The Relevance of Nonparametric and Semi-Parametric Statistics to HCI

Jacob O. Wobbrock
The Information School
DUB Group
University of Washington
Seattle, WA, USA
wobbrock@uw.edu

CHI 2017 Workshop on *Moving Transparent Statistics Forward*.
Denver, Colorado (May 6, 2017).

Abstract
Human-Computer Interaction (HCI) research methods are increasingly diverse, generating various types of data. Often, data are not appropriate for conventional parametric analyses of variance (ANOVAs). Instead, nonparametric and semi-parametric statistics may be more appropriate, but many such methods remain obscure and underutilized. This position paper promotes the view that non- and semi-parametric tests have a welcome place in HCI research. Moreover, they should be accompanied by effect sizes, which are often not apparent or readily available for many of these tests. Making nonparametric and semi-parametric tests known, along with their effect size calculations, is an important priority for HCI research methods.

Author Keywords
Statistics; nonparametric; semi-parametric; quantitative research methods; effect sizes.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. G.3. Probability and statistics: Nonparametric statistics.

Position Statement
Human-Computer Interaction (HCI) is a wide-ranging field, with computer scientists, social scientists, and designers variously studying technology and human behavior, and creating and evaluating interactive

Method	*Data*
Likert scales	Ordinal responses
Coded observations, interviews, diaries	Nominal responses, counts
Surveys	Ordinal responses, nominal responses, counts
Demographic questionnaires	Exponential distributions (e.g., income)
Human performance experiments, usability tests	Lognormal distributions (e.g., times), Poisson distributions (e.g., errors), binomial distributions (e.g., success/fail outcomes)

Table 1: Various research methods in HCI giving rise to certain types of quantitative data suited to nonparametric or semi-parametric statistical tests.

technologies. Given the range of these research endeavors, it is no surprise that the types of data examined are also diverse and wide-ranging (Table 1). To make sense of these data, inferential statistics are often employed. Owing to a many factors, including background, training, expediency, and tradition, researchers often utilize familiar parametric analyses of variance (ANOVAs). Unfortunately, ANOVAs are often inappropriate or undesirable due to response types or assumption violations. Add to this the recognized problems of null hypothesis significance testing (NHST) [4], and the misinterpretation of *p*-values [13], and there is a need for HCI statistical methods to improve.

One path forward is to better understand and employ nonparametric and semi-parametric statistical tests (see Table 2). Although such tests remain within the NHST paradigm, for many studies of novel technologies, where priors for Bayesian methods may be unavailable [4], non- and semi-parametric tests offer important alternatives to ANOVAs that avoid common pitfalls. For example, statistical power is often increased by ANOVA procedures relative to their nonparametric counterparts, exacerbating the chances of Type I errors and the reliance on *p*-values even when practically significant differences may be absent.

Also of relevance to HCI, but largely unrecognized and unused, are so-called "semi-parametric" statistical models, like generalized linear models (GLMs) [7] and generalized linear mixed models (GLMMs) [2,9]. These powerful, flexible models have become computationally feasible and can accommodate various data distributions and response types (e.g., nominal responses using logistic regression). To date, relatively few HCI papers employ GLMs or GLMMs.

Finally, along with the results of nonparametric and semi-parametric statistical tests, the reporting of effect sizes can convey practical significance and facilitate comparisons. Reporting effect sizes has been urged for ANOVA reports as well [13], but a challenge for non- and semi-parametric tests is that effect size calculations are often not obvious or readily available. Shedding light on such calculations and encouraging their use will improve HCI statistical practice.

Author's Short Biography

I am an Associate Professor in the Information School and, by courtesy, in the Department of Computer Science & Engineering at the University of Washington. I direct the Mobile & Accessible Design Lab and am a founding member of the DUB Group and the Master of Human-Computer Interaction & Design program.

My research seeks to understand people's interactions with computers and information, and to improve those interactions through design and engineering. My specific research interests include interaction techniques, human performance measurement and modeling, HCI research and design methods, mobile computing, and accessible computing.

Concerning statistics, I have published a book chapter with Matthew Kay on nonparametric statistics [12], created a popular statistics independent study[1], and taught an online Coursera course called *Designing, Running, and Analyzing Experiments*[2], which has

[1] http://depts.washington.edu/madlab/proj/ps4hci/
[2] https://www.coursera.org/learn/designexperiments

Factors, Levels	*B/W*	*Tests*
1, 2	B	Mann-Whitney [6]
1, 2	W	Wilcoxon signed-rank [10]
1, >2	B	Kruskal-Wallis [5]
1, >2	W	Friedman [1]
>1, ≥2	B	ART [3,8,11]; GLM [7]
>1, ≥2	W	ART [3,8,11]; GLMM [2,9]

Table 2: Some useful nonparametric and semi-parametric statistical tests for various (B)etween- or (W)ithin-subjects experimental designs in HCI. Notes: "ART" stands for Aligned Rank Transform, "GLM" stands for generalized linear model, and "GLMM" stands for generalized linear mixed model. This table is adapted from [12].

enrolled over 10,000 learners since February 2016. I also published a CHI 2011 paper on *ARTool* [11] for enabling Aligned Rank Transform nonparametric tests.

I have authored over 120 peer-reviewed publications, receiving 12 best paper awards and 7 honorable mentions. I am on the editorial board of ACM Transactions on Computer-Human Interaction. I received my B.S. and M.S. from Stanford University, and my Ph.D. from Carnegie Mellon University.

Intended Contributions to the Workshop

In this CHI 2017 workshop, I would like to contribute to conversations on useful and appropriate statistical methods, offering insight into nonparametric and semi-parametric tests, and voicing support for reporting effect sizes. I would also offer support for other "new" methods, such as Bayesian methods [4], which promise to benefit HCI as the field reconsiders its statistical practices. I would also like to contribute to author and reviewer guidelines that aim to help both roles more effectively communicate and consider statistical results in HCI. It is an exciting time to be reflecting on these issues, which are of great importance to HCI research.

References

1. Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association 32* (200), pp. 675-701.
2. Gilmour, A.R., Anderson, R.D. and Rae, A.L. (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika 72* (3), pp. 593-599.
3. Higgins, J.J. and Tashtoush, S. (1994). An aligned rank transform test for interaction. *Nonlinear World 1* (2), pp. 201-211.
4. Kay, M., Nelson, G.L. and Hekler, Eric B. (2016). Researcher-centered design of statistics: Why Bayesian statistics better fit the culture and incentives of HCI. In *Proc. CHI 2016*. New York: ACM Press, pp. 4521-4532.
5. Kruskal, W.H. and Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association 47* (260), pp. 583-621.
6. Mann, H.B. and Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics 18* (1), pp. 50-60.
7. Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. Journal of the Royal Statistical Society, Series A 135 (3), pp. 370-384.
8. Salter, K.C. and Fawcett, R.F. (1993). The ART test of interaction: A robust and powerful rank test of interaction in factorial models. *Comm. in Statistics: Simulation and Computation 22* (1), pp. 137-153.
9. Stiratelli, R., Laird, N. and Ware, J.H. (1984). Random-effects models for serial observations with binary response. *Biometrics 40* (4), pp. 961-971.
10. Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin 1* (6), pp. 80-83.
11. Wobbrock, J.O., Findlater, L., Gergle, D. and Higins, J.J. (2011). The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. In *Proc. CHI 2011*. New York: ACM Press, pp. 143-146.
12. Wobbrock, J.O. and Kay, M. (2016). Nonparametric statistics in human-computer interaction. Chapter 7 in J. Robertson and M.C. Kaptein (eds.), *Modern Statistical Methods for HCI*. Switzerland: Springer, pp. 135-170.
13. Yatani, K. (2016). Effect sizes and power analysis in HCI. Chapter 5 in J. Robertson and M.C. Kaptein (eds.), *Modern Statistical Methods for HCI*. Switzerland: Springer, pp. 87-110.