Tidsskriftet Tidsskrift for Den norske legeforening

Pairwise comparisons of three groups

MEDICINE AND NUMBERS

STIAN LYDERSEN

stian.lydersen@ntnu.no
Stian Lydersen, PhD, professor of medical statistics at the Regional Centre for Child and Youth Mental Health and Child Welfare, Department of Mental Health, Norwegian University of Science and Technology.
The author has completed the ICMJE form and declares no conflicts of interest

In some studies, the researcher wants to compare three or more groups. This could, for example, be a randomised controlled trial including several treatments. In this case, it will be relevant to conduct pairwise comparisons between the groups.

If the study includes three groups – A, B and C – up to three pairwise comparisons can be conducted in the form of hypothesis tests. And, if the study includes four groups – A, B, C and D – up to six pairwise comparisons are possible: A-B, A-C, A-D, B-C, B-D and C-D. When there are several hypotheses, it will be relevant to control for the *family-wise error rate* (FWER), so that the probability of wrongly asserting that there is a difference between at least one pair of groups does not exceed an overall significance level, usually 5 %. In principle, this can be done by calculating a *p*-value for each of the pairwise comparisons and then adjusting these *p*-values using one of the methods that have been previously described in this column (1). However, when we are making pairwise comparisons, there are methods available that take the pairwise structure of the hypotheses into account, and that have substantially higher statistical power.

Various methods

Different methods are available for making such pairwise comparisons. However, the choice of method depends on a number of assumptions: should the pairwise comparisons be made between all the groups, or, for example, only with a control group. Are the variances equal or unequal? Are the groups of equal or different size? An overview in (2) lists a total of 16 different recommended methods for different sets of assumptions. If the data are normally distributed, Tukey's test is recommended for all pairwise comparisons, or Dunnett's test for comparisons with only a control group. However, this recommendation is valid only for groups of approximately equal size and equal variance. Choosing an appropriate method can be difficult, even when the data are normally distributed.

Three groups

If the study includes only three groups, as is often the case, there is a much simpler procedure that does not even rely on assumptions concerning distribution or group size: First, the global *p*-value is calculated for the null hypotheses that all three groups are identical. Second, an unadjusted *p*-value is calculated separately for each of the three pairwise comparisons. Finally, each of these three *p*-values is adjusted by replacing it with the global *p*-value if the global *p*-value is higher. This is illustrated in the example below. This procedure always controls for the family-wise error rate (3), but many researchers seem to be unaware of this fact. Even when the data are normally distributed and Tukey's test could be used, this simple method will give a statistical power that is at least as high as Tukey's test for three groups (4).

If the data are normally distributed, we can estimate the global *p*-value in a one-way analysis of variance and then make pairwise comparisons with *t*-tests. If non-parametric methods are used, we can first perform a global Kruskal-Wallis test, followed by pairwise Wilcoxon-Mann-Whitney tests. And, if the data are categorical, we can first perform Pearson's chi-squared test for three groups and then Pearson's chi-squared tests for each of the three pairwise comparisons.

Let us illustrate this with an example: Weider and colleagues compared the cognitive function in three groups of people – 41 with anorexia, 40 with bulimia and 40 healthy control persons (5), Table (3). Wechsler's intelligence scale (5) showed an average score (standard deviation) of 10.51 (3.26), 10.00 (2.42) and 11.85 (2.83) in the three groups respectively. The global *p*-value in the one-way analysis of variance was 0.014. The *p*-values for pairwise comparisons by some alternative methods are shown in Table 1. We see that by using this method, both the anorexia group and the bulimia group stand out as significantly

different from the control group with a significance level of 5 %. If Tukey's or Dunnett's test had been used, only the difference between the bulimia group and the control group would have been significant.

Table 1

Pairwise comparisons for Wechsler's intelligence scale between persons with anorexia (A), bulimia (B) and healthy control persons (K) (based on data from (5), Table 3). The global *p*-value from a one-way analysis of variance was 0.014. Unadjusted *p*-values were estimated by LSD (*least significant difference*), which is a generalisation of the *t*-test.

	Unadjusted *p*-value	Adjusted *p*-value		
Pair	LSD	Tukey	Dunnett	Maximum of global and unadjusted
A-B	0.422	0.701		0.422
A-K	0.038	0.094	0.069	0.038
B-K	0.005	0.013	0.009	0.014

Only for three groups

It must be emphasised that the method described here only controls for the family-wise error rate in three groups. For example, when three different treatments are compared to a control group, four groups are involved, and this procedure will not control for the family-wise error rate. In other respects the method is simple to apply, has high statistical power and can always be recommended for pairwise comparisons between three groups.

LITERATURE

1. Lydersen S. Adjustment of p-values for multiple hypotheses. Tidsskr Nor Legeforen 2021; 141. doi: 10.4045/tidsskr.21.0357. [CrossRef]

2. Kirk RE. Experimental design. Procedures for the behavioral sciences. 4. utg. Thousand Oaks: Sage Publications, 2013.

3. Levin JR, Serlin RC, Seaman MAA. Controlled, Powerful Multiple-Comparison Strategy for Several Situations. Psychol Bull 1994; 115: 153–9. [CrossRef]

4. Seaman MA, Levin JR, Serlin RC. New Developments in Pairwise Multiple Comparisons: Some Powerful and Practicable Procedures. Psychol Bull 1991; 110: 577–86. [CrossRef]

5. Weider S, Indredavik MS, Lydersen S et al. Intellectual function in patients with anorexia nervosa and bulimia nervosa. Eur Eat Disord Rev 2014; 22: 15–24. [PubMed][CrossRef]

Publisert: 11 October 2021. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.21.0359