XXVI Congresso Brasileiro de Engenharia Biomédica

Armação de Búzios – RJ – Brasil

October 21st to 25th, 2018

Performance of contrast correction methods for brain imaging

B A V Alberton[1], H R Gamba[1], A M Winkler[1,2]
[1] Pós-Graduação em Engenharia Elétrica e Informática Industrial (CPGEI), UTFPR, Curitiba, Brazil
[2]National Institutes of Health (NIH), Bethesda, United States of America
**biancaalberton@alunos.utfpr.edu.br*

Background, Motivation and Objective. The multiple testing problem appears in brain imaging in the context of the general linear model (GLM) in two forms: (1) a statistical test is done for each voxel, and (2) multiple contrasts using the same model are often tested. The first has been greatly studied and various different procedures have been proposed, e.g., Bonferroni, random field theory, non-parametric approaches and false discovery rate. The second arises when testing or contrasting multiple regressors, such as the difference between group means. Although there are methods to adjust p-values across contrasts, as Tukey's Honest Significance Difference (HSD), they are not available in imaging analysis software packages. Furthermore, Tukey's HSD assume that all possible contrasts among regressors are of interest (e.g. comparing all groups vs. all other groups), which often is not the case. A simple method for the adjustment of p-values that consider the multiple contrasts was proposed by [1] and consists of applying a permutation test to estimate the distribution of the extreme (e.g., distribution of the maxima) across contrasts of interest. With permutation it is also possible to correct across voxels and there is already a free tool that calculates it for arbitrary GLMs (Permutation Analysis of Linear Models - PALM [2]), which makes this procedure very attractive. The objective of this communication is to evaluate the performance of the correction across contrasts made by permutation in comparison to established methods.

Methods. One of the first procedures reported for correction of multiple comparisons between groups is Fisher's Least Significant Difference (LSD), which performs an *omnibus* ANOVA test to detect if there is any group difference followed by *post hoc* tests between each pair of groups through a t-test [3]. Although it has been proved that LSD is invalid in tests between more than three groups [4], this procedure is still extremely popular. In Tukey's HSD *post hoc* analyses are assessed not with reference to the Student's t distribution, but with the studentized range distribution [3]. An improved approach, that is exact, was proposed by [4] and also uses the studentized range distribution, albeit with one degree of freedom less than in Tukey's method. A different type of adjustment is made by the Šidák-Bonferroni procedure: adjusted p-values are produced considering that each contrast is independent from all others as $p_{adj} = 1-(1-p)^C$. We evaluated these methods in comparison with a permutation test with 1000 random permutations of simulated data in MATLAB. The data consisted of 10000 voxels with normally distributed random noise and was generated for a dataset of 3000 subjects divided in groups with different sizes. A calibrated signal was added to the first and second groups to produce an approximate power of 50%. The significance level for all tests was 5%. The q values of the studentized range distribution were computed using the R software and the contrast correction was done with PALM.

Results. Table 1 shows the family-wise error rate (FWER), its 95% confidence intervals (CI) [5] and power obtained for the various methods. Except for LSD, all tests have FWERs below the significance level of 5%, being Šidák-Bonferroni the procedure most conservative. For LSD, it can be noted that for more than 10 groups the FWER approaches values higher than 90% for the tests in which no signal was present (i.e., between groups other than those to which signal had been added). The procedure with FWER below 5% and highest power is Hayter's method.

Table *1*: FWER, its confidence interval and power of different methods for contrast correction.

	5 groups			15 groups		
Test	FWER	Confidence interval	Power	FWER	Confidence interval	Power
Permutation	1.82%	1.57% - 2.10%	33.13%	4.22%	3.84% - 4.63%	9.96%
Šidák-Bonferroni	1.27%	1.07% - 1.51%	27.80%	2.64%	2.34% - 2.97%	6.92%
Fisher's LSD	34.65%	33.09% - 36.24%	100.00%	99.25%	98.63% - 99.59%	100.00%
Tukey's HSD	1.62%	1.39% - 1.89%	30.72%	3.56%	3.21% - 3.94%	8.76%
Hayter	2.48%	2.19% - 2.80%	37.92%	3.98%	3.61% - 4.38%	9.64%

Discussion and Conclusions. We confirmed that Fisher's LSD does not control the FWER for the comparisons in which there is no signal when there is signal among other comparisons and, therefore, its use is not recommend. The correction across contrasts via permutation presents a FWER below the test level of 5%. This method also has one of the greatest power, smaller only than that of the Hayter's. Permutation has some advantages over the other procedures studied here: Tukey's and Hayter's methods can handle only positive test statistics, permutation does not make many assumptions about the data (such as being normally distributed), and it is the only method other than Šidák-Bonferroni that can be used to correct across both contrasts and voxels. Furthermore, permutation tests can be applied to specific contrasts, whereas Tukey's and Hayter's assume that all possible contrasts are of interest. Even though Šidák-Bonferroni procedure presents the lowest FWER, this method is more conservative than the others, which causes an increased false negative rate (Type II error) and results in reduced power. Therefore, we recommend the use of permutation to correct across contrasts. As future work, effects of non-orthogonal contrasts in the performance of these methods will be studied as well as the effect of and correcting over only a subset of group comparisons.

References. [1] Winkler, Anderson M., et al. "Non-parametric combination and related permutation tests for neuroimaging." *Human brain mapping* v.37 n.4 (2016): 1486-1511. [2] Winkler, Anderson M., et al. "Permutation inference for the general linear model." *Neuroimage v.* 92 (2014): 381-397. [3] Hochberg, Yosef, and Tamhane, Ajit. "Multiple comparison procedures." (1987). [4] Hayter, Anthony J. "The maximum familywise error rate of Fisher's least significant difference test." *Journal of the American Statistical Association* v81. n.396 (1986): 1000-1004. [5] Wilson, Edwin B. "Probable inference, the law of succession, and statistical inference." *Journal of the American Statistical Association v.* 22 n.158 (1927): 209-212.

Acknowledgment. This work was produced with a scholarship provided by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Support was also provided by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

Keywords. multiple comparisons; multiple testing; brain imaging; permutation tests; contrast correction.