Comparison of profile-likelihood-based confidence intervals with other rank-based methods for the two-sample problem in ordered categorical data

船渡川 伊久子 データ科学研究系 准教授

1. Introduction: Two-sample problem

Two-sample problem considers the difference of the means between two groups when the variances of the two groups are not assumed to be equal. Random allocation guarantees only equal variances of baseline measurements, and unequal values of variances of post-intervention measurements are often seen in actual studies. The following properties are well know in the t-test with equal variances, the Student's t-test. In the case of unequal sample sizes and unequal variances, the actual type I error rate is not at a nominal. The actual rate is over a nominal level in the case in which a group with a large sample size has a smaller variance, and that is under a nominal level in the case in which a group with a large sample size has a larger variance. Then, in the case of unequal sample sizes, the Welch's t-test or t-test with Satterthwaite's approximation is used instead of the Student's t-test.

2. Ordered categorical data

For ordered categorical data from randomized studies, the relative effect, the probability that observations in one group tend to be larger, has been considered appropriate for a measure of an effect size. Although the Wilcoxon–Mann–Whitney test is widely used to compare two groups, the null hypothesis is not just the relative effect of 50%, but the identical distribution between groups. The null hypothesis of the Brunner–Munzel test, another rank-based method used for arbitrary types of data, is just the relative effect of 50%. In this study, we compared actual type I error rates (or 1 – coverage probability) of the profile-likelihood-based confidence intervals for the relative effect and other rank-based methods in simulation studies at the relative effect of 50% (Funatogawa and Funatogawa, 2023). The profile-likelihood method, as with the Brunner– Munzel test, does not require any assumptions on distributions.

3. Actual type I error rates

Actual type I error rates of the profile-likelihood method and the Brunner–Munzel test were close to the nominal level in large or medium samples, even under unequal distributions. Those of the Wilcoxon–Mann–Whitney test and the proportional-odds model largely differed from the nominal level under unequal distributions with unequal dispersion (the proportional-odds assumption being violated), especially under unequal sample sizes. In small samples, the actual type I error rates of Brunner–Munzel test were slightly larger than the nominal level and those of the profile-likelihood method were even larger.

4. Paradoxical numerical example

We provide a paradoxical numerical example in the Table below: only the Wilcoxon–Mann–Whitney test was significant under equal sample sizes, but by changing only the allocation ratio, it was not significant but the profile-likelihood method and the Brunner–Munzel test were significant. This phenomenon might reflect the nature of the Wilcoxon–Mann–Whitney test in the simulation study, that is, the actual type I error rates become over and under the nominal level depending on the allocation ratio.

References

Brunner, E., A. C. Bathke, and F. Konietschke. 2018. Rank and Pseudo-rank procedures for independent observations in factorial designs using R and SAS. Switzerland: Springer.

Brunner, E., and U. Munzel. 2000. The nonparametric Behrens–Fisher problem: Asymptotic theory and a small-sample approximation. Biometrical Journal 42 (1):17–25. doi:10.1002/(SICI)1521-4036(200001)42:1<17:AID-BIMJ17>3.0.CO;2-U.

Funatogawa, I., and T. Funatogawa. 2023. Comparison of profile-likelihood-based confidence intervals with other rank-based methods for the two-sample problem in ordered categorical data. Journal of Biopharmaceutical Statistics. 33(3):371-385.

Table. 95% confidence intervals (CIs) of relative effects (REs) or odds ratios (ORs) and P-values obtained from the numerical example with cell probabilities $(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{q}_1, \hat{q}_2, \hat{q}_3)$=(0.45,0.2,0.35,0.15,0.6,0.25) and relative effect $\hat{\theta}$=0.43

Sample sizes (n_1, n_2)	Frequency x_1, x_2, x_3 y_1, y_2, y_3	Profile-Likelihood	Brunner–Munzel	Wilcoxon–Mann–Whitney	Proportional-odds model
(120,120)	54, 24, 42 18, 72, 30	RE = 0.43 CI = .360–.502	RE = 0.43 CI = .358–.502 P = 0.058	P = 0.046	OR = 1.63 CI = 1.02–2.61 P = 0.042
(80,160)[a)]	36, 16, 28 24, 96, 40	RE = 0.43 CI = .349–.514	RE = 0.43 CI = .346–.514 P = 0.103	P = 0.057	OR = 1.74 CI = 1.05–2.88 P = 0.033
(160,80)[b)]	72, 32, 56 12, 48, 20	RE = 0.43 CI = .363–.499	RE = 0.43 CI = .361–.499 P = 0.047	P = 0.061	OR = 1.55 CI = 0.95–2.55 P = 0.082

Funatogawa and Funatogawa 2023

a) The group with a larger variance had a smaller sample size.

b) The group with a larger variance had a larger sample size.