaceutical Statistics* 9.2 (2010), pp. 142–150. DOI: 10.1002/pst.384.

[2] Wellek, S. "A New Approach to Equivalence Assessment in Standard Comparative Bioavailability Trials by Means of the Mann-Whitney Statistic". In: *Biometrical Journal* 38.6 (1996), pp. 695–710. DOI: 10.1002/bimj.4710380608.

[3] Caldwell, A. R. *Robust TOST Procedures*. 2025. URL: https://cran.r-project.org/web/packages/TOSTER/vignettes/robustTOST.html.