Practical Statistics for Human-Computer Interaction: An Independent Study Combining Statistics Theory and Tool Know-How

Jacob O. Wobbrock
The Information School | DUB Group
University of Washington
Seattle, WA 98195 USA
wobbrock@uw.edu

ABSTRACT
As both a student and professor, I have observed that new doctoral students in Human-Computer Interaction (HCI) often lack the statistical fundamentals necessary for conducting experimental research in HCI. Courses offered in statistics departments can be deeply theoretical but difficult for students to apply, while help manuals for statistics software packages describe the application of analyses without providing the rationale behind them. Between these extremes lies a "practical middle," wherein students gain enough theory to understand statistical analyses, but also learn how to handle data using current software tools. This "practical middle" for statistics is not often taught in HCI curricula. To address this issue, I have developed a set of independent study modules designed for new doctoral students in HCI to rapidly gain the statistics know-how necessary for their research. The modules relate specifically to the kinds of data common in HCI experiments, often involving myriad trials, many crossed factors, repeated measures, and the need for nonparametric analyses. To date, 11 students have taken this independent study, and report feeling proficient in understanding and producing statistical results after completing the study.

Author Keywords
Statistics, evaluation, experiments, laboratory studies.

ACM Classification Keywords
K.3.2 [Computers and Education]: Computer and Information Science Education— *Curriculum.*

INTRODUCTION
Although the range of research methods used in Human-Computer Interaction (HCI) has expanded well beyond controlled laboratory studies, such studies remain an important staple of the field. Yet many curricula lack courses dedicated to developing the statistics expertise necessary for analyzing these studies. While courses on research methods exist in many HCI doctoral programs, the amount of time spent on statistics is often only a few classes or assignments—not enough to give students an appreciation of anything beyond the ubiquitous ANOVA.

Complicating this picture are the data arising from studies in HCI, which, owing to the nature of computer-based work, may contain tens of thousands of trials, multiple crossed independent factors, subjective ratings, and nonparametric measures possibly from error counts, ordinal scales, category proportions, and phenomena producing non-normal distributions. While students coming into HCI doctoral programs from the social and behavioral sciences such as psychology, sociology, anthropology, education, communication, or economics may have the skills to analyze such data, many students from computer science, information science, engineering, and design do not. The typical approach is to "outsource" these students' statistics educations to statistics or psychology departments, which, at the graduate level, usually offer deeply theoretical treatments over a narrow range of study designs. In contrast, what students in HCI need is a practical working knowledge over many study designs and statistical techniques.

To that end, I have spent two years developing *Practical Statistics for HCI*, a set of independent study modules and an explanatory answer key enabling students to pace themselves through theoretical issues and practical know-how with two common statistics packages, SAS JMP and IBM PASW (formerly SPSS). To date, 11 students have completed this independent study for credit, and report feeling proficient in both understanding and producing statistical results after completing the study.

PRACTICAL STATISTICS FOR HCI
Practical Statistics for HCI was designed to establish the "practical middle" between two pedagogical endpoints: on the one hand, formal graduate-level statistics courses offered by statistics departments, and on the other, help manuals and other documentation accompanying statistics

Boaster at the *Annual Workshop of the Human-Computer Interaction Consortium (HCIC '11).* June 14-18, 2011, Pacific Grove, California, USA.

software packages. The "practical middle" lies between these extremes, where students gain enough theoretical understanding to avoid making blunders, and enough tool know-how for carrying out analyses on data sets of the kind found in HCI.

There are, of course, similar treatments found in some text books. For example, the *Computational Handbook of Statistics* by Bruning and Kintz [1] offers step-by-step instructions for how to carry out ANOVAs and other tests by hand. Similarly, *Using SPSS for Windows and Macintosh* by Green and Salkind [6] offers similar stepwise instructions in the context of the SPSS software package. However, these and other texts are general treatments not specific to HCI, and lack devotion to the kinds of data that often arise in our field. As noted, these data may comprise tens of thousands of trials, be taken over many potentially interacting independent factors, have multiple layers of repeated measures (particularly for longitudinal studies), and be nonparametric in nature. They also may involve many *post hoc* comparisons, of which many may have been unplanned. Rarely do "practical" texts devote themselves to such advanced topics, perhaps assuming that anyone delving into such matters will be deeply devoted by them. Unfortunately, that is not the case with many new doctoral students in HCI, who are, naturally, seeking to contribute to HCI, not statistics or quantitative research methods, *per se*.

Content of the Independent Study

Practical Statistics for HCI consists of 10 modules designed to be completed in about 10 weeks, making it suitable for universities on either the quarter or semester system. Two documents, one with just questions and one with both questions and explanatory answers, are given to students at the start of the term, along with data sets in *.jmp, *.sav, and *.xlsx formats, the file extensions for the SAS JMP, IBM PASW (formerly SPSS), and Microsoft Excel programs. The 10 modules are:

1. Concepts and definitions
2. Understanding data
3. Exposure to ANOVA
4. Statistics tools
5. Repeated measures
6. Transformations
7. Nesting
8. Longitudinal data
9. Nonparametric tests
10. Categories, counts and proportions.

The parametric analyses covered in these modules include the *t*-test, paired samples *t*-test, ANOVA, and repeated measures ANOVA. Traditional fixed-effects models and newer mixed-effects models with fixed and random effects are both covered [4,11,13].

The nonparametric analyses covered include the Mann-Whitney *U* test [12], Wilcoxon signed-rank test [16], Kruskal-Wallis test [10], Friedman test [5], nominal and ordinal logistic regression [17], and one- and two-sample Chi-Square tests of proportions. In addition, some useful lesser-known analyses are covered, such as Poisson regression for rare event data [15], and the Aligned Rank Transform (ART) [7,18] for producing accurate interaction effects for nonparametric data, particularly data from repeated measures.

Other topics covered include *post hoc* comparisons, corrections for multiple comparisons (*e.g.*, Holm's sequential Bonferroni procedure [8]), data transformations [9], tests for distributional conformity (*e.g.*, Shapiro-Wilk *W* test [14], Kolmogorov *D* test [3]), nesting of factors, fitting learning curves to longitudinal performance data [2], detecting where learning stabilizes, and a discussion of covariance structures, among others.

Theoretical knowledge of the aforementioned topics is combined at all times with carrying out actual analyses in JMP and/or SPSS. The questions or the answer key provide step-by-step instructions for operating the statistics software packages being used.

Student Assessment

I have supervised 11 students as they have progressed through *Practical Statistics for HCI*, all of whom were from the computer science or information science doctoral programs at the University of Washington. Students were required to print their typed answers to questions, mark those answers themselves in hardcopy using the answer key, and turn in that hardcopy at the end of the quarter. I verify that the answers have been completed and marked, but I do not grade the content of each answer—that is up to the student. As this is an independent study, I make myself available to students throughout the quarter for help but do not serve as a grader, saving me time and putting the onus on the student to master the material. Grading is on a Credit/No Credit basis. Students report feeling proficient in both understanding and producing statistical results by the end of the independent study. The study is being considered for use at Harvard University and Stanford University.

Online Resources

Although it is well-tested and stable, *Practical Statistics for HCI* still evolves as new topics, problems, and explanations are occasionally added. The project page for the study is http://depts.washington.edu/aimgroup/proj/ps4hci/. As of this writing, the version date is April 28, 2011.

FUTURE WORK

A current limitation of *Practical Statistics for HCI* is that it is based on two proprietary statistics packages, SAS JMP and IBM PASW (formerly SPSS). Currently, a graduate student at Stanford University is planning to create a version for **R**, which is a free open source statistics tool with a devoted community.

Topics not currently included in the independent study but planned for addition in the future are power analyses, effect sizes (Pearson *r*, Cohen's *d*, η^2), inter-rater agreement (Cohen's κ), and instrument consistency/reliability (Cronbach's α), among others.

CONCLUSION

As in any mature scientific field, the need for rigorous and appropriate methodology is paramount. The proper use of statistics in HCI is vital to ensuring the trustworthiness of our field's contributions, especially if those contributions are to be well regarded by communities outside HCI itself. As instructors in a multidisciplinary field, we must educate our students in a variety of areas, both substantive and methodological, and do so while acknowledging that students' research interests lie in HCI and not in contributing disciplines, *per se*. The independent study described herein, *Practical Statistics for HCI*, is an attempt to enable students to quickly become proficient in understanding and producing statistical results in HCI. Other research methods used in HCI may benefit from resources like this one that aim for the "practical middle" between pure theory on the one hand, and sheer practice on the other.

ACKNOWLEDGMENTS

The author thanks the students who entrusted their statistics education, at least in part, to the independent study described in this paper. The author also thanks Darren Gergle, Krzysztof Z. Gajos, Leah Findlater, Jessica J. Tran, James J. Higgins, and Howard Seltman.

This work was supported in part by National Science Foundation grant IIS-0952786. Any opinions, findings, conclusions or recommendations expressed in this work are those of the author and do not necessarily reflect those of the National Science Foundation.

REFERENCES

[1] Bruning, J.L. and Kintz, B.L. (1997). *Computational Handbook of Statistics*, 4th ed. New York: Longman.

[2] Card, S.K., Moran, T.P. and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, New Jersey: Lawrence Erlbaum.

[3] Conover, W.J. (1972). A Kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association 67* (339), 591-596.

[4] Frederick, B.N. (1999). Fixed-, random-, and mixed-effects ANOVA models: A user-friendly guide for increasing the generalizability of ANOVA results. In *Advances in Social Science Methodology*, B. Thompson (ed.). Stamford, Connecticut: JAI Press, 111-122.

[5] 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), 675-701.

[6] Green, S.B. and Salkind, N.J. (2011). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*, 6th ed. Upper Saddle River, New Jersey: Pearson Education.

[7] Higgins, J.J. and Tashtoush, S. (1994). An aligned rank transform test for interaction. *Nonlinear World 1* (2), 201-211.

[8] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6* (2), 65-70.

[9] Hoyle, M.H. (1973). Transformations: An introduction and a bibliography. *International Statistical Review 41* (2), 203-223.

[10] Kruskal, W.H. and Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association 47* (260), 583-621.

[11] Littell, R.C., Henry, P.R. and Ammerman, C.B. (1998). Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science 76* (4), 1216-1231.

[12] 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), 50-60.

[13] Schuster, C. and von Eye, A. (2001). The relationship of ANOVA models with random effects and repeated measurement designs. *Journal of Adolescent Research 16* (2), 205-220.

[14] Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika 52* (3 & 4), 591-611.

[15] Vermunt, J.K. (1997). *Log-linear Models for Event Histories*. Thousand Oaks, California: Sage Publications.

[16] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin 1* (6), 80-83.

[17] Winship, C. and Mare, R.D. (1984). Regression models with ordinal variables. *American Sociological Review 49* (4), 512-525.

[18] Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J. (2011). The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)*. Vancouver, British Columbia (May 7-12, 2011). New York: ACM Press, 143-146.