TEACHING STATISTICIANS AND APPLIED RESEARCHERS STATISTICAL METHODS FOR ANALYSIS OF DATA FROM RATING SCALES. EXPERIENCES FROM JOINT RESEARCH COURSES IN RATING SCALE DATA ANALYSIS

Elisabeth Svensson
Örebro University
Sweden

The widespread use of rating scales in applied research fields implies need of statistical methods applicable to various types of studies involving ordinal response data. The aim of this paper is to present a teaching model of joining research courses in rating scale data analysis for statisticians and applied scientists together in order to stimulate inter-disciplinary communication. The participants experienced the complexity of applied research problems that involve subjective assessments on scales and also some of the possibilities and limitations of novel and classical statistical methods of analysis.

INTRODUCTION

Rating scales are commonly used as measurement instruments in various kinds of questionnaires and quality of life assessments but also in clinical trials for evaluation of outcome or functioning. There is also an ongoing development of statistical methods for analysis of ordinal data. Hence, the users of rating scales and the statisticians need each other. The information delay concerning methodological development of statistical methods and lack of inter-disciplinary communication can be bridged by means of joint courses for statisticians and users of rating scales.

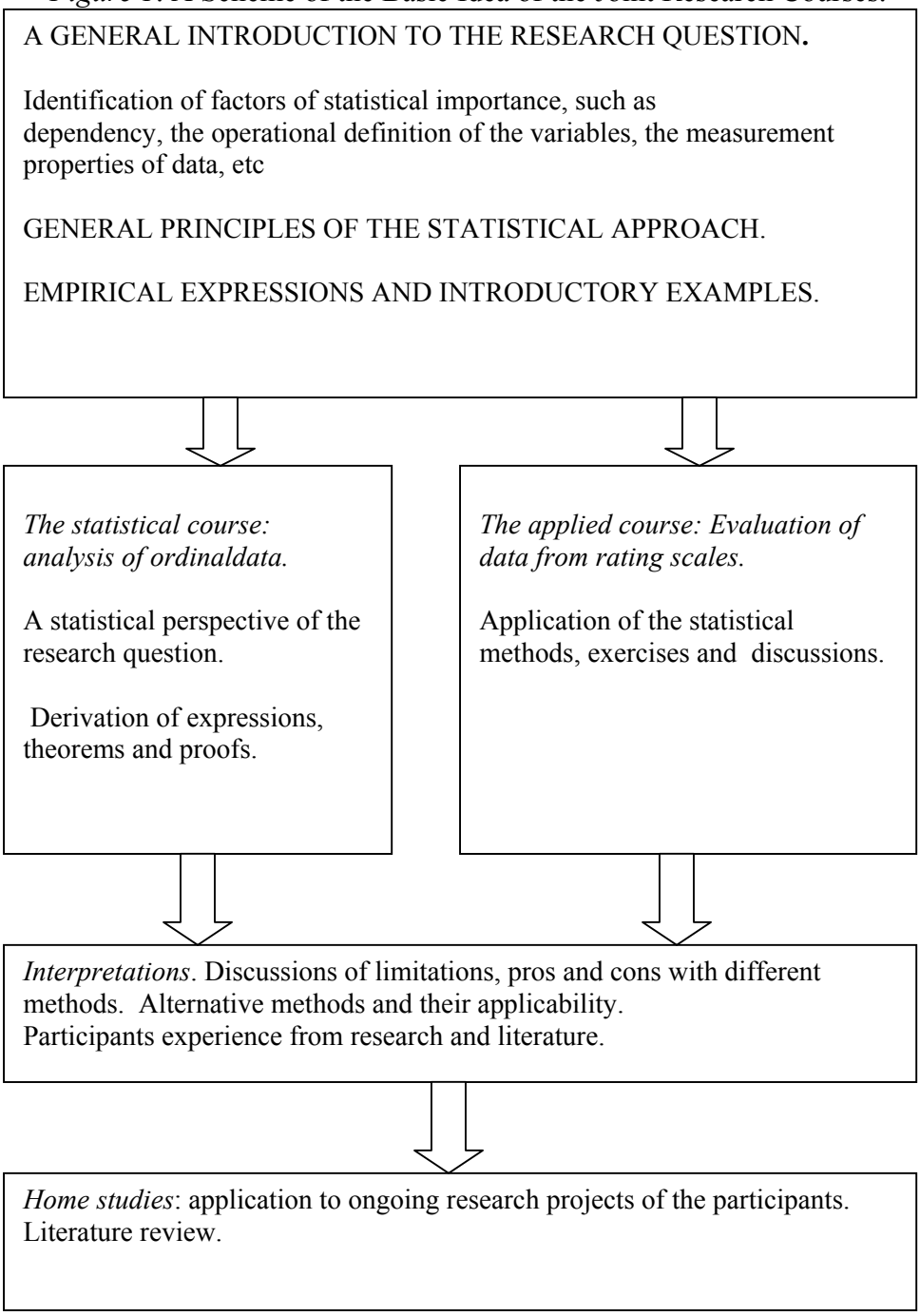
The aim is to present a teaching model in biostatistics that provides a link between biostatistical theory and applied sciences. The theoretical background of a statistical approach to analysis of rank-invariant data and its practical use were presented in joint research courses in rating scale data analysis for applied scientists and statisticians.

THE TEACHING MODEL

The idea of running joint research courses for statisticians and applied scientists is globally applicable and concerned in this case, courses in rating scale statistics with focus on statistical methods for evaluation of paired ordinal data. Ordinal data are invariant under monotone ordered transformations, which means that categorical labels, even ordinal numbers, indicate only a rank order and not a mathematical value (Stevens, 1946). Thus, the results of statistical treatments used should not be affected by any kind of ordered relabelling of the scale categories. This means that statistical methods that are based on calculations of sums of and differences between categories are not suitable.

Little attention is paid in basic statistical books and courses for statisticians and non-statisticians to the relationship between the measurement properties of data and the choice of statistical methods (Hand, 1996). The overall aim of the joint research course is that the participants shall gain a mutual understanding of a research question and of the measurement process behind the rating scale data, but also of the fact that the measurement properties of ordinal data influence the choice of statistical methods of analysis. The participants shall also gain knowledge about various statistical approaches for analysis of ordinal data obtained from quality assessments, follow-up studies and parallel group designs. The course is interactive in the sense that the users of rating scales must apply statistical methods presented in the course, when appropriate, to their own studies with assistance from the participating statisticians. There are also subject specific parts in the course; a mathematical statistical part for the statisticians, and a part of problem solving and application of statistical methods to various types of rating scale assessments for the applied scientists, see Figure 1.

Figure 1. A Scheme of the Basic Idea of the Joint Research Courses.



MAIN TOPICS

The main topics are the same in both courses and the general principles are presented and discussed by all participants and discussed before the separation. The course for statisticians contains mathematical derivations and methodological discussions, while the course for applied scientists focuses on understanding, practical use and interpretation.

The measurement process in qualitative assessments. The operationalisation process is to identify observable indicators of the concept to be studied and leads to the choice of measurement instrument. The lack of standardised rules of measuring qualitative variables implies that a large variety of multi-item instruments and single scales are used for the same variable. The use of multi-dimensional instruments, parallel item scales and hierarchical scales for assessment of various types of variables is discussed. The measurement properties of the responses are commonly identified as being binary, categorical or ordinal. Various approaches for construction of global scores for multi-item scales are also discussed (Svensson, 1998a; Svensson, 2001).

Statistical descriptions of uni- and bivariate data sets. Different types of tables, graphs and measures of describing ordinal data and dichotomised data are presented, discussed and interpreted.

Statistical methods for binary data. Statistical methods for binary and dichotomised data from dependent and independent groups of observations are presented and interpreted. In the course for statisticians various approaches to dichotomise ordinal data and the use of relative risks, odds ratios and logit links were thoroughly discussed (Agresti, 1984).

Statistical methods for inter- and intra-rater agreement. This is one of the main topics of the course and presents the rank-invariant approach for a comprehensive evaluation of systematic and occasional disagreement in paired assessments (Svensson,1993; Svensson, 1998b). Quality concepts of scale assessments and classical methods are also discussed (Agresti, 1988; Agresti, 1992; Svensson, 1993). As illustrated by Figure 1, parallel lectures of the mathematical and the applied approach are performed.

MEASURES OF INTER-SCALE CONCORDANCE

Classical methods and a new approach to evaluation of order consistency in scale assessments (Agresti, 1984; Svensson, 2000a, Svensson 2000b) are presented. Some of the measures in the novel rank-invariant approach can be added to the correlation family suggested by Daniels (1944), which is discussed among the statisticians.

EVALUATION OF CHANGE IN ORDINAL ASSESSMENTS

Statistical methods for analysis of change in ordinal data with regard to the rank-invariant properties were presented (Sonn & Svensson, 1997; Svensson, 1998c)

EXERCISES

All exercises had to be documented and discussed among all participants. The exercises A-C can be modified to suit statisticians. The statisticians also shall perform additional mathematical exercises within each topic.

A. Give a short written presentation of *the research project*, especially concerning the design and the use of measurement instruments and scales.
B. Report on the *measurement process* especially concerning the operationalisation of the qualitative variables used in the research project.
C. Review at least 10 *scientific papers* from the research field of interest with focus on the main purpose of the study, the design and the measurement process. Present the scales, instruments and the statistical methods used in a table and note, if any, motivations for the choice of statistical methods.
D. *Applied scientists*: make a list of the most common methods for evaluation of reliability and validity of rating scale assessments and for analysis of ordinal data in various types of studies that a found in papers and/or experienced. *Statisticians:* present the statistical assumptions on data for their use.

THE FINAL EXAMINATION

The examination in the applied course is optional. The participants can choose an ordinary written examination or apply appropriate statistical methods for paired data to their own research problem, when applicable. The participating statisticians are allowed, to some extent, to act as consultants for the applied students to support calculations of the measures and standard error estimates.

EXPERIENCES GAINED FROM THE JOINT RESEARCH COURSE 2001

Thirty participants (5 statisticians) fulfilled the joint research course including the final examination. They represented a range of disciplines such as physiotherapy, occupational therapy, sociology, nursing research, psychology, environmental medicine, agriculture and sports. The participants have asked for this course and they travelled from different universities in Sweden.

Some of the doctoral students from the applied course presented the result of the examinations as working papers in a symposium open for statisticians and scientists in applied research fields. The presentations concerned for example evaluation of change in a variable measured by scales, inter-rater agreement, order-consistency between visual analogue scale assessments and discrete scale assessments of the same variable and comparisons of sum scores and alternative treatments of data from multi-item instruments.

CONCLUSIONS

The interactive, inter-disciplinary combined course was very stimulating for all participants but also valuable for the ongoing research. Several participants increased their self-confidence concerning the handling of data from rating scales As the methods were applied to specific research questions of the participants throughout the course, they realised the pros and cons regarding relevance, interpretability, information and utility with different statistical methods. The wide range of measurement instruments represented in the course, the integration of statisticians and the interdisciplinary composition of other participants and that the participants were on different levels in the research process were not only an educational challenge but also illustrated the complexity of solving problems that include rating scales.

REFERENCES

Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: John Wiley & Sons.

Agresti, A. (1988). A model for agreement between ratings on an ordinal scale. *Biometrics*, *44,* 539-548.

Agresti, A. (1992). Modelling patterns of agreement and disagreement. *Statistical Methods in Medical Research*, *1*, 201-218.

Daniels, H.E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, *33*, 129-135.

Hand, D.J. (1996). Statistics and the theory of measurement. *J R Statist. Soc A*., *159*, 445-492.

Sonn, U., & Svensson, E.(1997). Measures of individual and group changes in ordered categorical data: application to the ADL Staircase. *Scand J Rehab Med*, *29*, 233-242.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, *103*, 677-680.

Svensson, E. (1993). *Analysis of systematic and random differences between paired ordinal categorical data* (dissertation). Stockholm: Almqvist & Wiksell International.

Svensson, E. (1998a). Teaching the measurement process in biostatistics. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W.K. Wong (Eds.), *Proceedings of the 5th International Conference on Teaching Statistics* (pp. 1257- 1262). Singapore: International Statistical Institute.

Svensson, E. (1998b). Application of a rank-invariant method to evaluate reliability of ordered categorical assessments. *Journal of Epidemiology and Biostatistics*, *3*, 403-409.

Svensson E. (1998c). Ordinal invariant measures for individual and group changes in ordered categorical data. *Statistics in Medicine*, *17*, 2923-2936.

Svensson, E. (2000a). Comparison of the quality of assessments using continuous and discrete ordinal rating scales. *Biometrical Journal*, *42*,417-434.

Svensson, E.(2000b). Concordance between ratings using different scales for the same variable. *Statistics in Medicine*,*19*, 3483-3896.

Svensson, E. (2001). Construction of a single global scale for multi-item assessments of the same variable. *Statistics in Medicine*, *20*, 3831-3846.