ABSTRACT

While statistics is recognized as an important topic in the area of human-computer interaction (HCI) and other scientific fields, it is also a topic that spurs continuous debate. The current practice is that, despite extensive developments in the area of statistics in the last decades, most practitioners stick to the most simple parametric methods. Increasingly, we see the argument arise that scientists will have to resort to more advanced methods, which are unfortunately only available in advanced statistical packages that require a specialized syntax and a substantial understanding of the underlying statistical principles. This paper introduces Illmo, a program for performing statistics in a more intuitive way. In order to operate the program successfully, the user only needs to understand a single statistical principle, i.e., the likelihood as a goodness-of-fit measure between the observed data and the proposed statistical model. Illmo is unique in the sense that its visual interface not only provides extensive graphical renderings of the data analysis results, but also assists explicitly in navigating between different (but related) statistical techniques.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces; J.2 [Physical Science and Engineering]; Mathematics and Statistics

Keywords
Statistics, Interaction, Likelihood

1. INTRODUCTION

The title of this paper has, quite purposely, been chosen to be open for at least two interpretations. The first possible interpretation is that this paper would be dealing with how the Human-Computer Interaction (HCI) community currently resolves issues related to statistics, such as which methods are deemed most appropriate for analyzing both objective measurements, such as performance times, and subjective measurements, such as rating scales, and discuss the most frequently experienced problems with these statistical methods. Although interesting in its own right, this is unlikely to provide a suitable topic for a conference on Advanced Visual Interfaces. The second, more appropriate, interpretation of the title is that the author takes the position that many of the problems reported with selecting and applying statistical methods are to a large extent due to the inappropriate interfaces offered by current statistical software packages. More precisely, these packages usually generate ample data analysis output, in both textual and graphical form, but provide little or no assistance with choosing the appropriate statistical method and/or with selecting the options offered by these methods. This paper therefore proposes a new, strongly visually-oriented, interface that provides an alternative access to many existing statistical methods, as well as incorporates some new ones in the process.

The perfect introduction to, and motivation for, the current paper can actually be found in a recent BLOG@CACM by Julie Robertson, entitled “Stats: We’re Doing It Wrong” (http://cacm.acm.org/blogs/blog-cacm/107125-stats-were-doing-it-wrong/fulltext). Robertson takes the position that performing proper statistics is very relevant for HCI, as well as for many other disciplines, such as psychology. Specifically, she refers to a paper by Kaptein et al. [3] that reports that 45% of the papers in the CHI’2009 proceedings used Likert scales to collect data, but only 8% of them used proper non-parametric statistics for the analysis of such ordinal data. She points at the unresolved debate, also raised in the paper of Kaptein et al., on which non-parametric method should be preferred, and remarks that, even if a researcher would want to apply more advanced statistical methods, he will have a hard time doing so. Several of these advanced methods are simply not supported by SPSS (Statistical Package for the Social Sciences), which is by far the most popular statistical software in the field of HCI [2]. The suggestion that she comes up with is to resort to programs such as R or SAS instead. Unfortunately, these are likely to require substantially more effort than SPSS to learn to operate correctly, as these packages have been developed with statistical experts in mind for whom the command-line syntax used is not a major obstacle. This resembles the comparison between command-line interfaces, which can be extremely powerful in the hands of experts that use them frequently, and graphical user interfaces (GUIs), aimed at non-expert or non-frequent users. This paper introduces a new program, called Illmo (for Interactive Log Likelihood...
MOdeling), for bridging the gap. On the one hand, Illmo provides access to modern statistical methods that have become feasible because of the increased computing power, but does this through a dedicated GUI that allows to easily and quickly manipulate the available statistical options. A side-effect of the Illmo interface is that it helps users to understand the relationship between different statistical methods, which has potential educational benefits. Last but not least, Illmo also offers options that, as far as known, are not included in any existing statistical software.

One of the problems of existing methods is that many of them are overly specific, i.e., they offer exact solutions in cases where the data satisfy the assumptions underlying the methods. The most popular methods, i.e., Analysis of Variance (ANOVA) and linear regression, for instance assume that the data in the different observed conditions are normally distributed with distinct averages, but equal standard deviation (the latter condition is called heteroscedasticity). How to proceed when such conditions are unrealistic is less clear, so that most non-experts simply don’t bother, as was evidenced by the above example of the CHI’2009 papers. Another problem is that many users also have difficulties interpreting the outcomes of statistical methods, as they tend to use specialized vocabulary. While most authors are by now quite familiar with establishing whether or not differences are statistically significant, new APA (American Psychological Association) style guidelines require authors to also publish effect sizes, which creates problems of its own. For example, one widespread problem was reported in a recent paper by Nieuwenhuis et al. [5]. They observed that many (close to 50%) of the papers investigated incorrectly assessed the effect size of differences of differences (e.g., the difference between conditions 1 and 2 larger or smaller than the difference between conditions 3 and 4, and if so, by how much) (see also http://www.guardian.co.uk/commentisfree/2011/sep/09/bad-science-research-error).

The first step towards realizing Illmo was to settle on a way of making statistics understandable to a broad audience. Illmo doesn’t rely on dedicated methods but uses a more general, albeit slightly approximate, approach, known as log-likelihood modeling, pioneered by R.A. Fisher in the beginning of the 20th century [6, 4]. The most evident advantage of this approach is that users only need to get acquainted with a few basic principles (that are summarized in the next section) in order to understand a wide variety of statistical methods. The goal of the paper is not to settle the (somewhat philosophical) dispute on which approach is better, i.e. using an exact method for data that is expected to meet a priori (and often unchecked) assumptions, or an approximate method, which is asymptotically correct in case of a large number of observations, and that provides a uniform approach for a very large class of methods. For the time being, the primary goal is to offer an alternative for existing approaches. Another unique characteristic of Illmo is that it integrates all aspects of the statistical analysis in a single interface, including both the model specification and the data analysis output (in both textual and graphical form). Last but not least, in order to reduce problems with incorrect interpretation, statistical inference and effect size estimation are also supported through graphical means. Within the limited space of this short paper, only a few of the many features of Illmo can be illustrated.

2. THE LOG LIKELIHOOD PRINCIPLE

In Figure 1 we show the interface of the program Illmo after opening a data set (indicated by $A_{k_i}(j)$) consisting of repeated time measurements (repetition indexed by $j$) under 9 different conditions (indexed by $i$). The data was taken from an actual Fitts’ law experiment where all combinations of 3 target distances and 3 target sizes were offered to 8 subjects. In this example, the time measurements of different subjects are treated as repetitions (which corresponds to an across-subject analysis where there is only one value for the attribute index $k$). The example is used to illustrate the statistical principles that the users are expected to get acquainted with when using Illmo.

The log-likelihood principle means that Illmo relies on a goodness-of-fit measure, called the log-likelihood criterion (LLC), to quantify the difference between the transformed data (indicated by $A_{k_i}(j)$) and an assumed statistical model. Both the data transformation and the model may include parameters that are optimized to make LLC as small as possible, i.e., to obtain the best possible fit within the class of models under consideration. The default model is that the original data for distinct conditions, after possible transformation by a monotonically increasing function (that is independent of the observed condition $i$), is normally distributed with average values (indicated by $a_{ki}$) that are condition-dependent, and a fixed standard deviation $\sigma_k$. Other models that assume a different distribution or a condition-dependent standard deviation are however also possible, and can be assessed through a dialog box that is triggered by the ”Noise” button.

As indicated in the top-left of Figure 1, the LLC requires 3 inputs for its calculation, the (transformed) data $A_{k_i}(j)$, the noise model (including the standard deviation $\sigma_k$, as parameter) and the model averages $a_{ki}$. The optimization of all parameters is performed by Illmo upon loading the data and the result is presented in both graphical and textual form. Note that Illmo automatically selects the two graphical renderings that are deemed most relevant at this stage, but that there are many more options for the user to select from. The graph in the lower-left shows the cumulative histograms of both the actual (transformed) data $A_{k_i}(j)$ and the theoretical model, in casu a Gaussian distribution with average $a_{ki}$ and standard deviation $\sigma_k$. This allows the user to visually inspect the correspondence between actual data and statistical model. The graph in the lower-right shows a scatterplot of the actual data and the model averages in different conditions. As ANOVA analyses are still the standard, the text box in the middle offers such an analysis, from which for instance the correlation coefficient, which is frequently used to express the effect size [2], can be obtained. Note that by also offering such standard analysis results, the above-raised (philosophical) discussion is avoided. In this way, users can also assess the (usually small) differences between the alternative approaches.

More detailed information about the model fit can be obtained by clicking the LLC button to open the dialog box shown in Figure 2. More extensive statistical tests can for instance be triggered from within this dialog box, such as an ANOVA including post-hoc tests between all pairs of conditions using the T-test (for planned comparisons) or the Q-test (for unplanned comparisons). The LLC dialog box also provides access to one of the most powerful features of the LLC, and hence also of the Illmo program itself.
Figure 1: The Illmo interface after loading a data set with repeated time measurements. The upper part of the window shows the current model (which has a default linear transformation on the data) and the outcome of the LLC optimization. The bottom part of the window shows, from left to right: cumulative histograms, as derived from the data (black/green) and from the theoretical model (red/blue) - the case of the "fixed stimulus" (7) is highlighted in green/blue; a text window with the result of an ANOVA analysis; and a scatterplot showing both the individual data points (black) and the model averages (red).

any parameter in the model, as well as for some combinations of parameter values\(^1\), Illmo can calculate both the log-likelihood function (LLF) and the log-likelihood profile (LLP). The LLF shows how the LLC varies as a function of the selected parameter while keeping the other parameters in the model fixed to their optimal value. This is a fairly easy and fast computation. The LLP takes substantially more time to compute, as it shows how the LLC varies as a function of the selected parameter when the other parameters are re-optimized for each new value of this parameter. In the example of Figure 2, the displayed LLF and the LLP are for the difference between the averages in conditions 8 and 9. Estimates for the 95% confidence interval (CI) for this difference are obtained by intersecting these curves at the value of \(\chi^2_{0.05}(1) = 3.84\). In the example case, the LLF and the LLP almost overlap, which illustrates that this difference is as good as independent of the other parameters in the model. Note that the CI estimate derived from the LLP is always larger and a better estimate for the CI than the estimate derived from the LLF, but of course the former is derived at a much higher computational cost (which is why the calculation of the LLP can be switched on or off in the Illmo program). In this specific case of pairwise comparison of estimated averages, the exact boundaries of the 95% confidence interval are known and supplied by the T-test, with the result displayed in the text box. It is worthwhile to note that the process of statistical inference, which is often perceived as quite obscure by people who are not statistical experts, reduces to a simple inspection of how the shape of the LLC varies around its optimal value when using the log-likelihood approximation.

3. ADDITIONAL FEATURES

It is impossible to discuss all the existing and planned features of the Illmo program in the limited space available within this paper. We therefore suffice with a brief description:

- Non-parametric and nonlinear statistics are included by applying non-linear transformations on the data. The "BoxCox map" in Figure 1 applies a power-like transformation, characterized by a few parameters. The "spline" function can be either a rank-order transformation (in case of non-parametric statistics) or a monotonic spline function characterized by the slopes at a

\(^1\)Illmo can determine the LLF and LLP for a single average value \(a(i)\), for a difference \(d(i_2, i_1) = a(i_2) - a(i_1)\) in average values, and for two kinds of differences of differences, either involving three, \(d(i_2, i_1) - d(i_3, i_2)\), or four, \(d(i_2, i_1) - d(i_3, i_4)\), different conditions.
Figure 2: The dialog box that is triggered by the "LLC" button provides access to more advanced statistical analysis methods, such as a complete ANOVA analysis, including (post-hoc) pairwise tests. The buttons labeled "LLP" are used to request LLF and LLP calculations. The graph displays the result of such a calculation for the difference in averages for the pair of conditions (8,9). The quantitative results are reported in the text box, where the CI derived from the LLP can be compared to the CI according to a standard (planned) T-test or (unplanned) Q-test.

- Multiple and linear regression is available in Illmo (accessible through the "Model" tab) by specifying a linear relationship between the statistical averages $a_{ki}$ and a priori known stimulus characteristics. Multiple regression can be turned into multidimensional scaling (MDS) [1], by optimizing a common model across a number of attributes.

- Illmo can handle discrete data (such as Likert-scale data). This only requires a minor modification in the calculation of the LLC. Specifically, in order to establish the probabilities according to the model for the different discrete categories, the values of the cumulative distributions predicted by the model (as shown in Figure 1) need to be evaluated at the discrete data levels used (denoted by $Q_a(k)$ in Figure 1).

- The current version of the program only models attribute data, which arises when conditions are measured individually. Illmo will be extended to also include pairwise comparisons, where measurements are collected for pairs of conditions. The measurements may be either signed values (called comparisons, in which case the ordering in the pair of conditions matters) or unsigned values (called dissimilarities, in which case the ordering in the pair of conditions is irrelevant).

4. REFERENCES