lrdif: A SAS macro for logistic regression tests for differential item functioning

Karl Bang Christensen
National Institute of Occupational Health, Denmark
May 27, 2004

Abstract
This note describes a SAS macro that can be used for logistic regression tests for differential item functioning.

1 Introduction
Absence of differential item functioning is a basic assumption in most item response theory models, and a necessary requirement in scale validation. This note describes a SAS macro to test for DIF using logistic regression techniques (Swaminathan & Rogers, 1990; Zumbo, 1999).

2 Logistic regression tests for differential item functioning
Let $x_1,\ldots,x_k$ denote the answers to a set of items and let $y_1,\ldots,y_l$ denote a set of exogenous variables like gender and age. The basic assumption behind scale construction is the existence of an underlying unobservable (latent) variable $\theta$ and requiring absence of differential item functioning can be written

$$p(x_i|y_j, \theta) = p(x_i|\theta),$$
this means that given the latent variable $\theta$, the item responses are independent of exogenous variables: item responses differ between groups (e.g. gender, age or job groups) only through differences in the latent variable. Logistic regression analysis is often used to investigate the relationship between discrete ordinal responses (for example, normal, mild, and severe) and a set of explanatory variables. Logistic regression models can thus be used to test that there is no association between item responses and exogenous variables when controlling for the total score, i.e. the absence of differential item functioning (Swaminathan & Rogers, 1990; Zumbo, 1999).

2.1 Syntax

The macro fits a proportional odds model (McCullagh, 1980) using the LOGISTIC procedure in SAS, and is called with the statement

```sas
%lrdif(data=test, items=ITEM1 ITEM2 ITEM3, exo=GENDER AGE, outfile=LRDIF, plot=NO);
```

where the data set 'data' contains the 'items' (that are coded 0, 1, 2, . . .), the exogenous variables 'exo' and the results are written to the data set 'outfile'. The number of cases used in each analysis is also included in the 'outfile'. For dichotomous exogenous variables the macro also includes an option (plot=YES) for plotting observed mean item scores (and 95 % confidence limits for the mean) stratified by the total score for the two values of the of the exogenous variable.

3 Comments

Problems have been known to occur if variable names are longer than eight characters.

References
