CONFIRMATORY ANCOVA: Manual for the Use of Selection.exe (version 2.1)

Introduction

This manual explains how to use the accompanying program "Selection.exe". This program calculates Bayes factors for a set of models that are defined by (about) equality and/or inequality constraints on the (adjusted) means. For details of the method, please read Klugkist, Laudy, and Hoijtink (KLH2005; see reference below) or Chapter 4 and 10 in:

Bayesian Evaluation of Informative Hypotheses (2008). Hoijtink, H., Klugkist,I., Boelen, P.A. (Eds.), Springer: New York. (HKB2008)

This program is free. However, when publishing results obtained with this program, you are obliged to refer to:

Klugkist, I., Laudy, O. and Hoijtink, H. (2005). Inequality Constrained Analysis of Variance: A Bayesian Approach. Psychological Methods, 10, 477-493.

This is *version 2.1* of the software. It can handle maximally 2 covariates. In the future updates may become available. This tutorial will describe and illustrate the options available in *version 2.1*.

VERSION 2.1:

1. In *version 2.0*, a small bug was found that caused a failure to produce results in rare cases. In *version 2.1* this error has been fixed.

COMPARED TO VERSION 1.0, SOME IMPORTANT CHANGES ARE MADE:

2. In *version 2.0*, hypotheses that contain a mix of equalities and inequalities can also be specified.
3. In *version 2.0*, the user can choose between 'relevant differences' between means (see KLH2005) or specify a difference between means of zero, i.e., an approximation of the strict equality (see Chapter 4 (Section 4.3.3) of HKB2008).
4. *Version 2* is not restricted to a maximum of 8 groups.

Installation

After unzipping (use 'extract to folder…'), three examples are provided in three subfolders. This tutorial will guide you through the examples in the three folders. To do analyses on your own data, copy one of the folders, and modify *data.txt* (the data-file[1]) and *ini-selection.txt* (input file for selection.exe) such that they apply to your data. Note that:

- The names *data.txt* and *ini-selection.txt* are fixed and cannot be changed. These files have to be text files (also known as ASCII files).
- The format of these files should not be changed, that is, do not add empty lines, or delete the first line containing labels.
- The data-file *data.txt* should be complete, that is, missing data are not allowed.

[1] In, for instance, SPSS you can save data (after deleting all columns that are not of interest and placing the remaining columns in the correct (see below) order) as an ASCII file via "save as" and name the resulting file *data.txt*.

Depending on the speed of your computer, computations may take a while. The software will regularly inform you about its progress.

Example1_anova

Go to the subfolder *Example1_anova*. It contains simulated data from a population in which the means for group 1 to 4 are ordered from smallest to largest.

Open *data.txt*. Columns in the data-file *have to be* separated by (one or more) spaces.

- The first column contains the number of the group to which a person belongs. With four groups the numbers *must be* 1, 2, 3 and 4.
- The second column contains the score on the dependent variable.

Close *data.txt.*

Open *ini-selection.txt*.
The first lines of this file contain input for *selection.exe.* Below the exclamation marks (!!!!!), an explanation of the required input is given. Study this file to understand the input. Note that for each constrained hypothesis, an Ngroup × Ngroup matrix must be provided, with one empty line between subsequent matrices. The first matrix in this example is:

```
0 -1 -1 -1
1  0 -1 -1
1  1  0  -1
1  1  1   0
```

The first row provides the constraints of, respectively:
- μ_1 with μ_1 (this value must be zero, i.e., no constraint of a mean with itself)
- μ_1 with μ_2 (here the -1 implies: $\mu_1 < \mu_2$)
- μ_1 with μ_3 (here the -1 implies: $\mu_1 < \mu_3$).
- μ_1 with μ_4 (here the -1 implies: $\mu_1 < \mu_4$).

The second row states that:
- $\mu_2 > \mu_1$ (the value 1 means 'greater than')
- μ_2 has no constraint w.r.t. itself (a zero; as must always be the case on the diagonal)
- $\mu_2 < \mu_3$ and $\mu_2 < \mu_4$ (the -1 signs).

The third row states that $\mu_3 > \mu_1$, $\mu_3 > \mu_2$ and $\mu_3 < \mu_4$.
The last row states that $\mu_4 > \mu_1$, $\mu_4 > \mu_2$ and $\mu_4 < \mu_3$.

Summarizing: the first hypothesis states: $\mu_1 < \mu_2 < \mu_3 < \mu_4$

In a similar way, the second matrix states: $\mu_1 > \mu_2 > \mu_3 > \mu_4$

The third hypothesis in the example is that all four means are equal: $\mu_1 = \mu_2 = \mu_3 = \mu_4$.

Note that one Delta-value must be provided (in this example it is set at zero, that is, exact equalities are approached). This value is used for all hypotheses with equality constraints (here, just the third model). *Delta* can also denote the largest difference between two group means that are considered irrelevant (Delta>0).
Close *ini-selection.txt.*

Run *selection.exe* (by double-clicking[2]). The output file *output.txt* and *technical report.txt* will be created in the folder you are working in.

- In *output.txt*, Bayes factor values are provided, for each hypothesis that was specified in *ini-selection.txt* with the unconstrained, encompassing hypothesis.
- In *technical report.txt* additional information is reported, e.g. the encompassing prior that is used and some computational details.

The resulting Bayes factors show a clear support for the first hypothesis.

The calculation of posterior model probabilities for a set of models, from the Bayes factors provided in the output, is described in both KLH2005 (page 485) and HKB2008 (Chapter 4, page 64).

Example2_ancova

Open *data.txt.*
As can be seen in the first column, there are 4 groups. The second column contains the scores on the outcome variable (Y), and the third column the covariate (X1). Note that maximally 2 covariates can be added. The order of the columns must always be (with one or more spaces between columns):
Group Y X1 X2 (optional 2nd covariate)
Close *data.txt.*

Study *ini-selection.txt* to understand the input for *selection.exe*.
Matrix 1 specifies: $\mu_1 = \mu_2 = \mu_3 = \mu_4$;
Matrix 2 specifies: $\{\mu_1, \mu_3\} < \{\mu_2, \mu_4\}$;
Matrix 3 specifies: $\{\mu_1 = \mu_3\} < \{\mu_2 = \mu_4\}$;
Where each μ_j (j=1, …, 4) now represents a covariate adjusted mean (or -if the covariate is not centered- a covariate adjusted intercept; see HKB, Chapter10, p.212-213).
Note that equality constraints are evaluated as strict equalities (Delta = 0).
Close *ini-selection.txt.*

Run *selection.exe* and study the results in *output.txt* (and *technical report.txt*).

As can be seen, the first model is not supported by the data, but models 2 and 3 both are (BF>1). The support for the third hypothesis is much stronger than for the second (both seem to be supported by the data, but the third is much more specific).

The calculation of posterior model probabilities for a set of models, from the Bayes factors provided in the output, is described in both KLH2005 (page 485) and HKB2008 (Chapter 4, page 64).

[2] If the software does not run but gives a warning about a missing dll-file instead, write down the name of the file and search the internet with the terms: *name*.dll download. Download the dll file and save it in the folders with the program selection.exe. Try to run selection.exe again by double-clicking.

Example3_twoway_anova

Open *data.txt.*
As can be seen in the first column, there are 6 groups. The second column contains the scores on the outcome variable.
Close *data.txt.*

Study *ini-selection.txt* to understand the input for *selection.exe*.
The matrices representing the hypotheses/models are constructed assuming that the 6 means are organized in a two-way table as follows:

1	2	3
4	5	6

The first inequality constrained model (the first matrix) specifies the simple column effects: $\{\mu_1 < \mu_4\}$, $\{\mu_2 < \mu_5\}$, $\{\mu_3 < \mu_6\}$.

The second model specifies increasing means for the first row, but about equal means (Delta=0.5) for the second row (i.e., differences between means smaller than 0.5 are considered irrelevant): $\{\mu_1 < \mu_2 < \mu_3$, $\{\mu_4 = \mu_5 = \mu_6\}$.

Run *selection.exe* and study the results in *output.txt* (and *technical report.txt*). As can be seen, both models are clearly supported by the data, and the support for the second hypothesis is larger than for the first. The calculation of posterior model probabilities for a set of models, from the Bayes factors provided in the output, is described in both KLH2005 (page 485) and HKB2008 (Chapter 4, page 64).

> **Exercise:**
> Modify *selection.exe* by adding a third and fourth hypothesis. Specify the constraints matrices such that the third hypothesis combines the row and column effects as specified above. Let the last hypothesis state that all 6 means are about equal.
>
> **Answer**: the file *answer.txt* provides the correct ini-file. Note that to be able to run this ini-file, it has to be re-named *ini-selection.txt* (in which case it will overwrite the original ini-file).

Contact

If you run into problems using this software, please write an e-mail to i.klugkist@uu.nl containing:

- Your *data.txt* and *ini-selection.txt*.
- The version-number (this tutorial refers to Version 2.1) of the software.
- The problem.

If you want to do analyses that are not possible with this software, e-mail to i.klugkist@uu.nl and shortly describe your data and research-questions.