



A Fortran 90 Program for Confirmatory Analysis of Variance

Rebecca M. Kuiper
Utrecht University

Irene Klugkist
Utrecht University

Herbert Hoijtink
Utrecht University

Abstract

There are different confirmatory techniques to compare means, like hypothesis testing and (Bayesian) model selection. However, there is no software package in which these techniques are available. A Fortran 90 program is written, which enables researchers to apply these techniques to their data. Besides traditional hypotheses, like $H_0 : \mu_1 = \mu_2 = \mu_3$ and $H_u : \mu_1, \mu_2, \mu_3$, order-restricted hypotheses, like $\mu_1 > \mu_2 > \mu_3$ or $\mu_1 > \mu_2 = \mu_3$ or $\mu_1 > \mu_2 < \mu_3$, can be evaluated.

Keywords: (Bayesian) model selection, comparison of means, confirmatory analysis, Fortran 90, hypothesis testing, inequality constraints.

1. Introduction

Often researchers have a theory with respect to the ordering of the means in the experiment. These theories can be written as order-restricted hypotheses (e.g., $\mu_1 > \mu_2 > \mu_3$) and can be tested with confirmatory methods. In the context of comparing independent means, that is, analysis of variance (ANOVA), three approaches are distinguished:

- [Silvapulle and Sen \(2005, pp. 25–42\)](#) present the \bar{F} test. There are two types of \bar{F} tests. Namely the *ordered alternative* and the *ordered null*. In the *ordered alternative*, the classical null (H_0) is tested against an order-restricted hypothesis, for example, H_1 . In the *ordered null*, an order-restricted hypothesis, like H_1 , is tested against the classical alternative (H_u).
- [Anraku \(1999\)](#) introduces the order-restricted information criterion (ORIC). The ORIC can be used to select the best of a set of order-restricted hypotheses, like the hypotheses in (2).

- Klugkist, Laudy, and Hoijtink (2005) present a Bayesian model selection (BMS) criterion, which can be used in the same context as the ORIC.

For these approaches, user-friendly software is not available. In this paper software is introduced with which the three confirmatory approaches can be executed.

The model used in this paper and in the software is the ANOVA model:

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad (1)$$

where $i = 1, \dots, k$, $j = 1, \dots, n_i$, y_{ij} is the j th observation of the dependent variable for group i , which has n_i observations, μ_i is the mean of group i , and ϵ_{ij} is an error term. The error terms are independently and normally distributed with expected value 0 and variance σ^2 , that is, $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

As an example, consider the simple ANOVA with five groups, presented by Lucas (2003). In Section 3, the theoretical background of his research will be elaborated. Lucas expresses clear theories with respect to the ordering of means, leading to the following specific hypotheses:

$$\begin{aligned} H_0 : & \quad \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5, \\ H_1 : & \quad \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2, \\ H_2 : & \quad \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2, \\ H_3 : & \quad \mu_1, \mu_2, \mu_3, \mu_4, \mu_5. \end{aligned} \quad (2)$$

The hypotheses H_0 and H_3 are the classical hypotheses, the other two are order-restricted hypotheses. It is also possible to specify a set of models/hypotheses without the classical null $H_0 : \mu_1 = \dots = \mu_k$ and/or the classical alternative $H_u : \mu_1, \dots, \mu_k$. We recommend to include H_u (when doing model selection) as a safeguard for choosing a weak hypotheses (Kuiper and Hoijtink 2010). Furthermore, one should include H_0 only when there is real interest in H_0 .

Although software for exploratory approaches is widely available – e.g., classical hypothesis testing in SPSS (SPSS Inc. 2006) and model selection using information criteria, like AIC, in R with the package **nlme** (Pinheiro, Bates, DebRoy, Sarkar, and R Development Core Team 2009) – this is not the case for confirmatory approaches, that is, for evaluation of the four hypotheses in (2). The software presented in this paper can evaluate different types of hypotheses which can be formulated by $A\boldsymbol{\mu} \geq 0$, for some matrix A in which each row is a permutation of the k -vector $(-1, 1, 0, \dots, 0)$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^\top$.

Note that also factorial ANOVA models fit in the framework presented. For instance, a 2×3 -design can be represented by (1) with $k = 6$. Specific hypotheses about the six group means can again be formulated by $A\boldsymbol{\mu} \geq 0$ (as explained earlier). In a standard two-way ANOVA, three hypotheses are tested concerning the presence of a main effect of the first factor, a main effect of the second factor, and the presence of an interaction effect. If one or more of these effects are found, further evaluation is required to describe the direction of the effects (making the approach exploratory). In a confirmatory approach, in contrast to a exploratory approach, expected patterns are specified beforehand. For instance, $H : \{\mu_1 = \mu_2 = \mu_3\}, \{\mu_4 < \mu_5 < \mu_6\}$, is a prespecified and specific interaction effect. Competing (interaction) effects can also be specified.

In the next section, subsequently, the \bar{F} test, the ORIC, and Bayesian model selection will be shortly explained. In Section 3, a practical example is provided and analyzed using each of the three approaches. The appendix contains a user manual for the software.

2. Three confirmatory techniques for comparing means

2.1. Hypothesis testing using the F-bar (\bar{F}) statistic

In classical statistical testing the hypothesis *all means are equal* is tested against the alternative *not all means are equal*. This is usually tested with an F test using a one-way ANOVA. However, often researchers want to test a certain order restriction, because of a theory with respect to the order of the means in the experiment. See, for example, H_1 and H_2 in (2).

In [Silvapulle and Sen \(2005, pp. 25–42\)](#) the F test is modified, such that an order-restricted hypothesis can be tested. This test is called the F-bar (\bar{F}) test. It is possible to test the null hypothesis *all means are equal* (H_0) against an *ordered alternative*, like H_1 , and it possible to test an *ordered null* (H_1) against the alternative *all parameters are free* (H_u).

The \bar{F} test statistic is calculated by: $\bar{F} = \frac{RSS(H_{null}) - RSS(H_{alt})}{S^2}$, where $RSS(H)$ is the residual sum of squares under hypothesis H and $S^2 = (n_1 + \dots + n_k - k)^{-1} \sum_i \sum_j (y_{ij} - \bar{y}_i)^2$ is the mean square error, with $n_1 + \dots + n_k - k$ the error degrees of freedom. This is applied to the two types of test. For each of the two tests the $RSS(H_{null})$ and $RSS(H_{alt})$ will be elaborated on.

The first test is the *ordered alternative*, in this test $H_0 : \mu_1 = \dots = \mu_k$ is tested against an order restriction of the form $H_1 : A\boldsymbol{\mu} \geq 0$, for some matrix A in which each row is a permutation of the k -vector $(-1, 1, 0, \dots, 0)$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^\top$. In this test,

$$RSS(H_{null}) = \sum_i \sum_j (y_{ij} - \bar{y})^2,$$

where \bar{y} is the overall mean, and

$$RSS(H_{alt}) = \sum_i \sum_j (y_{ij} - \tilde{\mu}_i)^2,$$

where

$$\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_k)^\top = \operatorname{argmin}_{\boldsymbol{\mu} \in H_1} \sum_i \sum_j (y_{ij} - \mu_i)^2.$$

Since $\sum_i \sum_j (y_{ij} - \mu_i)^2$ can be rewritten as

$$\sum_i \sum_j (y_{ij} - \mu_i)^2 = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 + \sum_i n_i (\bar{y}_i - \mu_i)^2 = C(\mathbf{y}) + q(\boldsymbol{\mu}),$$

where \bar{y}_i is the i th group/treatment mean, \mathbf{y} is the matrix consisting of the elements y_{ij} , and

$$q(\boldsymbol{\mu}) = \sum_i n_i (\bar{y}_i - \mu_i)^2 = (\bar{\mathbf{y}} - \boldsymbol{\mu})^\top \operatorname{diag}\{n_1, \dots, n_k\} (\bar{\mathbf{y}} - \boldsymbol{\mu}),$$

it holds that

$$\tilde{\boldsymbol{\mu}} = \operatorname{argmin}_{\boldsymbol{\mu} \in H_1} q(\boldsymbol{\mu}). \quad (3)$$

This constrained minimization problem, where the objective function $q(\boldsymbol{\mu})$ is quadratic in $\boldsymbol{\mu}$ and the (equality and inequality) constraints are linear in $\boldsymbol{\mu}$, is a *quadratic programming*

problem. There are efficient computer algorithms for this minimization problem, here the **IMSL** subroutine **QPROG** (Visual Numerics 2003, pp. 1307–1310) is used in the Fortran 90 program.

The second test is the *ordered null* in which a null of the form $H_1 : A\boldsymbol{\mu} \geq 0$ (as explained earlier) is tested against the alternative *no restrictions on the $\mu_i \forall i$* , that is, $H_u : \mu_1, \dots, \mu_k$. Then,

$$RSS(H_{null}) = \sum_i \sum_j (y_{ij} - \tilde{\mu}_i)^2$$

and

$$RSS(H_{alt}) = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2.$$

As with classical hypothesis testing, p values must be determined. The exact p value, for the \bar{F} , can be obtained via simulation. In the ANOVA model, the errors are normally distributed, that is $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$. Therefore, the simulation consists of the following three steps (Silvapulle and Sen 2005, pp. 32–33 and 40):

1. Generate independent observations z_{ij} ($i = 1, \dots, k$ and $j = 1, \dots, n_i$) from the standard normal distribution $\mathcal{N}(0, 1)$.
2. Compute \bar{F} for the generated data.
3. Repeat the previous two steps R_p times. In the program, the default value of R_p is $R_p = 100,000$. Calculate the number of times the \bar{F} statistic, calculated in Step 2, exceeds the sample value of the \bar{F} statistic, this number is denoted by M . The p value is calculated by M/R_p .

When the p value is smaller than the nominal α -level, often set equal to 0.05, the null hypothesis is rejected. Thus, in the *ordered alternative*, when $p < \alpha$, H_0 is rejected and in the *ordered null*, when $p < \alpha$, the order-restricted hypothesis is rejected.

2.2. Model selection using order-restricted information criterion

Anraku (1999) proposes the order-restricted information-criterion (ORIC). It can be used to select the best of a set (\mathcal{M}) of models/hypotheses H_m , $m \in \mathcal{M}$. The set of hypotheses can contain H_0 , H_u , and order-restricted hypotheses of the form $A\boldsymbol{\mu} \geq 0$, for some matrix A in which each row is a permutation of the k -vector $(-1, 1, 0, \dots, 0)$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^\top$.

Like other information criteria, the ORIC is based on the log likelihood ($\log L$) and a penalty term (PT): $ORIC = -2 \log L + 2PT$. The hypothesis with the smallest ORIC value is the preferred hypothesis.

The maximum likelihood estimators (mle's) $\hat{\boldsymbol{\mu}}_m$ and $\hat{\sigma}_m^2$ for H_m , $m \in \mathcal{M}$ are the values of $\boldsymbol{\mu}$ and σ^2 , respectively, that maximize the log likelihood for H_m :

$$\log L(\hat{\boldsymbol{\mu}}_m, \hat{\sigma}_m^2 | y) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\hat{\sigma}_m^2) - \frac{1}{2\hat{\sigma}_m^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_{mi})^2, \quad (4)$$

where $\hat{\sigma}_m^2 = \frac{1}{N} \sum_i \sum_j (y_{ij} - \hat{\mu}_{mi})^2$, and $N = \sum_i n_i$. Because of the order restrictions, the order-restricted mle $\tilde{\boldsymbol{\mu}}_m$ must be found (Anraku 1999). Since the term $\hat{\sigma}_m^2$ cancels out the

term $\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_{mi})^2$, the maximization of the likelihood actually comes down to minimizing $\log(\sigma_m^2)$, subject to $\mu_i - \mu_{i'} \geq 0$, for $i, i' = 1, \dots, k$ (Silvapulle and Sen 2005, pp. 42–43). Which results in the order-restricted mle $\tilde{\boldsymbol{\mu}}_m$ as defined in (3).

The penalty term of H_m (PT_m) is equal to:

$$PT_m = 1 + \sum_{l=1}^{q_m-1} LP_{ml}(q_m - 1, V_m) l,$$

where V_m is explained below, $LP_{ml}(q_m - 1, V_m)$ is a level probability, that is, the probability that there are l distinct mean values / levels among the $q_m - 1$ means in the order-restricted mle, and $q_m - 1 \leq k$ the number of distinct μ_i 's in Hypothesis m . The distinct number of parameter values equals q_m , because of the unknown variance term. For example, in case $k = 5$ and the hypothesis is $H_1 : \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$, there are 4 distinct mean values, namely μ_5 and μ_3 are a distinct value and μ_1, μ_4 and μ_2 are each a distinct value, and an unknown variance term. Thus, $q_1 = 5$. Note that the number of observations of the four distinct mean values are: $\tilde{n}_1 = n_3 + n_5$, $\tilde{n}_2 = n_1$, $\tilde{n}_3 = n_4$ and $\tilde{n}_4 = n_2$. The computation of the level probabilities (corresponding to the restrictions) can be done via simulation (Silvapulle and Sen 2005, pp. 78–81), consisting of 5 steps (where for convenience the subscript m is left out):

1. Generate \mathbf{Z} (of dimension $q - 1$) from $\mathcal{N}(0, V)$, where $q - 1$ equals the number of distinct μ_i values in the hypothesis. It holds that $V = \text{diag}\{1/\tilde{n}_1, \dots, 1/\tilde{n}_{q-1}\}$, where \tilde{n}_l is the number of observations of group l ($l = 1, \dots, q - 1$).
2. Compute $\tilde{\mathbf{Z}}$ via (3), that is, $\tilde{\mathbf{Z}} = \text{argmin}_{\boldsymbol{\mu} \in H} (\mathbf{Z} - \boldsymbol{\mu})^T V^{-1} (\mathbf{Z} - \boldsymbol{\mu})$, where H is the order-restricted hypothesis.
3. Determine the number of distinct values in $\tilde{\mathbf{Z}}$, called levels, denote this by s .
4. Repeat the previous steps R_{PT} times. In the program, the default value of R_{PT} is $R_{PT} = 100,000$.
5. Estimate the level probability $LP_l(q - 1, V)$ by the proportion of times s is equal to l ($l = 1, \dots, q - 1$).

The penalty term can thus be seen as the expected number of distinct parameters, that is, the expected number of distinct mean values plus 1 (because of the unknown variance term).

2.3. Bayesian model selection

Klugkist *et al.* (2005) present the (Bayesian) encompassing prior approach for order-restricted hypotheses in ANOVA. The model selection criterion used is the Bayes factor (Kass and Raftery 1995; Chib 1995), which is the ratio of marginal likelihoods of two hypotheses, say H_m and $H_{m'}$:

$$BF_{mm'} = \frac{L(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}) p(\boldsymbol{\mu}, \sigma^2 | H_m) / p(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}, H_m)}{L(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}) p(\boldsymbol{\mu}, \sigma^2 | H_{m'}) / p(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}, H_{m'})}, \quad (5)$$

where $p(\boldsymbol{\mu}, \sigma^2 | H_m)$ and $p(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}, H_m)$ are the prior and posterior distribution of the model parameters, respectively, which will be elaborated upon later.

In the encompassing prior approach, a prior $p(\boldsymbol{\mu}, \sigma^2 | H_u)$ is specified for the unconstrained hypothesis $H_u : \mu_1, \dots, \mu_k$. The prior distribution of any hypothesis H_m nested in H_u follows from the encompassing prior, using:

$$p(\boldsymbol{\mu}, \sigma^2 | H_m) = p(\boldsymbol{\mu}, \sigma^2 | H_u) \left(\frac{I_{\boldsymbol{\mu} \in H_m}}{\int p(\boldsymbol{\mu}, \sigma^2 | H_u) I_{\boldsymbol{\mu} \in H_m} d\boldsymbol{\mu} d\sigma^2} \right), \quad (6)$$

where the indicator function $I_{\boldsymbol{\mu} \in H_m}$ has the value one if the argument is true, that is, if the parameter values are in accordance with the constraints imposed by H_m , and zero otherwise.

The encompassing prior is specified as follows (Klugkist *et al.* 2005):

- All model parameters are a priori considered to be independent, that is,

$$p(\boldsymbol{\mu}, \sigma^2) = p(\mu_1) \times \dots \times p(\mu_k) \times p(\sigma^2).$$

- The prior distributions for all means are equal, that is,

$$p(\mu_1) = \dots = p(\mu_k).$$

- As will be shown in the sequel, for each parameter a relatively uninformative, conjugate prior will be specified, that is, $p(\mu_i) \sim \mathcal{N}(\mu_0; \tau_0^2)$ for $i = 1, \dots, k$, and $p(\sigma^2) \sim \text{Inv-}\chi^2(1; \sigma_0^2)$.

In sum, the encompassing prior for the ANOVA model is:

$$p(\boldsymbol{\mu}, \sigma^2 | H_u) = \prod_{i=1}^k \mathcal{N}(\mu_0; \tau_0^2) \times \text{Inv-}\chi^2(1; \sigma_0^2). \quad (7)$$

Combination of (6) and (7) gives the prior distribution (up to proportionality) of any order-restricted hypothesis H_m :

$$p(\boldsymbol{\mu}, \sigma^2 | H_m) \propto \prod_{i=1}^k \mathcal{N}(\mu_0; \tau_0^2) I_{\boldsymbol{\mu} \in H_m} \times \text{Inv-}\chi^2(1; \sigma_0^2). \quad (8)$$

In a similar way as in (6), the posterior of any hypothesis H_m is

$$p(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}, H_m) = p(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}, H_u) \left(\frac{I_{\boldsymbol{\mu} \in H_m}}{\int p(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}, H_u) I_{\boldsymbol{\mu} \in H_m} d\boldsymbol{\mu} d\sigma^2} \right). \quad (9)$$

The posterior distribution is proportional to the density of the data times the prior distribution, that is

$$p(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}, H_m) \propto L(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}) \times \prod_{i=1}^k \mathcal{N}(\mu_0; \tau_0^2) I_{\boldsymbol{\mu} \in H_m} \times \text{Inv-}\chi^2(1; \sigma_0^2). \quad (10)$$

The encompassing posterior $p(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}, H_u)$ is (10) where the indicator function always equals one.

Klugkist *et al.* (2005) have shown that (5) when using the prior in (6) and subsequent posterior in (9) leads to a simple form for the Bayes factor for a nested hypothesis H_m with the unconstrained hypothesis H_u :

$$BF_{mu} = \frac{c_m}{d_m}, \quad (11)$$

where c_m and d_m are the last terms (between the large brackets) of (6) and (9), respectively. The inverse of these constants, that is, c_m^{-1} and d_m^{-1} are the proportions of the encompassing prior and posterior, respectively, in agreement with the constraints of hypothesis H_m . Estimation of these proportions is straightforward using sampling. This means that in the context of order-restricted ANOVA, Bayes factors can be obtained without the – often burdensome – estimation of marginal likelihoods.

Specification of the encompassing prior

To complete the specification of the prior distribution, values must be assigned to the hyper-parameters μ_0 , τ_0^2 , and σ_0^2 . Klugkist and Hoijtink (2007) showed that Bayes factors for hypotheses formulated using inequality constraints among parameters are insensitive to the exact specification, as long as the encompassing prior is relatively vague. However, the results for hypotheses containing equality constraints are sensitive to the choice of τ_0^2 . Although we want relatively uninformative priors, that is, a large τ_0^2 , too large values result in Bartlett's or Lindley's paradox (cf. Lindley 1957; Bernardo and Smith 1994). Hence, to get reasonable values for the hyper-parameters, the prior specification is data-based. A Gibbs sample (Smith and Roberts 1993) is drawn from the unconstrained posterior $p(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}, H_u) = L(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}) \times p(\boldsymbol{\mu}, \sigma^2 | H_u)$, where $p(\boldsymbol{\mu}, \sigma^2 | H_u) \propto 1$. Summaries of the posterior sample provide values for the hyper-parameters according to the following choices:

- For σ_0^2 , the posterior mean of σ^2 is used. This provides a value that is reasonable for the data at hand and, with 1 degree of freedom, a posterior that is hardly affected by the prior.
- To obtain μ_0 and τ_0^2 , the information about each of the μ 's in the posterior sample is combined as follows: Based on the posterior sample, a credibility interval for each μ_i ($i = 1, \dots, k$) is determined by: $\bar{\mu}_i \pm pv \cdot s_{\mu_i}$, where $\bar{\mu}_i$ and s_{μ_i} are the mean and standard deviation of the sampled values for μ_i , respectively, and pv stands for *prior vagueness*. With the pv value it is specified which interval is used (e.g., $pv = 2$ provides the 95% credibility interval) and allows the user a choice for the amount of vagueness in the encompassing prior. Subsequently, the smallest lower bound (lb) and the largest upper bound (ub) of the k intervals define one broad interval containing all reasonable values for each of the μ 's. From this interval, $\mu_0 = (lb + ub)/2$ and $\tau_0 = (ub - lb)/2$ are specified.

For hypotheses containing equality constraints, the value specified for pv will affect the resulting Bayes factors. Larger pv values provide more support for hypotheses containing equality constraints (i.e., Lindley's paradox). Specification of the pv value by the user also provides the option to investigate prior sensitivity by running the program several times with different values. This is recommended if one or more of the hypotheses contain equality constraints among the means.

Estimation of the Bayes factor

The computation of Bayes factors of order-restricted hypotheses versus the unconstrained hypothesis is straightforward by taking a sample from the unconstrained prior (7) and a sample from the unconstrained posterior, that is, (10) with $I_{\boldsymbol{\mu} \in H_m} = 1$ for all $\boldsymbol{\mu}$. Samples are obtained by application of the Gibbs sampler (Smith and Roberts 1993). The proportions of prior and posterior iterations in agreement with the order-restricted hypotheses provide estimates for c_m^{-1} and d_m^{-1} , respectively. However, for a hypothesis containing at least one strict equality (e.g., $\mu_1 = \mu_2$), the direct application of this approach would result in the problematic outcome $d_m^{-1} = 0$ and $c_m^{-1} = 0$. Therefore, the program evaluates ‘about equality’ constraints instead, that is, $|\mu_1 - \mu_2| < \delta$ for a positive small δ . Two options are provided in the software: Researchers have the opportunity to investigate ‘relevant differences’ between means by specifying a non-zero δ , or strict equality constrained hypotheses can be evaluated, in which case δ approaches zero in a stepwise method. The latter requires an extension of the basic approach and includes constrained sampling. This will be elaborated in the next section.

For hypotheses that only require unconstrained sampling, estimation of Bayes factors in the software is based on a minimum of R_{BMS} iterations from both prior and posterior. The default value of R_{BMS} is $R_{BMS} = 500,000$. The posterior sample is taken after discarding 1,000 iterations that serve as burn-in. Note that, sampling from the unconstrained prior does not require a burn-in period, since all parameters are a priori independent.

For highly constrained hypotheses, the default value of $R_{BMS} = 500,000$ iterations may be insufficient to get stable estimates of d_m and c_m . Therefore, some additional rules are incorporated in the program, when the default setting is chosen: For more than 6 groups, the number of iterations from prior and posterior are doubled; for more than 10 groups, the number of iterations from prior and posterior are set at 5 million. Furthermore, another additional rule is incorporated (whether the default setting is used or not): if a minimum of 100 prior ‘hits’ (iterations in agreement with the constraints) is not reached for each hypothesis in the set, more iterations are added until this is the case. For an elaboration on the investigation of the stability and Monte Carlo errors of Bayes factors computed via (11), see Klugkist and Hoijtink (2007).

Stepwise estimation for small δ

For small values of δ , the estimation as just described would be rather inefficient. Furthermore, for $\delta = 0$ it would give the result $c_m^{-1} = d_m^{-1} = 0$. Therefore, a procedure is applied where the unconstrained samples are evaluated with a not too small initial δ value, denoted δ_0 , followed by a procedure that decreases δ_0 in a stepwise way, using $\delta_r = \delta_{r-1}/3$, for steps $r = 1, \dots, R$ (see also Klugkist 2008).

The stepwise procedure is based on the following product rule for the Bayes factor of H_m with the unconstrained H_u :

$$\begin{aligned} BF_{mu} &\approx BF_{m_{\delta_0}u} \times BF_{m_{\delta_1}m_{\delta_0}} \times \dots \times BF_{m_{\delta_R}m_{\delta_{R-1}}} \\ &= \frac{c_{m_{\delta_0}u}}{d_{m_{\delta_0}u}} \times \frac{c_{m_{\delta_1}m_{\delta_0}}}{d_{m_{\delta_1}m_{\delta_0}}} \times \dots \times \frac{c_{m_{\delta_R}m_{\delta_{R-1}}}}{d_{m_{\delta_R}m_{\delta_{R-1}}}}. \end{aligned} \quad (12)$$

Consider the case where H_m includes both inequality and strict equality ($\delta = 0$) constraints. The notation m_{δ_r} is used to denote the constraints of H_m , where the desired value $\delta = 0$ is replaced by a larger value δ_r ($r = 0, \dots, R$).

The first Bayes factor in (12), $BF_{m_{\delta_0}u} = \frac{c_{m_{\delta_0}u}}{d_{m_{\delta_0}u}}$, requires sampling from the unconstrained prior and posterior and counting the number of iterations in agreement with all constraints in m_{δ_0} , that is, the order restrictions as well as the equalities evaluated with δ_0 . The second and subsequent steps require constrained sampling, that is, sampling from (8) and (10). Consider, for instance, $BF_{m_{\delta_1}m_{\delta_0}} = \frac{c_{m_{\delta_1}m_{\delta_0}}}{d_{m_{\delta_1}m_{\delta_0}}}$, where $(c_{m_{\delta_1}m_{\delta_0}})^{-1}$ denotes the proportion of iterations, sampled from the prior of H_m with δ replaced by δ_0 , that are in agreement with H_m with δ replaced by δ_1 ($\delta_1 < \delta_0$). Similarly, $d_{m_{\delta_1}m_{\delta_0}}$ is based on a sample from the posterior constrained to the area that complies with the constraints of m_{δ_0} . Constrained sampling is also done by application of the Gibbs sampler, with inverse probability sampling to obtain samples in agreement with the constraints (see Gelfand, Smith, and Lee 1992).

In each step, H_m is evaluated with a smaller δ_r value. The stepwise Bayes factors represent change in the estimated BF_{mu} , as a consequence of the decrease in δ_r . At a certain point, a further decrease of the value for δ_r no longer changes the Bayes factor, that is, for large enough R , $BF_{m_{\delta_R}m_{\delta_{R-1}}} \rightarrow 1$ (Berger and Delempady 1987). This implies that a good approximation of BF_{mu} with *exact* equalities is obtained.

To obtain an efficient estimation procedure, it is important to start with a large enough δ_0 . In the software, the starting value is prior-based and equals $\tau_0/2$ (τ_0 for more than 8 groups) unless this value is smaller than the user-specified δ in which case δ is evaluated directly. Sampling from the unconstrained prior and posterior (first step) is as explained in the previous section. In each subsequent step ($r = 1, \dots, R$), samples are drawn from the constrained priors and posteriors after a burn-in of 100 iterations. The number of iterations from each constrained prior is minimally 500,000. If necessary, up to 1 million iterations are done until the number of hits reaches 500. If after 1 million iterations the number of hits is below 100, more samples are drawn until 100 hits are obtained. From each constrained posterior, samples are drawn until 500 hits are obtained, with a maximum of 1 million iterations.

The number of steps, R , is determined by one of two stopping rules: convergence of the estimate of the final Bayes factor is assumed if two subsequent $BF_{m_{\delta_r}m_{\delta_{r-1}}}$ values deviate less than 0.05 from 1. When $\delta \neq 0$, the last step of the procedure is performed as soon as $\delta_r \leq \delta$ (if $\delta_r < \delta$, in the final step it is set at δ).

Interpretation of the results

The software estimates Bayes factors for each order-restricted hypothesis with the unconstrained hypothesis using (11) or (12). The Bayes factor for the comparison of two order-restricted hypotheses, say H_m and $H_{m'}$ can be computed, using:

$$BF_{mm'} = \frac{BF_{mu}}{BF_{m'u}}.$$

A Bayes factor provides the amount of support of one hypothesis compared to another. If, for instance $BF_{mm'} = 6$, the support for H_m is 6 times as large as for $H_{m'}$. Likewise, $BF_{mm'} = 0.5$ shows that the support for H_m is 2 times as small as the support for $H_{m'}$.

Furthermore, the software provides posterior model probabilities (*pmp*), representing the relative support for each hypothesis in a finite set of hypotheses (\mathcal{M}). To obtain *pmp*'s from

Bayes factors, prior model probabilities must be specified, representing the degree of belief in each hypothesis before observing the data. A usual – objective – choice are equal prior probabilities for all hypotheses, that is, $p(H_m) = 1/M$, for $m \in \mathcal{M}$, where M denotes the total number of hypotheses. This prior specification, which is also adopted in the software, leads to the following equation for $pmp(H_m)$:

$$pmp(H_m) = \frac{BF_{mu}}{\sum_{m \in \mathcal{M}} BF_{mu}}.$$

Posterior model probabilities can be computed including or excluding the unconstrained hypothesis. As an example, consider H_0 , H_1 and H_2 from (2). In the software, the user can specify the presence of just these 3 hypotheses (i.e., $M = 3$) and the pmp 's are computed excluding the unconstrained hypothesis:

$$pmp(H_m) = BF_{mu} / (BF_{0u} + BF_{1u} + BF_{2u}).$$

Alternatively, one can also explicitly add the unconstrained hypothesis as a hypothesis of interest ($M = 4$). The resulting pmp values, denoting the unconstrained hypothesis by H_3 , are:

$$pmp(H_m) = BF_{mu} / (BF_{0u} + BF_{1u} + BF_{2u} + BF_{3u}),$$

for $m \in \mathcal{M}$ and with $BF_{3u} = 1$.

3. Example based on Lucas (2003)

The three approaches for confirmatory ANOVA will be illustrated using data from the research of Lucas (2003). In this study, the interest lies in the amount of influence a leader has on his/her group members. The experiment contained five experimental groups: (1) a group with a randomly selected male leader, (2) a group with a randomly selected female leader, (3) a group with a male leader selected on ability, (4) a group with a female leader selected on ability, and (5) a group with a female leader selected on ability after institutionalization of female leadership. The institutionalization is done by showing the participants a film in which it is normal to have female leadership and females do well as leaders. The resulting group means and standard deviations of the influence of the leader are shown in Table 1.

The research question of Lucas (2003) is: “Can institutionalization of female leadership reduce the influence gap between woman and men by legitimating structures of female leadership?” The expectations of Lucas (2003) are in short:

Group	Mean influence	SD	n
1	2.33	1.86	30
2	1.33	1.15	30
3	3.20	1.79	30
4	2.23	1.45	30
5	3.23	1.50	30

Table 1: Group means and standard deviations (SD) of influence (Lucas 2003).

- Male leaders (group 1 and 3) have higher influence over participants than female leaders (group 2 and 4, respectively), *ceteris paribus*.
- Leaders appointed on ability (group 3 and 4) have higher influence over participants than leaders appointed randomly (group 1 and 2, respectively), *ceteris paribus*.
- Institutionalized female leaders selected on ability (group 5) have higher influence over participants than ‘normal’ female leaders selected on ability (group 4), or than randomly selected female leaders (group 1).
- Institutionalized female leaders selected on ability (group 5) have (almost) the same influence over participants as male leaders appointed on ability (group 3).

These expectations can be represented by the hypothesis $H_1 : \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$, where μ_i represents the mean influence of the leader in group i .

Another hypothesis of interest can be:

- Leaders chosen on basis of ability score higher than leaders selected at random (so, group 3 scores higher than group 1 and group 4 scores higher than group 2).
- Male leaders selected at random (group 1) have an higher influence than female leaders selected on competence (group 4).
- There is no difference in influence of female leaders selected on competence in case of institutionalization (group 5) or in the ‘normal’ case (group 4).

This can be represented by $H_2 : \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2$.

In addition, the traditional null and alternative hypothesis, H_0 and H_3 from (2), respectively, will be used to illustrate the \bar{F} test, the order-restricted information criterion and Bayesian model selection.

3.1. Results using the \bar{F} test

Using the \bar{F} test for the evaluation of the four hypotheses specified in the Lucas’ example (see (2)), five tests are performed (Table 2). First, H_0 is tested against H_3 . This test results in a p value smaller than 0.001, rejecting H_0 . Second, two tests are done with respect to H_1 : H_0 is tested against the order-restricted H_1 and H_1 is tested against the unconstrained hypothesis H_3 . These tests result in a p value smaller than 0.001 and a p value of 0.995,

Hypotheses tested			\bar{F}	p value
H_0	against	H_3	30.27	< 0.001
H_0	against	H_1	30.26	< 0.001
H_1	against	H_3	0.01	0.995
H_0	against	H_2	22.91	< 0.001
H_2	against	H_3	7.36	0.070

Table 2: The \bar{F} tests of the four specified hypotheses.

Hypothesis	$\log L_m$	Penalty	ORIC
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$	-292.27	2.00	588.54
$H_1 : \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$	-278.05	3.19	562.49
$H_2 : \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2$	-281.76	3.14	569.79
$H_3 : \mu_1, \mu_2, \mu_3, \mu_4, \mu_5$	-278.05	6.00	568.10

Table 3: The ORIC values of the four specified hypotheses.

respectively, both favoring H_1 . Third, two tests are done with respect to H_2 : H_0 is tested against H_2 and H_2 is tested against H_3 . The resulting p values are a p value smaller than 0.001 and a p value of 0.07, respectively, both favoring H_2 . So, both H_1 and H_2 are preferred. Since no direct comparison between order-restricted hypotheses is possible with the \bar{F} test, nothing can be concluded with respect to an overall preferred hypothesis

3.2. Results using ORIC

The ORIC consists of a fit/likelihood part and a complexity/penalty part. In Table 3 the likelihood, penalty term, and ORIC values are given for H_0 , H_1 , H_2 and H_3 . The penalty for H_0 equals 2, because there are two distinct parameters: all means are equal, which represents one distinct parameter, and the parameter σ^2 . Analogously, the penalty for H_3 is 6: there are five distinct means (since there are no restrictions among the means) and one variance parameter. In hypotheses H_1 and H_2 there are also five means, but these hypotheses contain inequality constraints, as opposed to H_0 and H_3 . The penalties of these order-restricted hypotheses are calculated as explained earlier. The values of the penalties are respectively 3.20 and 3.13. The hypothesis with the smallest value for the ORIC is the preferred hypothesis. Thus, in the example, H_1 is the preferred hypothesis according to the ORIC.

3.3. Results using BMS

The results using BMS are presented in Table 4. For each hypothesis, the Bayes factor comparing that hypothesis with the unconstrained hypothesis (H_3) is presented for three different values of pv . The equality constraints are evaluated as strict equalities, that is, $\delta = 0$ is approximated using the stepwise approach explained in Section 2.3.

As was explained in Section 2.3, the encompassing prior is specified to be low informative and is based on the observed data. The resulting encompassing prior for the Lucas data and

Model	$pv = 1$		$pv = 2$		$pv = 3$	
	BF	pmp	BF	pmp	BF	pmp
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$	0.0	0.00	0.0	0.00	0.0	0.00
$H_1 : \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$	57.8	0.96	67.9	0.96	80.6	0.97
$H_2 : \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2$	1.4	0.02	1.5	0.02	1.8	0.02
$H_3 : \mu_1, \mu_2, \mu_3, \mu_4, \mu_5$	1.0	0.02	1.0	0.01	1.0	0.01

Table 4: Bayes factors (BF) and posterior model probabilities (pmp) for different prior specifications.

$pv = 1$ is:

$$p(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \sigma^2 | H_3) = \prod_{i=1}^5 \mathcal{N}(2.28; 1.53) \times \text{Inv-}\chi^2(1; 2.50).$$

The priors using $pv = 2$ and $pv = 3$ differ only with respect to the variance of the normal distributions, with values 2.32 and 3.27, respectively.

Irrespective of the choice made for the prior, H_1 is clearly the most supported order-restricted hypothesis. The corresponding posterior probabilities (using equal prior model probabilities) for the four hypotheses are 0.00, 0.97, 0.02 and 0.01, respectively (for each pv value). These results again lead to the conclusion that H_1 is the best of the four models/hypotheses considered, as was also concluded when using the ORIC. Comparison of the three prior specifications shows that, for the hypotheses at hand, the results are not very sensitive to the specification of the prior.

Acknowledgments

This research is financed by the Dutch organization for scientific research NWO-VICI-453-05-002.

References

- Anraku K (1999). “An Information Criterion for Parameters under a Simple Order Restriction.” *Biometrika*, **86**, 141–152.
- Berger JO, Delempady M (1987). “Testing Precise Hypotheses.” *Statistical Science*, **2**, 317–352.
- Bernardo JM, Smith AFM (1994). *Bayesian Theory*. John Wiley & Sons, Hoboken, NJ.
- Chib S (1995). “Marginal Likelihood from the Gibbs Output.” *Journal of the American Statistical Association*, **90**, 1313–1321.
- Gelfand AE, Smith AFM, Lee T (1992). “Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling.” *Journal of the American Statistical Association*, **87**, 523–532.
- Kass RE, Raftery AE (1995). “Bayes Factors.” *Journal of the American Statistical Association*, **90**, 773–795.
- Klugkist I (2008). *Bayesian Evaluation of Informative Hypotheses*, chapter Encompassing Prior Based Model Selection for Inequality Constrained Analysis of Variance. Springer-Verlag, New York.
- Klugkist I, Hoijtink H (2007). “The Bayes Factor for Inequality and About Equality Constrained Models.” *Computational Statistics & Data Analysis*, **51**, 6367–6379.
- Klugkist I, Laudy O, Hoijtink H (2005). “Inequality Constrained Analysis of Variance: A Bayesian Approach.” *Psychological Methods*, **10**, 477–493.

- Kuiper RM, Hoijtink H (2010). “Comparisons of Means Using Exploratory and Confirmatory Approaches.” *Psychological Methods*, **15**, 69–86.
- Lindley DV (1957). “A Statistical Paradox.” *Biometrika*, **44**, 187–192.
- Lucas JW (2003). “Status Processes and the Institutionalization of Women as Leaders.” *American Sociological Review*, **68**, 464–480.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Development Core Team (2009). “nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-92.” URL <http://CRAN.R-project.org/package=nlme>.
- Silvapulle MJ, Sen PK (2005). *Constrained Statistical Inference*. John Wiley & Sons, Hoboken, NJ.
- Smith AFM, Roberts GO (1993). “Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods.” *Journal of the Royal Statistical Society B*, **55**, 3–24.
- SPSS Inc (2006). *SPSS for Windows, Release 15*. SPSS Inc., Chicago, IL. URL <http://www.spss.com/>.
- Visual Numerics (2003). *IMSL Fortran Library User’s Guide MATH/LIBRARY Volume 2 of 2: Mathematical Functions in Fortran*.

A. ConfirmatoryANOVA.exe user manual

This user manual will describe and illustrate the options available in `ConfirmatoryANOVA.exe` (published along with this manuscript and also available at <http://www.fss.uu.nl/ms/Kuiper/>). This program is made in Fortran 90 using the Intel Visual Fortran Compiler 9.1 for Windows. This compiler uses **IMSL** 5.0.

To make the program more user-friendly, an interface is made with use of C# in Microsoft Visual Studio 2005 (Appendix B). The interface calls the exe file made in Fortran 90 and it makes the appropriate input file needed for the Fortran 90 program. The input for the Fortran 90 program is described in this appendix. Those interested in only the interface are referred to Appendix B.

`ConfirmatoryANOVA.exe` is free of use. However, when results obtained with this program are published, please refer to this paper. In the program the following methods can be performed:

- the \bar{F} test,
- the order-restricted information criterion (ORIC),
- Bayesian model selection (BMS).

A.1. Modification input files

No matter what analysis should be performed two text files have to be modified (such that they apply to your data), namely `Input.txt` and `Data.txt`.

It should be noted that:

- The names of the text files 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 and do not delete lines containing labels.
- The data in `Data.txt` should be complete, that is, missing data are not allowed.

First half of Input.txt

First of all, you must denote which analyses should be performed. This has to be done in `Input.txt`. A certain analysis will be performed if in the line below the name of that analysis a 1 is filled in. It will not be performed, when a 0 is filled in.

When the ORIC and the \bar{F} test should be performed and BMS should not, the first half of `Input.txt` should look as follows (when using the default values for the seed value and number of iterations):

```
Seed value and number of iterations (>0) for Fbar test, ORIC, and BMS
123 100000 100000 500000
Perform F bar test, ORIC, BMS (1 = yes, 0 = no)
1 1 0
```

We will come back to the seed value and number of iterations in Section A.5.

Data.txt

Second of all, group membership and the corresponding data must be given in `Data.txt`, where in the first column the group numbers must be given and in the second column the corresponding data point y_{ij} . The order of the group numbers and data points does not matter as long as the group number corresponds to the data point in the same row. The following are three examples of how `Data.txt` could look:

1	3.58	1	3.58	3	1.39
1	-0.15	2	1.67	2	1.85
...	...	3	1.39	5	4.58
2	1.67	4	1.57	5	1.38
2	1.85	5	1.38	3	4.53
...	...	1	-0.15	4	2.97
3	1.39	2	1.85	1	-0.15
3	4.53	3	4.53	2	1.67
...	...	4	2.97	4	1.57
4	1.57	5	4.58	1	3.58
4	2.97
...
5	1.38
5	4.58
...

The group numbers do not need to be sequential. For example, if you have a **SPSS** or **Excel** file with data for 10 groups and in the current analysis you want to compare only 5 groups (which are not the first five). In that case, you can just copy the appropriate data from the **SPSS** or **Excel** file to a text file without adjusting the group numbers. In the software, the group numbers will be made sequential and the output will be given for these adjusted sequential group numbers. For example, if you have data with group numbers 1, 4, 5, 6, and 8 (whether the data are in order or not), these will become group numbers 1, 2, 3, 4, and 5, respectively. Note that, in specifying the restrictions (see the next sections), you need to use the adjusted sequential group numbers.

From the data the number of groups and the number of observations per group are determined.

A.2. Basic elements of writing constraints

In performing an \bar{F} test, in determining the ORIC or in doing BMS, all the hypotheses of interest, like H_0 to H_3 in (2), must be given explicitly. Note that it is also possible to specify a set of hypotheses without the classical null $H_0 : \mu_1 = \dots = \mu_k$ and/or alternative $H_u : \mu_1, \dots, \mu_k$. However, we recommend to include the alternative H_u (when doing model selection), since it can be used to protect against choosing a weak hypothesis (Kuiper and Hoijtink 2010). Note that one should include H_0 only when there is real interest in H_0 .

When using the ORIC or doing BMS several models/hypotheses are compared to each other. In the \bar{F} test, the order-restricted hypotheses (like H_1 and H_2 in (2)) are tested against H_0

and H_u . If the classical null and/or the alternative are included in the set of hypotheses, the classical null will be tested against the classical alternative.

The basic elements, for writing down the hypotheses of interest are:

1. Representation of an equality sign (=)

Suppose the hypothesis of interest is $\mu_5 = \mu_3$, that is, $\mu_5 = \mu_3, \mu_1, \mu_2, \mu_4$. The ordering of the group numbers in this restriction is represented by: 5 3 1 2 4. The restriction is represented by: 1 1 0 0 0, where the 1s indicate that mean 5 and 3 belong to set 1 and are equal to each other, and where 0 indicates that the corresponding mean is unrestricted. N.B. in a restriction the first set is always labeled as 1, the second as 2 (and so on).

2. Representation of a greater than sign (>)

Suppose the hypothesis of interest is $\mu_1 > \mu_3, \mu_2, \mu_4, \mu_5$. The ordering of the group numbers in this restriction is represented by: 1 3 2 4 5. The restriction is represented by: 1 -3 0 0 0, where -3 means that mean 3 is smaller than mean 1. Thus, it represents $\mu_3 < \mu_1$, which is equal to $\mu_1 > \mu_3$. Here again 1 indicates that mean 1 belongs to set 1. Because of the inequality restriction between mean 1 and 3, mean 3 belongs to set 2 (the importance of this will be made clear in the next section).

3. Representation of a smaller than sign (<)

Suppose the hypothesis of interest is $\mu_1 < \mu_2, \mu_3, \mu_4, \mu_5$. The ordering of the group numbers in this restriction is represented by: 1 2 3 4 5. The restriction is represented by: 1 -1 0 0 0, where -1 means that mean 2 is greater than mean 1. Thus, it represents $\mu_2 > \mu_1$, which is equal to $\mu_1 < \mu_2$.

Every hypothesis can be represented by these basic elements in one or more restrictions. For example, $\mu_5 > \mu_3 < \mu_1, \mu_2 < \mu_4$, can be represented by the restrictions:

$$\mu_5 > \mu_3, \mu_1, \mu_2, \mu_4,$$

$$\mu_5, \mu_3 < \mu_1, \mu_2, \mu_4,$$

$$\mu_2 < \mu_4, \mu_5, \mu_3, \mu_1,$$

which can be represented by:

Ordering of means in restriction

5 3 1 2 4

5 3 1 2 4

2 4 5 3 1

(Order) Restrictions

1 -3 0 0 0

0 1 -1 0 0

1 -1 0 0 0

A.3. Combinations of basic elements

Often the hypothesis of interest can be represented in a smaller number of restrictions than when using only the basic elements. The following shortcuts can be used:

1. $\mu_5 = \mu_3 = \mu_1, \mu_2 = \mu_4$

Because of the equality constraints (“=”), means 5, 3 and 1 belong to set 1. Therefore, means 2 and 4 belong to set 2. Thus, this hypothesis can be represented by the following ordering of the group numbers and corresponding restriction:

```
Ordering of means in restriction
5 3 1 2 4
(Order) Restrictions
1 1 1 2 2
```

2. $\{\mu_5 = \mu_3 = \mu_1\} > \{\mu_2 = \mu_4\}$

Because of the equality constraints, mean 5, 3 and 1 belong to set 1, and means 2 and 4 belong to set 2. Because of the constraint between mean 1 and 2, mean 2 belongs implicitly to set 2. This hypothesis can be represented by the following ordering of the group numbers and corresponding restriction:

```
Ordering of means in restriction
5 3 1 2 4
(Order) Restrictions
1 1 1 -3 2
```

The 1s indicate the equality constraints between mean 5, 3 and 1, the -3 represents the inequality constraint “ $>$ ” between mean 1 and 2, and the 2 indicates the equality constraints between mean 2 and 4 (because mean 2 implicitly belongs to set 2). N.B. the restrictions $\mu_5 = \mu_1, \mu_5 > \mu_2, \mu_3 > \mu_2, \mu_5 > \mu_4, \mu_3 > \mu_4, \mu_1 > \mu_4$ do not have to stated explicitly, these will hold since it holds that $\mu_5 = \mu_3, \mu_3 = \mu_1, \mu_1 > \mu_2$ and $\mu_2 = \mu_4$.

3. $\mu_5 = \mu_3 > \mu_1 > \mu_2 = \mu_4$

This hypothesis can be represented by the following ordering of the group numbers and corresponding restriction:

```
Ordering of means in restriction
5 3 1 2 4
(Order) Restrictions
1 1 -3 -3 3
```

The 1s indicate the equality constraint between mean 5 and 3. The -3 s represents the inequality constraint “ $>$ ” between mean 3 and 1 and between 1 and 2. Note that mean 1 implicitly belongs to set 2 and mean 2 implicitly to set 3. Therefore, the equality constraint between mean 2 and 4 is represented by the 3, because mean 2 and 4 belong to set 3.

4. $\mu_5 = \mu_3 > \mu_1 > \mu_2 = \mu_4$

Likewise, this hypothesis can be represented by the following ordering of the group numbers and corresponding restriction:

```
Ordering of means in restriction
5 3 1 2 4
```

(Order) Restrictions

1 1 -3 -1 3

5. $\mu_5 = \mu_3 > \mu_1 > \mu_2, \mu_4$

This hypothesis can be represented by the following ordering of the group numbers and corresponding restriction:

Ordering of means in restriction

5 3 1 2 4

(Order) Restrictions

1 1 -3 -3 0

Note that, mean 4 is a free parameters, that is, means which are not restricted at all. The free parameters are denoted by a “0” and no group numbers are assigned (directly or indirectly). Here, as in the previous two examples, mean 2 belongs to group 3 and mean 4 belongs to another group. However, this is not group 4. Another example is given next.

6. $\mu_5 = \mu_3, \mu_4, \mu_1 > \mu_2$

This hypothesis can be represented by the following ordering of the group numbers and corresponding restriction:

Ordering of means in restriction

5 3 4 1 2

(Order) Restrictions

1 1 0 2 -3

The 1s indicate the equality constraint between mean 5 and 3. Note that, as mentioned in the previous example, no group number is assigned to free parameters. So, mean 4 belongs to another group than all the other means, but no group number is assigned. A 0 is filled in. Mean 1 belong to the next group, that is, group 2. The -3 represents the inequality constraint “ $>$ ” between mean 1 and 2. Note that mean 2 indirectly belongs to group 3.

7. $\mu_5 = \mu_3 > \mu_1 < \mu_2 = \mu_4$

This hypothesis can be represented by the following ordering of the group numbers and corresponding restriction:

Ordering of means in restriction

5 3 1 2 4

(Order) Restrictions

1 1 -3 -1 3

8. $\mu_5 = \mu_3 > \mu_1 = \mu_2 > \mu_4$

This hypothesis can be represented by the following ordering of the group numbers and corresponding restriction:

Ordering of means in restriction

5 3 1 2 4

(Order) Restrictions

```
1 1 -3 2 -3
```

The 1s indicate the equality constraints between mean 5 and 3, the -3 s represents the inequality constraint “ $>$ ” between mean 3 and 1 and between 2 and 4. Note that mean 1 implicitly belongs to set 2. Therefore, the equality constraint between mean 1 and 2 is represented by the 2.

In the program `ConfirmatoryANOVA.exe` error messages are built in to detect wrongly stated hypotheses. But sometimes wrongly stated hypotheses are not detected, because the stated hypothesis represents another existing hypothesis. When accidentally a 3 is given instead of a 2, another existing hypothesis is stated. Namely, the restriction 1 1 -3 3 -3 represents the hypothesis $\mu_5 = \mu_3 > \mu_1, \mu_2 > \mu_4$. So, care must be taken in writing down the hypothesis of interest.

9. $\mu_5 = \mu_3 > \mu_1, \mu_2 > \mu_4$

This hypothesis can be represented by the following ordering of the group numbers and corresponding restriction:

Ordering of means in restriction

```
5 3 1 2 4
```

(Order) Restrictions

```
1 1 -3 3 -3
```

A.4. Equalities and about equalities in BMS

In the \bar{F} -test, in the ORIC and in BMS for strict equalities ($\delta = 0$), the two hypotheses $\mu_1 = \mu_2 = \mu_3$ (specified by 1 restriction, with ordering of means 1 2 3, and (order) restrictions 1 1 1) and $\mu_1 = \mu_2, \mu_2 = \mu_3$ (specified by 2 restrictions, both with ordering of means 1 2 3, and (order) restrictions 1 1 0 and 0 1 1) are equivalent.

However, the BMS approach in the program also provides the option to specify about equality constraints ($\delta > 0$). In that case, the second hypothesis is evaluated using $|\mu_1 - \mu_2| < \delta$ and $|\mu_2 - \mu_3| < \delta$, whereas the first hypothesis adds a third constraint: $|\mu_1 - \mu_3| < \delta$. The results of the first and second hypothesis may differ and therefore careful consideration of the formulation of hypotheses is important.

A.5. Set the seed value and number of iterations

The calculation of the p value of the \bar{F} and the penalty of the ORIC (i.e., PT) and BMS are sampling based approaches. For example, when generating data from a normal distribution (in order to determine the p value of the \bar{F} or the penalty of the ORIC), a seed value is needed. When using the same seed value, the same data will be ‘sampled’. When looking at another seed value in a rerun of the same problem, one can also see how stable the results are. Thus, the p value of the \bar{F} , the penalty of the ORIC (i.e., PT), and the results of BMS can differ for various seed values.

In case a result is not stable, the number of iterations needs be set higher. In the \bar{F} test, the p value depends on the number of iterations R_p . When using the ORIC, the penalty is dependent on the number of iterations R_{PT} . When doing BMS, the Gibbs sampler is used,

which is based on a minimum of R_{BMS} iterations. These values can also be set in the input, namely in the second line of `Input.txt` (see Section A.1.1). The default values of the number of iterations in each method are: $R_p = 100,000$, $R_{PT} = 100,000$, and $R_{BMS} = 500,000$.

Note that the higher the number of iterations the higher the computing time. If one lowers the number of iterations (in order to lower the computing time), one must be aware that this probably affects the stability of the results. Furthermore, when the initial number of iterations for BMS (i.e., R_{BMS}) is lowered, the computing time is not necessarily decreased, because of the requirement of a minimum of 100 prior hits.

A.6. Error messages

In the program `ConfirmatoryANOVA.exe` error messages are built in to detect wrong stated hypotheses. However, it does not detect all wrongly stated hypotheses, since the stated hypothesis can represent another existing hypothesis, as is made clear in Section A.3.

It is also possible to state other input wrongly. For example, the wrong number of restrictions is given. When making a mistake, an informative warning will be given.

However, it is possible to make a mistake that we have not foreseen. In that case, check the input and compare it to the data. If you cannot solve the problem, send the input and data file to `r.m.kuiper@uu.nl`.

A.7. Modification of the second half of `Input.txt`

For all three methods (i.e., the \bar{F} test, the ORIC, and BMS), all hypotheses of interest must be given explicitly. In case BMS is performed, two additional specifications need to be made: The desired δ (for exact equalities specify $\delta = 0$, for an about equality any positive number can be specified) and the prior vagueness pv (default recommendation $pv = 2$; any positive number may be specified). This must be done in the second half of `Input.txt`. If the hypotheses of interest are the set of hypotheses of Lucas stated in (2), `Input.txt` has the format shown in Table 5 (with comments in angle brackets).

A.8. Save and close

When you have modified `Input.txt` (such that it applies to your data), you should save and close it.

A.9. Run `ConfirmatoryANOVA.exe`

When `ConfirmatoryANOVA.exe` is runned, the output file `Output.txt` will be created in the folder you are working in.

`Output.txt`

`Output.txt` gives the results of the requested analyses. In Section B.5, the output is given when using the interface for the Lucas example described in this paper (provided that all three analyses, that is, the \bar{F} test, the ORIC, and BMS, are performed).

In case the interface is not used, the output will be a bit different, namely in two ways:

1) The hypotheses of interest will be displayed in the way they are filled in `Input.txt`. When

```

Number of models to be compared
<Fill in the number of models / hypotheses you want to compare; e.g.,>
4
Number of restrictions per model
<Fill in, for every model / hypothesis, the number of restrictions that represent that
model / hypothesis; e.g.,>
1
2
1
1
Ordering of means in restriction
<Fill in the ordering of the means / group numbers for each restriction for every model /
hypothesis. The orderings per restrictions are separated by an "enter". The ordering
consists of the numbers 1 to "the total number of groups". For more details see
"Basic Elements of Writing Constraints" and "Combinations of Basic Elements"; e.g.,>
1 2 3 4 5
5 3 1 2 4
3 4 2 1 5
3 1 4 5 2
1 2 3 4 5
(Order) Restrictions
<Fill in the restrictions. This must be done in a certain manner, which is explained in
"Basic Elements of Writing Constraints" and "Combinations of Basic Elements";
e.g.,>
1 1 1 1 1
1 1 -3 -3 0
1 -3 -3 0 0
1 -3 -3 3 -3
0 0 0 0 0
When BMS is performed, an interval for equality relations (delta) is needed
and a parameter for prior vagueness (pv)
<Fill in  $\delta \geq 0$  and  $pv > 0$ ; e.g.,>
0.0 2.0

```

Table 5: Example Input.txt file with comments in angle brackets.

using the interface, the hypotheses of interest are stated in terms of " μ_i ", ">", "<", "=", and ",".

2) The numbering of the hypotheses is different.

When using the interface and when H_0 is included in the set of hypotheses, H_0 will become Hypotheses 1 in the output. When H_u is included in the set of hypotheses, H_u will become Hypotheses 2, when H_0 is also included, and Hypothesis 1, when H_0 is not included. The other hypotheses will also be adjusted to the appropriate hypothesis number. So, an order-restricted hypotheses H_1 will become Hypotheses 3 (when both H_0 and H_u are specified) or Hypotheses 2 (when only H_0 or H_u is specified); et cetera.

B. User manual of ConfirmatoryANOVA.exe with interface

B.1. Read, write or copy data

First of all, the data should be entered. Press on the **Data** button in the **ConfirmatoryANOVA.exe** form (see Figure 1) to go the **DataInput** form (see Figure 2). The data can be entered manually, by copying it from a file (say an **SPSS** or **Excel** file) or by reading it from a text (i.e., **.txt**) file. See also Section A.1.2 for a description of the data format.

When the data are read from a file or, otherwise, after clicking on the **OK** button, the data are validated. In case of invalid data (e.g., in case no number is entered or in case a row has only a group number or has only a data point), the corresponding lines are made red. The invalid data should be corrected (by adjusting the data in the textbox/field or by adjusting the **.txt** file and rereading the adjusted file) or deleted (e.g., when all lines should be deleted, by clicking on the **Clear Invalid Data** button). After the adjustments, press the **OK** button. The data will be validated again. Note that, in case of deleting data, the number of observations per group (which is shown in the **ConfirmatoryANOVA.exe** form) will be adjusted automatically. In case the data are valid, you return to the **ConfirmatoryANOVA.exe** form (see Figure 1).

The screenshot shows the main interface of the Confirmatory ANOVA software. The window has a blue title bar with the text "Confirmatory ANOVA". The interface is organized into several functional areas:

- Data Input Section:** Includes a "Data" button, a "Reset" button, and a "Clear All" button. There are also checkboxes for "Fill in data", "Number of groups (k > 1)", and "Number of observations in the groups".
- Help and Information Section:** Contains a "Tutorial" link, a "Note that a pdf viewer is needed" message, and an "About" section with version information (Version 0.1 Beta) and contact details for Rebecca Kuiper.
- Method Selection Section:** Titled "Select the methods you want to perform", it includes checkboxes for "Confirmation", "Fbar", "ORIC", and "BMS".
- Model Specification Section:** Titled "Specify models for confirmatory methods", it has checkboxes for "H0: $\mu_1 = \dots = \mu_k$ ", "Ha: μ_1, \dots, μ_k ", and "Add (order-restricted) hypothesis/-es".
- Execution Section:** Features a "Run" button, a "Settings" button, and a field for "Name of output file:" with "Output" and ".txt" entered.
- Progress Report:** A large empty box on the right side of the interface.

Figure 1: ConfirmatoryANOVA.exe form.

Figure 2: DataInput form.

B.2. Specify methods

From the data (and group membership), the number of groups and number of observations per group are determined. Then, you must denote which analyses should be performed. A certain analysis will be performed if the corresponding checkbox is checked.

In case BMS is performed, two additional specifications need to be made (in a popup-panel): The desired δ and the prior vagueness pv . It holds that $\delta \geq 0$ and $pv > 0$. When specifying exact equality restrictions in BMS, δ must be set to $\delta = 0$. When specifying about equality restrictions in BMS, δ must be set to $\delta > 0$. In the latter case, one should carefully specify the restrictions (see Section A.4). The default recommendation of pv is $pv = 2$.

For all three methods, the hypotheses of interest must be given explicitly (in the then appearing panel). Specifying the order-restricted hypotheses will be explained in the next section. When you do not want to specify any order-restricted hypothesis, you should uncheck the **Add (order-restricted) hypothesis/-es** checkbox (see Figure 1). One should also specify whether one wants to evaluate the classical null hypothesis (i.e., $H_0 : \mu_1 = \dots = \mu_k$) and the classical alternative (i.e., $H_a : \mu_1, \dots, \mu_k$), here also called the unconstrained hypothesis (see Figure 1).

As discussed in Section A.5, the values for the seed value and the number of iterations can be specified. This is done in the **Settings** form (see Figure 3) appearing when pressing the **Settings** button.

More details on specifying the restrictions are given next.

Settings

Fill in the settings. [Tutorial](#) Note that a pdf viewer is needed.
 In case no value is inserted or Cancel-button is pressed, the default value(s) will be used.

Seed value

Number of iterations for

- the penalty of the ORIC

- the p-value of the Fbar test

- Gibbs sampler in BMS

Figure 3: Settings form.

B.3. Specifying the order-restricted hypotheses

An (order-restricted) hypothesis, say H_1 , can be specified in the panel appearing when clicking on the **Add** button in the **Edit Hypothesis 1** panel (see Figure 4). Then fill in the hypothesis, that is, fill in the group numbers and the corresponding constraints between the means of these group. Note that a hypothesis can consist of multiple ‘restrictions’. To add another restriction in, say H_1 , you must press the **Add** button in the **Edit restriction(s) in H1** panel (see Figure 4).

As mentioned before, the numbers in a certain restriction always consists of the numbers 1 to “the number of groups” (in the example, 5) and each number is used precisely once. It should be noted that in case of 5 means and you want to evaluate $H_1 : \mu_3 > \mu_1 > \mu_4$, you should fill in $H_1 : \mu_3 > \mu_1 > \mu_4, \mu_2, \mu_5$ or, better “3 > 1 > 4, 2, 5”. If the entry is a number greater than “the number of groups” or the entry is not an integer, then the corresponding textbox will be made red and an error message is given in the **ConfirmatoryANOVA.exe** form. The check on using every number only once is done after pressing the **Run** button. In that case, an error message will be given in the **Progress Report** panel and in **Error.txt** (see also the next subsection).

One can also specify the classical H_0 and/or the classical H_a as a hypothesis of interest. Note that one should include H_0 only when there is real interest in H_0 (Kuiper and Hoijtink 2010). We recommend to include H_a (when doing model selection) as a safeguard for choosing a weak hypotheses (Kuiper and Hoijtink 2010). H_0 and H_a do not need to be specified explicitly, one can just check the corresponding checkboxes (see Figure 1 and Figure 4).

B.4. Error messages

Before pressing the **Run** button, one should check whether the requirements are met. When

The screenshot shows the 'Confirmatory ANOVA' software window. The interface is divided into several sections:

- Input Section:** Includes checkboxes for 'Fill in data', 'Number of groups (k > 1)' (set to 5), and 'Number of observations in the groups' (set to 30 for each of five groups). There are 'Reset' and 'Clear All' buttons.
- Method Selection:** A section titled 'Select the methods you want to perform' with checkboxes for 'Confirmation', 'Fbar', 'ORIC', and 'BMS'. Below this, 'Specification needed for BMS' includes 'Delta (>= 0)' (set to 0) and 'Pv (>0)' (set to 2).
- Hypothesis Specification:** A section titled 'Specify models for confirmatory methods' with checkboxes for 'H0: mu1 = ... = muk', 'Ha: mu1, ..., muk', and 'Add (order-restricted) hypothesis/es'. Below this are 'Edit Hypothesis 1' and 'Edit restriction(s) in H1' buttons, each with 'Add' and 'Remove' sub-buttons.
- Restriction Editor:** A central area for defining hypotheses. It shows 'Restriction 1: 5 = 3 > 1 > 2 . 4' and 'Restriction 2: 3 > 4 > 2 . 1 . 5'. A dropdown menu is open over the second restriction, showing options '>', '<', and '='.
- Execution and Output:** A 'Run' button, a 'Settings' button, and a 'Name of output file:' field containing 'Output.txt'. A 'Progress Report' panel is also visible.
- Help and About:** 'Tutorial' and 'About' buttons are present. The 'About' box contains version information (Version 0.1 Beta) and contact details for Rebecca Kuiper.

Figure 4: Specifying H_1 .

not all requirements are met and you press the Run button, a popup appears with the text “Not all requirements are met yet.”. The requirements are that the data are valid, the specification of δ and pv (if needed) is correct, and that the hypotheses (if any) are specified correctly. In case a requirement is met the checkbox is checked; otherwise, it is not checked and an error message is given in the `ConfirmatoryANOVA.exe` form.

Furthermore, error messages are built in to detect other wrongly stated input (e.g., when not using every number precisely once in a (order) restriction). A popup will appear with the text “Error in input. ... This run will be stopped.”. These error messages will be given in the `Progress Report` panel in the `ConfirmatoryANOVA.exe` form (see Figure 1 and Figure 4) and in `Error.txt` (which will be created in the folder where `ConfirmatoryANOVA.exe` is saved in). In most cases the methods will not be performed and no output will be given.

When, for some reason, the error is not detected, a popup will appear with the text “The program has stopped. No methods are performed. Check input. ...”. In that case no output will be given. One should look at the input again, especially the input needed for the method during which the error occurred. When looking at the `Progress Report` panel one can get a better idea of during which method the error has occurred. It should be noted that the methods are performed in a fixed order: the \bar{F} test is performed first, then the ORIC and the

programm ends with BMS (when all methods are performed). The progress report tells you when the method has started (e.g., “The Fbar test is running...”) and when it is ended (e.g., “Fbar is performed”).

B.5. Output

Let the hypotheses of interest be the set of hypotheses specified in (2). As mentioned before, in the output, these hypotheses will be referred to as H_1 , H_3 , H_4 , and H_2 corresponding to H_0 , H_1 , H_2 , and H_3 , respectively, in (2).

When `ConfirmatoryANOVA.exe` is done, the output file `Output.txt` (or the name you entered in the `Name of output file` textbox) will be created in the folder where `ConfirmatoryANOVA.exe` is saved in. The output file gives the results of the requested analyses. In the example, the output for the three methods (i.e., \bar{F} , ORIC, and BMS), with $\delta = 0.3$ and $pv = 2$, is:

This program is free of use. However, when results obtained with this program are published, please refer to:

Rebecca M. Kuiper, Irene Klugkist, and Herbert Hoijsink (2010).
 A Fortran 90 Program for Confirmatory Analysis of Variance.
 Journal of Statistical Software, 34(8), 1-31.
 URL <http://www.jstatsoft.org/v34/i08/>.

N.B. This paper is available upon request (R.M.Kuiper@uu.nl).

Summary of observed data

Group number, means, standard deviations, and sample sizes per group

```
1 2.33 1.86 30
2 1.33 1.15 30
3 3.20 1.79 30
4 2.23 1.45 30
5 3.23 1.50 30
```

Restricted means

```
Group number:  1 2 3 4 5
Sample means:  2.33 1.33 3.20 2.23 3.23
```

```
Hypothesis 1 2.46 2.46 2.46 2.46 2.46
Hypothesis 2 2.33 1.33 3.20 2.23 3.23
Hypothesis 3 2.33 1.33 3.21 2.23 3.21
Hypothesis 4 2.60 1.33 3.20 2.60 2.60
```

The hypotheses of interest are stated below.

- Fbar test -
 <See Section 2.1>

Results of the Fbar test for the null hypothesis 1 and the unconstrained hypothesis 2

Hypotheses numbers	Fbar value	p-value
1 versus 2	30.27	0.00

Results of the ordered alternative Fbar test

Ordered-hypothesis number	Fbar value	p-value
H0 versus 3	30.26	0.00
H0 versus 4	22.91	0.00

Results of the ordered null Fbar test

Ordered-hypothesis number	Fbar value	p-value
3 versus Hu	0.01	1.00
4 versus Hu	7.36	0.07

Residual sum of squares

Hypothesis 1	432.53
Hypothesis 2	0.00
Hypothesis 3	357.85
Hypothesis 4	375.99

The hypotheses of interest are stated below.

- ORIC -
 <See Section 2.2>

The value of the Order-Restricted Information Criterion (ORIC) =
 $-2 * \log \text{likelihood} + 2 * \text{penalty}$:

for Hypothesis 1,	$ORIC = -2 * -292.27 + 2 * 2.00 = 588.54$
for Hypothesis 2,	$ORIC = -2 * -278.05 + 2 * 6.00 = 568.10$
for Hypothesis 3,	$ORIC = -2 * -278.05 + 2 * 3.19 = 562.49$
for Hypothesis 4,	$ORIC = -2 * -281.76 + 2 * 3.14 = 569.79$

The preferred hypothesis, according to the Order-Restricted Information

Criterion, of the hypotheses to be compared is hypothesis number 3.

The hypotheses of interest are stated below.

- BMS -

<See Section 2.3>

The resulting Bayes factor values (of the order-restricted hypothesis versus the unconstrained model) and the posterior model probabilities (with respect to the whole set of models) are:

Hypothesis 1	0.00	0.00
Hypothesis 2	1.00	0.01
Hypothesis 3	67.94	0.96
Hypothesis 4	1.52	0.02

The preferred hypothesis, according to Bayesian model selection, of the hypotheses to be compared is hypothesis number 3.

The hypotheses of interest are stated below.

Specification of the encompassing prior:

For all means, the same normal prior with mean
2.28
 and variance
2.32
 is used.

For the residual variance, a scaled inverse chi-square with
 degrees of freedom
1.00
 and scale parameter
2.50
 is used.

- The hypotheses of interest -

Hypothesis 1 (= 'H0')

Restriction 1: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

Hypothesis 2 (= 'Ha')

Restriction 1: $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$

Hypothesis 3

Restriction 1: $\mu_5 = \mu_3 > \mu_1 > \mu_2$, μ_4 Restriction 2: $\mu_3 > \mu_4 > \mu_2$, μ_1 , μ_5

Hypothesis 4

Restriction 1: $\mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2$ **Affiliation:**

Rebecca M. Kuiper

Department of Methodology and Statistics

Utrecht University

P.O. Box 80140

3508TC Utrecht, The Netherlands

E-mail: r.m.kuiper@uu.nlURL: <http://www.fss.uu.nl/ms/Kuiper/>