February 25, 2021

UMEÅ UNIVERSITY

Review of Adéla Hladká doctoral thesis:
Statistical models for detection of differential item functioning.

Marie Wiberg, Professor, Umeå University, Sweden

This thesis is well written about the important topic Differential Item Functioning (DIF), i.e. when an item function differently for different groups of the same ability. The different groups can for example be men and women. Hladká has managed to fill an important research gap in DIF research which is impressive especially as there are already a large number of DIF methods around. This thesis contains several new scientific results and has several strengths in its four chapters. From her thesis it is obvious that she has a deep understanding of the field and the mathematics and statistics in the thesis are theoretically sound.

In the first chapter, generalized logistic regressions models for detecting DIF for binary items are proposed. This is an extension from the usual logistic regression method for detecting DIF in binary items. The specific new parts are the inclusion of guessing and/or inattention. The inclusion of guessing and/or inattention is important as that has been lacking in previous DIF methods. Hladka also presents how to estimate the item parameters as well as two new algorithms which simplifies the achievement of estimates. One is an EM algorithm and the other one builds on the parametric link function. Her simulation studies illustrate that the proposed methods work well in comparison to previous methods. A special strength of this chapter is the implementation of the proposed models into the freely available software R in terms of the difNLR package. The clear descriptions make the method available for a larger audience. Overall, the proposed methods have good properties and work well in the presented simulation studies. Hladká also gives clear descriptions of her derived results, for example of the asymptotic properties and the estimation methods.

In the second chapter generalized logistic regression models for DIF detection for polytomous items is proposed. This chapter is basically an extension of the first chapter for the case of polytomous items. The chapter include both theoretical and applied work in terms of the R package difNLR and examples and simulation studies are given which are very useful if you want to use the proposed methods. Again, the derived results are important and presented clearly and I believe this research is useful for many other researchers. In the future it would also be interesting if she would explore for example the use of the generalized partial credit model which is useful for many items which are part of tests given at schools.

The third chapter contains a nonparametric comparison of regression curves when detecting DIF based on kernel smoothing. In this chapter, an estimate of an optimal weight function for a test statistic to identify DIF is proposed and discussed. The chapter is interesting and gives valuable results as a strength of using a nonparametric method is that one does not need to know the functions exact shape and if there are between groups differences one can still detect them

even if a parametric model is not fitted. The nonparametric approaches are compared with logistic regression in a simulation study which illustrates the usefulness for detection of DIF.

The final fourth chapter focuses on DIF detection when one examines the items, item by item and decides whether to keep the item or not in the test. The novel approach here is that item purification and multiple comparison corrections are examined at the same time which has not been done before as these parts are usually examined separately. A simulation study is used to examine this important combination. The DIF effect sizes are also discussed. This chapter is interesting and very important as it guides and help both researchers and practioners when working with items. In the future it would be interesting if she could implement these ideas with her proposed method in the previous chapter.

Overall, this thesis contains several well conducted simulations studies where different conditions are explored. The new methods are compared with several well-known methods so it is easy to compare the properties of the proposed method with previous methods. The results show that the new methods could be preferred in the examined situations. A strength with Hladkás thesis is that she does not only gives the mathematics and statistics behind the models, she also has taken an active part of developing an R package so the methods are easy to carry out in practice. This is very important as if we do not make new methods easily available to other researchers it is a large risk that new and improved methods are never used by others. Another strength is the included easy to follow examples with actual code and explanations. The research in the thesis is important in the DIF methods field as there has not been well developed methods for the situation, especially polytomous items, which Hladká is proposing. The implementation in a freely available software is of high importance as it facilitates the spread of the methods outside the area of statistics and educational tests. Thus, the results from this thesis can be used in all areas which conduct surveys with questionnaires of the item types examined in the thesis. In the future I would suggest that more things are added to the R package. For example, to allow to use IRT ability as a matching criterion would be very useful.

Summing up, this thesis is very strong as it has new theoretical proposals with sound properties, large simulation studies to examine different condition and good description of usable R code. Some of the papers are already published in international peer reviewed journals and the content of the thesis is scientifically sound as well as contains new important results. The examined topics and how it has been examined are very convincing and from my point of view the thesis conform with the requirements in our field and proves that the author has a high ability for creative scientific work. I consider the thesis by Hladká to be of high quality and definitely meets the requirement for a doctoral dissertation

Yours sincerely,

Marie Wiberg
Professor in Statistics