ACCURACY OF THE GAUSSIAN APPROXIMATIONS TO DISTRIBUTION OF THE WILCOXON STATISTICS. A GRAPHICAL APPROACH.

Carine A Bellera[1], Marilyse Julien[2], James A Hanley[3]

[1] *Clinical Research and Epidemiology Unit, Institut Bergonié, Regional Comprehensive Cancer Center, Bordeaux, France.*
[2] *Department of Mathematics and Statistics, McGill University, Montreal, Canada.*
[3] *Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada.*

The Wilcoxon tests are usually taught as non-parametric alternatives for the 1- and 2-sample Student t-tests in situations where the data appear to arise from non-Gaussian distributions, or where sample sizes are so small that we cannot check if they do. In the past, critical values, based on exact tail values, were presented in tables, often laid out in a way that saves space but makes them confusing to look up. This is unfortunate given the logic behind these tests is so easy to teach. Recently a number of textbooks have bypassed the tables altogether, and reverted to Gaussian approximations to the exact distributions. In the context of non-Gaussian data, students find the use of Gaussian statistics to test hypotheses concerning non-Gaussian data confusing. Moreover, different textbooks and software give different advice regarding the sample sizes at which such approximations are accurate. We survey what textbooks and software manuals say about this, and display graphically the shapes of these exact distributions. The plots illustrate that the Gaussian approximation holds with sample sizes much smaller that the ones usually suggested by the books, and we thus let users make their own rules of thumb. Finally, we also show how the exact distributions can easily be generated 'on the fly' in today's software.