The Mid p-value and Exact Statistical Inference on Hardy-Weinberg Equilibrium

Graffelman, J.[1]

[1]jan.graffelman@upc.edu, Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain.

Statistical theory for continuous variables implies that the p-value obtained in a hypothesis test has a uniform distribution if the null hypothesis is true. In genetics, genomics, genetic epidemiology and related fields hypothesis tests are often repeated many times, and it is of interest to study the distribution of the p-values. Q-Q plots of p-values (against the uniform distribution) have become an important tool, and are often used to identify the most salient features of a dataset. However, if the variables under study are discrete, with only a moderate number of possible outcomes, then the distribution of the p-value is also discrete, and no longer uniform. In the case of exact statistical tests for Hardy-Weinberg equilibrium (HWE), the distribution of the p-values often shows a spike at the value 1, in particular for small samples with a low minor allele frequency. The p-value is usually defined as the probability of observing the obtained test-statistic or something more extreme with respect to the null hypothesis. Exact tests for HWE are known to be conservative in the sense that for low minor allele counts they underrate the nominal significance level α. The mid p-value, defined as *half* the probability of the observed sample or something more extreme with respect to the null, can be used as an alternative in these circumstances. Under the null, the mid p-value is shown to have expectation $\frac{1}{2}$, whereas the standard p-value has a larger expectation. Moreover, for exact HWE tests, the mid p-value is most close to the nominal significance level. In this contribution some properties of the mid p-value are discussed, and its use in tests for HWE is illustrated with genetic data.

Keywords: Power, exact test, minor allele frequency.

References:

Graffelman, J. and Moreno, V. (2013) The mid p-value in exact tests for Hardy-Weinberg equilibrium. *Statistical Applications in Genetics and Molecular Biology* 12(4), pp. 433–448. doi: 10.1515/sagmb-2012-0039.

Rohlfs, R. V. and Weir, B. S. (2008) Distributions of Hardy-Weinberg equilibrium test statistics. *Genetics* 180, pp. 1609–1616.