Inequality constrained hypotheses for ANOVA

Maaike van Rossum *, Rens van de Schoot, Herbert Hoijtink

Department of Methods and Statistics, Utrecht University

Abstract

In this paper a novel approach for the evaluation of inequality constrained hypotheses is described. An inequality constrained hypothesis (H_i) is an hypothesis with order restrictions between the parameters of interest. The Bayes factor is used to compare H_i with its complement H_c, that is all situations in which H_i is not true. This approach is applied to ANOVA models and the performance is evaluated by looking at the error probabilities, which are the counterparts of Type 1 and Type 2 errors in the traditional null hypothesis testing framework. Furthermore, the robustness of the approach proposed with respect to violations of the assumption of homogeneity of variances is evaluated. Two examples are also analyzed using this new approach. The overall result is that the approach works well, with low error probabilities for H_i for relatively small sample sizes. Furthermore, the approach is robust to violations to the assumptions of homogeneity of variances.

Keywords: Inequality constrained hypothesis, ANOVA, Bayes factor

---

*Correspondence should be addressed to Maaike van Rossum: Department of Methods and Statistics, Utrecht University, P.O. Box 80.140, 3508TC, Utrecht, The Netherlands; Tel.: +31 302534468; Fax: +31 2535797; E-mail address: M.L.vanRossum@students.uu.nl.

1 Introduction

This paper describes a novel approach for evaluating an inequality constrained (or informative) hypothesis, which is an hypothesis with order restrictions between the parameters of interest. Evaluating inequality constrained hypotheses is not new (see for example: Robertson, Wright, & Dykstra, 1988; Barlow, Bartholomew, Bremner, & Brunk, 1972; Silvapulle & Sen, 2005; Hoijtink, Klugkist, & Boelen, 2008), but in this paper the inequality constrained hypothesis will be compared to its complement, which has not been done before. This paper is restricted to the evaluation of inequality constrained hypotheses in the context of Analysis of Variance (ANOVA) models.

Many researchers have a specific expectation about the outcome of their research, which often can be formulated as an inequality constrained hypothesis. For example, in an hypothetical study with four means of interest, the expectation could be that, on average, Group 1 scores higher than Group 2, Group 2 scores higher than Group 3 and Group 3 scores higher than Group 4. Let μ denote the group mean, than the corresponding inequality constrained hypothesis is:

$$H_i : \mu_1 > \mu_2 > \mu_3 > \mu_4. \tag{1}$$

Although researchers frequently have such expectations, they often restore to testing the null hypothesis, which in this hypothetical example is:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4, \tag{2}$$

which is tested against the alternative hypothesis,

$$H_a : \mu_1 \ , \ \mu_2 \ , \ \mu_3 \ , \ \mu_4. \tag{3}$$

Testing Equation (2) against Equation (3) is not a straightforward evaluation of Equation (1). Even in combination with pairwise comparisons or planned contrasts, would this not render a straightforward evaluation of H_i (Hoijtink, Huntjes, Reijntjes, Kuiper, & Boelen, 2008).

There are already two approaches for a direct evaluation of inequality constrained hypotheses. Such as the $\bar{F}$-test, developed by Silvapulle and Sen (2005), which evaluates H_i by comparing it to H_a:

$$H_i \quad : \quad \mu_1 > \mu_2 > \mu_3 > \mu_4 \tag{4}$$
$$H_a \quad : \quad \mu_1 \ , \ \mu_2 \ , \ \mu_3 \ , \ \mu_4.$$

The $\bar{F}$-test is an adaptation of the F-test for ANOVA models, for the evaluation

of a single inequality constrained hypothesis (see also Kuiper & Hoijtink, 2010).

The second approach is Bayesian evaluation of informative hypothesis using Bayesian Model Selection, which is described by Klugkist, Laudy, and Hoijtink (2005), Hoijtink, Klugkist, and Boelen (2008) and Kuiper and Hoijtink (2010) in the context of ANOVA. With this approach H_i is compared to another competing inequality constrained hypothesis: H_i'. For software for the $\bar{F}$-test and Bayesian evaluation of informative hypothesis for ANOVA models, see the paper by Kuiper, Klugkist, and Hoijtink (2010). Readers interested in the software can also visit *www.fss.uu.nl/ms/informativehypothesis*.

Thus, in the $\bar{F}$-test H_i is compared to H_a, which encompasses H_i. However, it is not logical to compare H_i to an hypothesis that also encompasses H_i, furthermore, researchers do not always have competing hypotheses. Therefore, in the current paper, one single H_i will be evaluated by comparing it to it's complement H_c:

$$
\begin{aligned}
H_i &: \quad \mu_1 > \mu_2 > \mu_3 > \mu_4 \\
\text{vs} & \\
H_c &: \quad \text{not } H_i.
\end{aligned}
\tag{5}
$$

Note that the complement encompasses all the situations in which H_i is not true.

In the next section, two examples are introduced, which are used throughout this paper to illustrate the approach proposed in Equation (5). Section 3 elaborates on the generalization of inequality constrained hypotheses. Section 4 describes how H_i can be compared to H_c by making use of the Bayes factor. Section 5 introduces error probabilities for the comparison of H_i versus H_c. Section 6 evaluates the robustness of the approach proposed with respect to violations of the assumption of homogeneity of variances. In Section 7, the inequality constrained hypotheses of the two examples introduced in the next section, are evaluated using the novel approach proposed in the current paper.

2 Examples

This paper focuses on the evaluation of inequality constrained hypothesis in the context of ANOVA models:

$$
y_i = \sum_{j=1}^{J} \mu_j d_{ij} + \varepsilon_i,
\tag{6}
$$

Table 1: Descriptives Example 1

	Mean	σ^2	N
μ_1	0.54	0.096	30
μ_2	0.20	0.048	30
μ_3	0.29	0.063	30
μ_4	0.19	0.068	30

where y_i is the observation of the dependent variable of person i ($i = 1, \ldots, N$), μ_j is the mean of group j ($j = 1, \ldots, J$), and d_{ij} denotes the group membership of a person, with 0 denoting not being a member of the group and 1 denoting being a member of the group. The residuals, ε_i, for each group are assumed to be normally distributed with mean zero and variance σ^2.

Below, two examples are presented, one based on Geraerts et al. (2009) and one on Van de Schoot, Velden, Boom, and Brugman (2010). In both articles, the researchers formulated a clear expectation. These two examples are used to show how expectations can be translated into inequality constrained hypotheses.

2.1 Example 1

The first example is a study by Geraerts et al. (2009), who investigated the cognitive mechanisms underlying recovered memory experiences of childhood sexual abuse. Four different groups participated in the experiment: Group 1: memory recovered in therapy; Group 2: memory recovered spontaneously; Group 3: continuous memory and Group 4: a control group containing persons who reported no memory of abuse. In their introduction, Geraerts et al. (2009) specifically state the following expectation:

> "We predicted that people reporting childhood sexual abuse memories recovered during suggestive therapy would score high on a measure of susceptibility to false memories ... people who report spontaneously recovered memories of abuse ... would score similarly to control subjects on false-memory tasks. (p. 93)"

Clearly, the researchers have one expectation they want to test and this expectation can be translated into the following inequality constrained hypothesis:

$$H_{i1} : \mu_1 > \{\mu_2, \mu_3, \mu_4\}, \tag{7}$$

which states that the mean false recall rate in Group 1 (memory recovered in therapy) is larger than in the other three groups. There are no predictions with respect to the relative size of the false recall rate within the other three groups.

4

False-recall rate was measured as the average proportion of false recall of words over several different experiments. The mean proportions are given in Table 1.

Despite their expectation, Geraerts et al. (2009) analyzed their data using traditional null hypothesis testing in which they compared the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ with the alternative hypothesis H_a. The result of traditional null hypothesis testing is either that H_0 can not be rejected, or that H_0 can be rejected. Because rejection of H_0 is not yet evidence in favor of H_i, pairwise comparisons (see, for example, Toothaker, 1993) or contrast analysis (see, for example, Rosenthal & Rosnow, 1985) can be used for further analysis. However, the resulting procedure does not provide a straightforward evaluation of H_i.

The aim of this paper is not to criticize traditional null hypothesis testing (the interested reader is referred to: Cohen, 1994; Wagenmakers, 2007; Hoijtink, Huntjes, et al., 2008), nor to give a comparison of different methods (Kuiper & Hoijtink, 2010). The aim of this paper is to propose a method for the comparison of an inequality constrained hypothesis with its complement using the Bayes factor, and to investigate the properties of this method.

2.2 Example 2

The second example comes from a study by Van de Schoot et al. (2010). They investigated the association between popularity and antisocial behaviour in a large sample of young adolescents from preparatory vocational schools in the Netherlands. Five, so-called, sociometric status groups were defined: Group 1: a controversial group of adolescents, Group 2: a rejected group, Group 3: an average group, Group 4: a popular group, and Group 5: a neglected group of adolescents. Each sociometric status group has been characterized by distinct behavioural patterns which influence the quality of social relations (Newcomb, Bukowski, & Pattee, 1993).

It could, for example, be expected that the controversial adolescents would report more signs of anti-social behaviour than the rejected group, the rejected group would report more signs than the average group, the average group would report more signs than the popular group and that the popular group would report more signs of anti-social behaviour than the neglected group. These expectations can be summarized in the following inequality constrained hypothesis:

$$H_{i2} : \mu_1 > \mu_2 > \mu_3 > \mu_4 > \mu_5, \tag{8}$$

which states that the mean reported signs of anti-social behaviour for Group 1 is larger than for Group 2, for Group 2 is larger than for Group 3, for Group 3

Table 2: Descriptives Example 2

	Mean	σ^2	N
μ_1	0.3023	0.54	29
μ_2	0.0537	0.35	205
μ_3	0.0508	0.35	954
μ_4	-0.0203	0.36	215
μ_5	-0.0907	0.29	88

is larger than for Group 4 and for Group 4 is larger than Group 5. Anti-social behaviour was measured with a questionnaire of eight items scored on a 4-point scale. Table 2 shows for each group, the mean of the factor scores obtained analyzing these eight items.

3 Inequality constrained hypotheses

The previous section described two examples of inequality constrained hypotheses. This section shows that these inequality constrained hypotheses can be generalized by formulating them as follows:

$$H_i : \mathbf{R}\boldsymbol{\mu} > \mathbf{0}, \tag{9}$$

where $\mathbf{R}$ denotes a $K \times P$ matrix containing real numbers, and $\mathbf{0}$ denotes a vector of length K.

For the inequality constrained hypothesis of Example 1, $H_{i1} : \mu_1 > \{\mu_2, \mu_3, \mu_4\}$:

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}, \ \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} \text{ and } \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{10}$$

which renders:

$$H_{i1} = \begin{bmatrix} \mu_1 - \mu_2 > 0 \\ \mu_1 - \mu_3 > 0 \\ \mu_1 - \mu_4 > 0 \end{bmatrix}, \tag{11}$$

which is equivalent to $H_{i1} : \mu_1 > \{\mu_2, \mu_3, \mu_4\}$.

For Example 2, $H_{i2} : \mu_1 > \mu_2 > \mu_3 > \mu_4 > \mu_5$ can be written in the

following form:

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \ \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{bmatrix} \text{ and } \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{12}$$

which renders:

$$H_{i2} = \begin{bmatrix} \mu_1 - \mu_2 > 0 \\ \mu_2 - \mu_3 > 0 \\ \mu_3 - \mu_4 > 0 \\ \mu_4 - \mu_5 > 0 \end{bmatrix}, \tag{13}$$

which is equivalent to $H_{i2} : \mu_1 > \mu_2 > \mu_3 > \mu_4 > \mu_5$.

The matrix $\mathbf{R}$ can also be used in other situations for specifying inequality constrained hypotheses. Consider, for example, $H_i : \mu_{11} - \mu_{21} > \mu_{12} - \mu_{22}$, that is a constrained interaction effect for a 2 by 2 factorial design. This hypothesis can be constructed using:

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix}, \ \boldsymbol{\mu} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} \text{ and } \mathbf{0} = \begin{bmatrix} 0 \end{bmatrix}. \tag{14}$$

Another example is $H_i : \mu_2 - \mu_1 > \mu_3 - \mu_2$, where μ_1 is a control group and μ_2 and μ_3 are treatment groups. This hypothesis states that, the difference between the two treatment groups is smaller than the difference between the control group and the first treatment group. This hypothesis can be constructed using:

$$\mathbf{R} = \begin{bmatrix} -1 & 2 & -1 \end{bmatrix}, \ \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \text{ and } \mathbf{0} = \begin{bmatrix} 0 \end{bmatrix}. \tag{15}$$

4 The Bayes Factor

The Bayes factor (Kass & Raftery, 1995) will be used for the comparison of H_i versus H_c. It is a measure of the relative support for both hypotheses. The Bayes factor for H_i versus H_c can be derived from the Bayes factor for H_i versus H_a. In this section the Bayes factor for H_i versus H_a is given and explained, and in Section 4.4 the Bayes for H_i versus H_c is derived.

As was shown by Klugkist and Hoijtink (2007), the Bayes factor of H_i versus

H_a can be written as

$$BF_{ia} = \frac{f_i}{c_i}, \tag{16}$$

where f_i can be interpreted as the fit and c_i as the complexity of H_i. Both components will now be further elaborated.

4.1 Complexity

Complexity, c_i, is measured as the proportion of the prior distribution of H_a in agreement with the constraints of H_i. According to the encompassing prior approach as proposed by Klugkist et al. (2005), Mulder, Klugkist, et al. (2009) and Mulder, Hoijtink, and Klugkist (2009), the prior for H_a in the context of ANOVA models is:

$$\begin{aligned}
h(\boldsymbol{\mu}, \sigma^2 | H_a) &= \mathcal{N}(\mu_1 | \mu_0, \tau_0^2) \times \ldots \times \mathcal{N}(\mu_J | \mu_0, \tau_0^2) \\
&\quad \times \text{Inv-Gamma}(\sigma^2 | a, b),
\end{aligned} \tag{17}$$

where μ_0 denotes the prior mean, τ_0^2 the precision and a and b are respectively, the shape and the scale parameter of the inverse gamma distribution. From this, it can be derived that:

$$h(\boldsymbol{\mu}, \sigma^2 | H_i) = \frac{1}{c_i} \times h(\boldsymbol{\mu}, \sigma^2 | H_a), \tag{18}$$

where c_i is the proportion of the prior distribution of H_a in agreement with H_i.

Note that in the encompassing prior approach the same prior distribution is chosen for each μ_j. Furthermore, a and b should be chosen in such a way that it results in a vague prior. As was shown by Hoijtink (2009), c_i is independent of μ_0 for hypotheses of the form $\mathbf{R}\boldsymbol{\mu} > \mathbf{0}$ if $\tau_0^2 \to 0$. This vague and default prior distribution for each μ_j renders a posterior distribution, and therefore also f_i, that are completely determined by the data. In this sense, the resulting Bayes factor can be called objective.

4.2 Fit

As was shown by Klugkist et al. (2005), Mulder, Klugkist, et al. (2009) and Mulder, Hoijtink, and Klugkist (2009), fit, f_i, is the proportion of the posterior distribution of H_a in agreement with the inequality constrained hypothesis H_i. The posterior distribution is proportional to the product of the prior and the likelihood function of the data. For ANOVA models, the likelihood is given by:

$$f(y | \boldsymbol{\mu}, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{(y_i - \sum_{j=1}^{J} \mu_j d_{ij})^2}{2\sigma^2} \right\}. \tag{19}$$

4.3 Calculation of complexity and fit

The previous two sections explained the complexity and fit and this section describes how they can be computed.

The complexity can be estimated using a sample with elements $t = 1, \ldots, T$ from $h(\boldsymbol{\mu}|H_a)$:

1. Sample μ_j^t from $\mathcal{N}(\mu_j|\mu_0, \tau_0)$ for $j = 1, \ldots, J$ and $t = 1, \ldots, T$

2. Estimate c_i by the proportion of $\boldsymbol{\mu}^t$ for $t = 1, \ldots, T$ in agreement with H_i

Complexity can be estimated using WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000), which is explained in Appendix A.

When following the above described approach with the specification of the priors from Section 4.1, it can be shown that the complexity of the hypotheses used in our examples can be calculated using the following guidelines:

- For $H_i : \mu_1 > \{\mu_2, \ldots, \mu_J\}$, $c_i = 1/J$

- For $H_i : \mu_1 < \ldots < \mu_J$, $c_i = 1/J!$

Consider Example 1 with $H_{i1} : \mu_1 > \{\mu_2, \mu_3, \mu_4\}$. If four means are sampled from the prior as specified in Equation (17), $\mathrm{P}(\mu_1 > \{\mu_2, \mu_3, \mu_4\}|H_a)$ is equal to $\mathrm{P}(\mu_2 > \{\mu_1, \mu_3, \mu_4\}|H_a)$, $\mathrm{P}(\mu_3 > \{\mu_1, \mu_2, \mu_4\}|H_a)$ or $\mathrm{P}(\mu_4 > \{\mu_1, \mu_2, \mu_3\}|H_a)$. Therefore each of these four equivalent models has the same complexity: $c_i = 1/4 = 25\%$.

In the same manner, fit can be estimated using a sample from the posterior distribution:

1. Sample $\boldsymbol{\mu}^t$ from the posterior distribution for $t = 1, \ldots, T$. This can be done using the Gibbs sampler.

2. Estimate f_i by the proportion of $\boldsymbol{\mu}^t$ for $t = 1, \ldots, T$ in agreement with H_i

Just like complexity, fit can also be estimated using WinBUGS, this is described in Appendix B.

To determine the size of the samples from the prior and posterior necessary to obtain stable estimates of c_i and f_i, the convergence of c_i and f_i can be monitored. For each example the convergence plots of c_i and f_i will be given in Section 7.

4.4 The Bayes Factor for H_i vs. H_c

Since H_c is the complement of H_i, it follows that $f_c = 1 - f_i$ and $c_c = 1 - c_i$. The Bayes factor of H_c versus H_a can then be written as:

$$BF_{ca} = \frac{f_c}{c_c} = \frac{1 - f_i}{1 - c_i}. \tag{20}$$

Combining Equation (16) and Equation (20), it follows that

$$BF_{ic} = \frac{BF_{ia}}{BF_{ca}} = \frac{f_i/c_i}{(1-f_i)/(1-c_i)}. \tag{21}$$

The Bayes factor can be interpreted as a relative measure of support. If $BF_{ic} > 1$, H_i is more supported by the data than H_c. If $BF_{ic} = 1$, none of the two hypotheses is preferred by the data. For $BF_{ic} < 1$, H_c is more supported by the data than H_i. Note that for Bayes factors the value 1 is a natural decision point, but it should be clear that a Bayes factor of 50 provides more support for an hypothesis than a Bayes factor of 1.05 does. By computing the estimates for c_i and f_i in WinBUGS, BF_{ic} can simply be computed using Equation (21).

5   Error Probabilities

The Bayes factor is a relative measure of support for H_i and H_c, where the decision which hypothesis is more supported by the data is based on the cut-off value 1. As is illustrated in Figure 1, if a data-set is sampled from a population in agreement with H_c, there is a non-zero probability, α_c, that the resulting Bayes factor is larger than 1. The probability of making a wrong decision for a data-set sampled from a population in agreement with H_c is given by:

$$\alpha_c = P(BF_{ic} > 1|H_c). \tag{22}$$

Similarly, if a data-set is sampled from a population in agreement with H_i, there is a non-zero probability, α_i, that the resulting Bayes factor is smaller than 1. The probability of making a wrong decision for a data-set sampled from a population in agreement with H_i is given by:

$$\alpha_i = P(BF_{ic} < 1|H_i). \tag{23}$$

These error probabilities (Garcia-Donato & Chen, 2005) are the counterparts of the Type I and Type II errors in the traditional null hypothesis testing framework.

In Section 5.1 and Section 5.2, a simulation study is executed, based on the inequality constrained hypotheses of the two examples, respectively. For a variety of situations the error probabilities are calculated. From multiple populations, in agreement with H_i or H_c, 1000 data matrices for different sample sizes are generated. Thereafter, the Bayes factor for each data matrix is computed, resulting in a distribution of Bayes factors for every stated population and sample size. In WinBUGS, 10,000 iterations are used to generate a sample
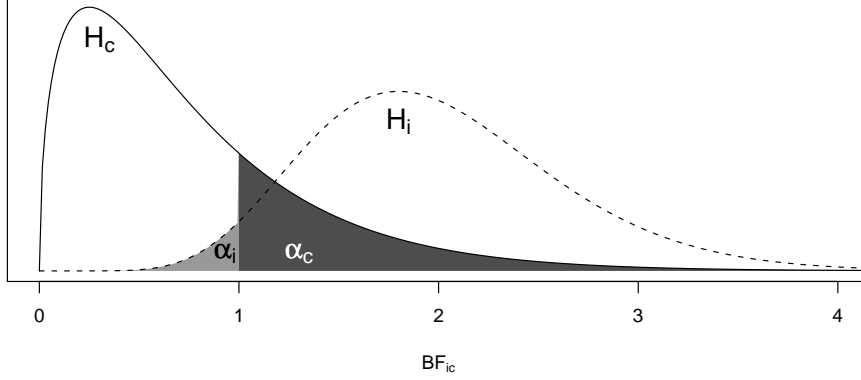
Figure 1: Error probabilities

from the posterior distribution with a burn-in period of 1,000 iterations.

5.1 Simulation 1

The first simulation is based on the inequality constrained hypothesis of Example 1:

$$H_{i1} : \mu_1 > \{\mu_2, \mu_3, \mu_4\}. \tag{24}$$

Five populations are used in this simulation. The population means are displayed in Table 3. Note that for each group in each population $\sigma^2 = 1.0$. For each population, data matrices with different sample sizes are generated ($N = 10, 20$ and 40). The first two populations are in agreement with H_i and are based on, respectively, a medium effect size of 0.5 for M_1 and a small effect size of 0.2 for M_2 according to Cohen's d (Cohen, 1988):

$$d = \frac{\mu_1 - \mu_2}{\sigma}. \tag{25}$$

The last three populations are in agreement with H_c. The third population (M_3) is the opposite of M_2, with a Cohen's d of -0.2. The fourth (M_4) and fifth population (M_5) have one and two violations, respectively, of the ordering expected under H_i and are based on a Cohen's d of 0.5.

In Table 3, the models are ordered in such a manner that they range from the most in agreement with H_i to the least in agreement with H_i. M_1 is more in agreement with H_i than M_2, because of the larger effect size of M_1. For the models in agreement with H_c, M_5, with two violations of the ordering expected under H_i, is expected to be the least in agreement with H_i. M_3 and M_4 both have one violation of the ordering expected under H_i.

11

Table 3: Population means used in Simulation 1

	μ_1	μ_2	μ_3	μ_4
Models in agreement with H_i				
M_1	0.5	0.0	0.0	0.0
M_2	0.2	0.0	0.0	0.0
Models in agreement with H_c				
M_3	-0.2	0.0	0.0	0.0
M_4	0.5	0.0	1.0	0.0
M_5	0.5	0.0	1.0	1.5

It is expected that, for the models in agreement with H_i, α_i will become smaller when the sample size is increased. Since M_1 is more in agreement with H_i than M_2, because of the larger effect size in M_1, it is expected that α_i for M_1 will be smaller than α_i for M_2. For the models in agreement with H_c the α_c is expected to become smaller with increased sample sizes. Furthermore, it can be expected that the models that are the least in agreement with H_i, so the most in agreement with H_c have the smallest α_c.

5.1.1 Results Simulation 1

Table 4 shows the results for this simulation. For both models in agreement with H_i, it can be seen that α_i becomes smaller with increased sample sizes. For example, for M_1 with a sample size of 10 $\alpha_i = 0.141$, with a sample size of 20 $\alpha_i = 0.060$ and with a sample size of 40 $\alpha_i = 0.012$. For the models in agreement with H_c, α_c becomes smaller with increased sample sizes.

As expected, α_i of M_1, for all sample sizes, is smaller than α_i of M_2. This is because M_1 is more in agreement with H_i. For the models in agreement with H_c, it can be seen that for each sample size α_c is the smallest for M_5, the model that is the least in agreement with H_i. Furthermore, for each sample size α_c is similar for both M_3 and M_4, the two models with one violation to the order expected under H_i. Thus in general, α_i is smaller for models that are more in agreement with H_i and α_c is smaller for models that are less in agreement with H_i.

Table 4 can be used to check how large the sample size needs to be in combination with the expected effect size to obtain reasonable error probabilities. For example, if a medium effect size is expected, the error probabilities are already small for $N = 20$ and larger. However, if a small effect size is expected, it can be seen that the error probabilities are quite high and therefore the possibility of making a wrong decision needs to be taken into account. Furthermore, the table shows that for models in agreement with H_c the error probabilities are

Table 4: Results of Simulation 1

	N	α_i	α_c
Models in agreement with H_i			
M_1	10	0.141	
	20	0.060	
	40	0.012	
M_2	10	0.366	
	20	0.344	
	40	0.204	
Models in agreement with H_c			
M_3	10		0.229
	20		0.144
	40		0.097
M_4	10		0.330
	20		0.153
	40		0.068
M_5	10		0.030
	20		0.003
	40		0.000

relatively small for a sample size of 30 and larger. So for this experiment with $H_{i1} : \mu_1 > \{\mu_2, \mu_3, \mu_4\}$ a sample size of 30 will lead to relatively small error probabilities.

5.2 Simulation 2

This second simulation study is a copy of the first simulation study but now based on the inequality constrained hypothesis of Example 2:

$$H_{i2} : \mu_1 > \mu_2 > \mu_3 > \mu_4 > \mu_5. \tag{26}$$

Again five populations are used in this simulation. The population means are displayed in Table 5. Note that for each group in each population $\sigma^2 = 1.0$. For each population, data matrices with different sample sizes are generated (N = 10, 20 and 40). The first two populations are in agreement with H_i. The first population, (M_1), is based on a Cohen's d of -0.5 (medium effect size), so the sample means decrease by 0.5 going from Group 1 to Group 5. The second population, (M_2), is based on a Cohen's d of -0.2 (small effect size). The last three populations are in agreement with H_c. M_3 and M_4 have one or two violations, respectively, of the ordering expected under H_i, and are based on a Cohen's d of -0.5. The fifth population (M_5) is exactly the opposite of M_2 with a Cohen's d of 0.2. This means that this model has five violations of the ordering expected under H_i.

Table 5: Population means used in Simulation 2

	μ_1	μ_2	μ_3	μ_4	μ_5
Models in agreement with H_i					
M_1	2.0	1.5	1.0	0.5	0.0
M_2	0.8	0.6	0.4	0.2	0.0
Models in agreement with H_c					
M_3	2.0	1.5	0.5	1.0	0.0
M_4	1.5	2.0	0.5	1.0	0.0
M_5	-0.8	-0.6	-0.4	-0.2	0.0

In Table 5, the five models are ordered in a way that the top model is the one most in agreement with H_i and the bottom one is the model least in agreement with H_i. This automatically means that the last model is also the most in agreement with H_c.

Like in Simulation 1, it is expected that, for the models in agreement with H_i, α_i should become smaller with increasing sample size. As well, it is expected that α_i for M_1 will be smaller than α_i for M_2, since M_1 is more in agreement with H_i than M_2, because the effect size of model M_1 is larger. For the models in agreement with H_c, α_c is expected to become smaller with increased sample sizes. Furthermore, it can be expected that for models less in agreement with H_i, α_c becomes smaller.

5.2.1 Results Simulation 2

Table 6 shows the results for the second simulation. For both models in agreement with H_i, α_i becomes smaller with increased sample sizes. For the models in agreement with H_c, α_c also becomes smaller with increased sample sizes.

As expected, α_i of M_1, for each sample size, is smaller than α_i of M_2, since M_1 is more in agreement with H_i because of the larger effect size. For the models in agreement with H_c, it can be seen that for each sample size α_c is the smallest for M_5, the model that is the least in agreement with H_i. Furthermore, for each sample size α_c is smaller for M_4 than for M_3, which is as expected since M_4 with two violations of the order expected under H_i is less in agreement than M_3 with only one violation. In general, α_i is smaller for models that are more in agreement with H_i and α_c is smaller for models that are less in agreement with H_i.

Table 6 can be used to determine the sample size needed, in combination with the expected effect size, to obtain reasonable error probabilities. For a population in agreement with H_i, for both a small and medium effect size, the error probabilities are already small for a sample size of 20 and larger. However,

Table 6: Results of Simulation 2

	N	α_i	α_c
Models in agreement with H_i			
M_1	10	0.002	
	20	0.000	
	40	0.000	
M_2	10	0.156	
	20	0.056	
	40	0.014	
Models in agreement with H_c			
M_3	10		0.823
	20		0.737
	40		0.550
M_4	10		0.527
	20		0.303
	40		0.095
M_6	10		0.009
	20		0.000
	40		0.000

it can be seen that with a sample size of 20, the error probabilities are still very high for populations in agreement with H_c, so there is still a high probability of making the wrong decision. Taking these error probabilities into account, a sample size of at least 40 is needed to keep the error probabilities under control.

6 Robustness

An important assumption of ANOVA models is homogeneity of variances. To further evaluate the performance of the comparison of H_i versus H_c using the Bayes factor, a simulation study is executed to determine the robustness with respect to violations of this assumption. P-values in the classical ANOVA are fairly robust to violations of the homogeneity of variance. When the sample sizes are within a ratio of 4:1 for the largest to smallest group size, then a ratio of 10:1 is acceptable for the largest to the smallest group variance (Tabachnick & Fidell, 2007). In this simulation study, both the ratio of the sample sizes as the ratio of the variances are varied in order to evaluate the robustness.

6.1 Simulation 3

Table 7 shows the population parameters and sample sizes used in this simulation. The models are all based on the first model from Simulation 2 with N = 20, with decreasing means based on Cohen's d of -0.5 and σ^2 of 1.0 per group. M_1

Table 7: Populations and group sizes used in Simulation 3

	Ratio		$\mu_1 = 2.0$	$\mu_2 = 1.5$	$\mu_3 = 1.0$	$\mu_4 = 0.5$	$\mu_5 = 0.0$
	Equal sample sizes and increasing σ^2						
M_1	1:1	N	20	20	20	20	20
	1:10	σ^2	0.182	0.591	1.000	1.409	1.818
M_2	1:1	N	20	20	20	20	20
	1:20	σ^2	0.095	0.548	1.000	1.452	1.905
	Increasing N and increasing σ^2						
M_3	1:4	N	8	14	20	26	32
	1:10	σ^2	0.182	0.591	1.000	1.409	1.818
M_4	1:4	N	8	14	20	26	32
	1:20	σ^2	0.095	0.548	1.000	1.452	1.905
M_5	1:8	N	4	12	20	28	36
	1:10	σ^2	0.182	0.591	1.000	1.409	1.818
M_6	1:8	N	4	12	20	28	36
	1:20	σ^2	0.095	0.548	1.000	1.452	1.905
	Decreasing N and increasing σ^2						
M_7	4:1	N	32	26	20	14	8
	1:10	σ^2	0.182	0.591	1.000	1.409	1.818
M_8	4:1	N	32	26	20	14	8
	1:20	σ^2	0.095	0.548	1.000	1.452	1.905
M_9	8:1	N	36	28	20	12	4
	1:10	σ^2	0.182	0.591	1.000	1.409	1.818
M_{10}	8:1	N	36	28	20	12	4
	1:20	σ^2	0.095	0.548	1.000	1.452	1.905

and M_2 are based on equal sample sizes but with unequal variances, increasing from small to large with a ratio of 1:10 for M_1 and 1:20 for M_2. The models M_3 through M_6 are based on both unequal sample sizes and variances, both increasing from small to large, with a ratio of 1:4 and 1:8 for the sample size and 1:10 and 1:20 for the variances. The models M_7 through M_{10} are also based on both unequal sample sizes and variances, but the sample sizes are decreasing from large to small while the variances are increasing from small to large.

As in Simulation 1 and 2, 1000 data matrices are generated from these populations and for each data matrix the Bayes factor is calculated, resulting in a distribution of BF_{ic} for every population from Table 7.

The results from this simulation are displayed in Table 8. For populations M_1 and M_2 it can be seen that α_i is not higher than for the original population with no violations.

The same holds for the populations with both increasing sample size and variances (M_3 - M_6), α_i values are not higher than for the original population with no violations.

Table 8: Results of Simulation 3	
	α_i
No violations	
	0.000
Equal sample sizes and increasing σ^2	
M_1	0.000
M_2	0.000
Increasing N and increasing σ^2	
M_3	0.000
M_4	0.000
M_5	0.000
M_6	0.000
Decreasing N and increasing σ^2	
M_7	0.014
M_8	0.016
M_9	0.016
M_{10}	0.020

The populations M_7 through M_{10}, with decreasing sample sizes and increasing variances, perform slightly worse than the original population with no violations. α_i values are a bit higher but are still rather small. Even for M_{10}, with decreasing sample sizes with a ratio of 8:1 and increasing variances with a ratio of 1:20, α_i is only 0.020.

Thus, looking at the error probabilities of H_i, the comparison of H_i versus its complement seems robust to violations of the assumption of homogeneity of variances.

7 Analysis of the Examples

7.1 Example 1

The inequality constrained hypothesis from Example 1 is $H_{i1} : \mu_1 > \{\mu_2, \mu_3, \mu_4\}$. Appendix A.1 and Appendix B.1 show how the complexity and fit for Example 1 are computed in WinBUGS. To determine the number of iterations necessary to obtain a stable estimate of the complexity and fit, the convergence should be inspected. Figure 2 and Figure 3 show the convergence plots of c_i and f_i for this example. The convergence plots show that 10,000 iteration is enough to obtain stable estimates of c_i and f_i. A burn-in period of 1,000 iteration is used.

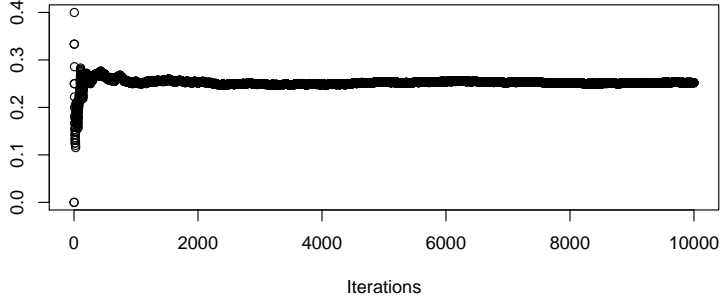The complexity for this example is 0.252 and the fit of this model is 0.999.

Figure 2: Convergence plot of the complexity of Example 1

Thus, according to equation (21), the Bayes factor is

$$\frac{0.999/0.252}{(1-0.999)/(1-0.252)} = 2965.286 \tag{27}$$

This means that the inequality constrained hypothesis: $H_{i1} : \mu_1 > \{\mu_2, \mu_3, \mu_4\}$ is 2965 times more supported by the data than its complement.

As discussed in Section 5.1.1, the error probabilities for this experiment are relatively small with a sample size of minimal 30 for each group. In this example data, $N = 30$ for each group, and moreover, the Bayes factor is 2965. This is so much evidence for H_i that the probability of making a wrong decision in this case is close to zero.

7.2   Example 2

The inequality constrained hypothesis of Example 2 is $H_{i2} : \mu_1 > \mu_2 > \mu_3 > \mu_4 > \mu_5$. Appendix A.2 and Appendix B.2 show how WinBUGS can be used
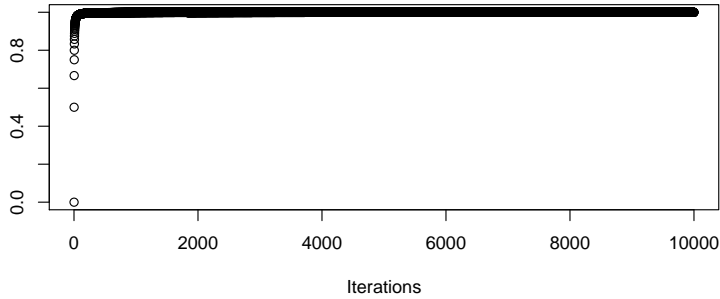

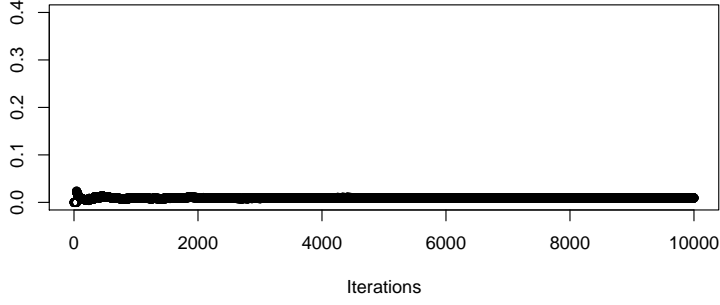Figure 3: Convergence plot of the fit of Example 1

18

Figure 4: Convergence plot of the complexity of Example 2

to obtain the complexity and fit for Example 2. Figure 4 and Figure 5 show the convergence plots for Example 2. The convergence plots show that 10,000 iteration is enough to obtain stable estimates of c_i and f_i. A burn-in period of 1,000 iteration is used.

The fit is 0.385 and the complexity of the model is 0.009. Thus, according to equation (21), the Bayes factor is

$$\frac{0.385/0.009}{(1 - 0.385)/(1 - 0.009)} = 68.931. \tag{28}$$

This means that the inequality constrained hypothesis of Example 2: $H_{i2} : \mu_1 > \mu_2 > \mu_3 > \mu_4 > \mu_5$ is 69 times more supported by the data than H_c.

As discussed in Section 5.2.1, a minimum sample size per group of 40 is needed to keep the error probabilities relatively small. In this example, there are unequal sample sizes, but this has no effect as shown in Simulation 3. Furthermore, N is only smaller than 40 in one group, the sample sizes of the other
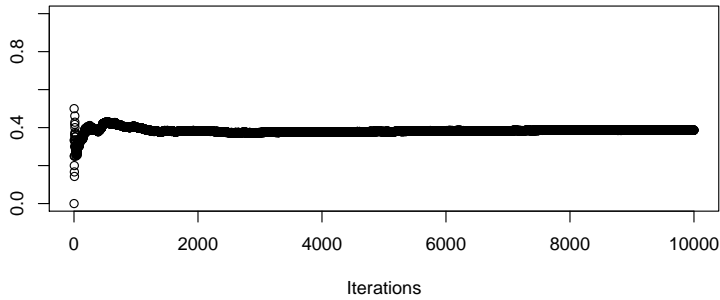

Figure 5: Convergence plot of the fit of Example 2

group are well above 40. The Bayes factor is 69, which is strong evidence for H_i. Therefore, the chance of making a wrong decision is also for this example very small.

8 Discussion

In this paper, a method is proposed for the comparison of an inequality constrained hypothesis with its complement. It was shown that the Bayes factor can be used to evaluate the comparison between H_i and H_c. Furthermore, simulation studies showed that the error probabilities depend on a combination of sample size and effect size. With larger samples sizes, the error probabilities decrease. Furthermore, with larger effect sizes, the population is more in agreement with H_i and therefore the error probabilities also decrease.

Moreover, it was shown that the new approach for evaluating H_i by comparing it to H_c is robust against violations of the assumption of homogeneity of variances.

Concluding, if the interest lies in the comparison of an inequality constrained hypothesis with its complement, than the approach proposed is a valuable addition to the existing methods.

Relevant Literature

Barlow, R. E., Bartholomew, D. J., Bremner, H. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions*. New York: Wiley.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (second ed.)*. Lawrence Erlbaum Associates.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003.

Garcia-Donato, G., & Chen, M. (2005). Calibrating bayes factor under prior predictive ditributions. *Statistica Sinica*, *15*, 359-380.

Geraerts, E., Lindsay, D., Merckelbach, H., Jelicie, M., Raymaekers, L., Arnold, M., et al. (2009). Cognitive mechanisms underlying recovered-memory experiences of childhood sexual abuse. *Psychological Science*, *20*, 92-98.

Hoijtink, H. (2009). Objective bayes factors for inequality constrained hypothesis. In *Manuscript submitted for publication*.

Hoijtink, H., Huntjes, R., Reijntjes, A., Kuiper, R., & Boelen, P. (2008). Bayesian evaluation of informative hypothesis. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), (p. Chap. 5). New York : Springer.

Hoijtink, H., Klugkist, I., & Boelen, P. A. (2008). *Bayesian evaluation of informative hypotheses*. New-York: Springer.

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773-795.

Klugkist, I., & Hoijtink, H. (2007). The bayes factor for inequality and about equality constrained models. *Computational Statistics and 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, R. M., & Hoijtink, H. (2010). Comparisons of means using exploratory and confirmatory approaches. *Psychological Methods*, *15*, 69-86.

Kuiper, R. M., Klugkist, I., & Hoijtink, H. (2010). A fortran 90 program for confirmatory analysis of variance. *Journal of Statistical Software.*, *34*, 1-31.

Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). Winbugs - a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325 - 337.

Mulder, J., Hoijtink, H., & Klugkist, I. (2009). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, *4*, 887-906.

Mulder, J., Klugkist, I., Schoot, R. Van de, Meeus, W., Selfhout, M., & Hoijtink,

H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, *53*, 530-546.

Newcomb, A., Bukowski, W., & Pattee, L. (1993). Children's peer relations: a meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin*, *113*, 99-128.

Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference.* New York : Wiley.

Rosenthal, R., & Rosnow, R. (1985). *Contrast analysis: focused comparisons in the analysis of variance?* Cambridge University Press.

Silvapulle, M. J., & Sen, P. K. (2005). *Constrained statistical inference. order, inequality and shape contraints.* Hoboken, NJ: Wiley.

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics.* Pearson.

Toothaker, L. E. (1993). *Multiple comparison procedures.* Newbury Park, CA: Sage.

Van de Schoot, R., Velden, R. van der, Boom, J., & Brugman, D. (2010). Can at-rik young adoloscents be popular and anti-social? sociometric status groups, anti-social behaviour, gender and ethnic background. *Journal of Adolescence*, *1*, 1-10.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779 - 804.

A   Computation of c_i in WinBUGS

This Appendix shows how WinBUGS can be used to compute the complexity. The WinBUGS code of both examples is given below.

A.1   WinBUGS code for Example 1

MODEL {

    # *Specification of the prior* [1]
    mu1~dnorm(0.0,0.001)
    mu2~dnorm(0.0,0.001)
    mu3~dnorm(0.0,0.001)
    mu4~dnorm(0.0,0.001)

    # *Calculation of the complexity* [2]
    c1< −step(mu1-mu2)
    c2< −step(mu1-mu3)
    c3< −step(mu1-mu4)
    complexity< −c1*c2*c3

}

[1] Here, the priors are specified with the parameters μ_0 and τ_0^2 as given in Equation (17). Note that specification of the prior for σ^2 is not necessary for the computation of c_i.

[2] As shown in Section 4.3, the proportion of $\boldsymbol{\mu}^t$ for $t = 1, \ldots, T$ in agreement with H_i needs to be computed. First, $c1 - c3$ checks if the relations between the separate means are in accordance with H_i, these are the same steps as given in Equation (11). Secondly, *complexity* checks if all the relations are in agreement with H_i. Finally, to obtain the proportion of $\boldsymbol{\mu}^t$ in agreement with H_i, the mean of *complexity* can be requested in WinBUGS.

A.2 WinBUGS code for Example 2

MODEL {

    # Specification of the prior
    mu1∼dnorm(0.0,0.001)
    mu2∼dnorm(0.0,0.001)
    mu3∼dnorm(0.0,0.001)
    mu4∼dnorm(0.0,0.001)
    mu5∼dnorm(0.0,0.001)

    # Calculation of the complexity [3]
    c1< −step(mu1-mu2)
    c2< −step(mu2-mu3)
    c3< −step(mu3-mu4)
    c4< −step(mu4-mu5)
    complexity< −c1*c2*c3*c4

}

---

[3] As in the WinBUGS code of Example 1, here the proportion of $\boldsymbol{\mu}^t$ for $t = 1, \ldots, T$ in agreement with H_i is calculated by the steps given in Equation (13).

B Computation of f_i in WinBUGS

This Appendix shows how the fit can be computes using WinBUGS. The Win-
BUGS code of both examples is given below.

B.1 WinBUGS code for Example 1

MODEL{

    #Specification of the likelihood [1]
    for(i in 1:n){
    mu[i] < − mu1*d1[i] + mu2*d2[i] + mu3*d3[i] + mu4*d4[i]
    y[i] dnorm(mu[i],inv-σ^2)}

    # Specification of the prior [2]
    mu1~dnorm(0.0,0.001)
    mu2~dnorm(0.0,0.001)
    mu3~dnorm(0.0,0.001)
    mu4~dnorm(0.0,0.001)
    inv-σ^2~dgamma(0.01,0.01)

    # Calculation of the fit [3]
    f1< −step(mu1-mu2)
    f2< −step(mu1-mu3)
    f3< −step(mu1-mu4)
    fit< −f1*f2*f3
    }

    # Specification of the data [4]
    list(n=120)

$y[]$	$d1[]$	$d2[]$	$d3[]$	$d4[]$
0.332	1	0	0	0
0.212	1	0	0	0
. . .				
. . .				
0.341	0	0	0	1
0.289	0	0	0	1

END

    # Specification of the initial values [5]
    list(mu1=1.0, mu2=1.0, mu3=1.0,mu4=1.0,inv-σ^2=1.0)

[1] For the computation of the fit, the likelihood function needs to be specified, this is based on the general likelihood function of ANOVA model as given in Equation (19).

[2] The priors for $\boldsymbol{\mu}$ are specified with the parameters μ_0 and τ_0^2 as given in Equation (17).

[3] As shown in Section 4.3, the proportion of $\boldsymbol{\mu}^t$ for $t = 1, \ldots, T$ from the posterior distribution in agreement with H_i needs to be computed. First, $f1 - f3$ checks if the relations between the separate means are in accordance with H_i, these are the same steps as given in Equation (11). Secondly, fit checks if all the relations are in agreement with H_i and finally to obtain the proportion of $\boldsymbol{\mu}^t$ in agreement with H_i the mean if fit can be requested in WinBUGS.

[4] The data need to be specified.

[5] The initial values for the parameters of the prior need to be specified.

B.2 WinBUGS code for Example 2

MODEL{

    #*likelihood*
    for(i in 1:n){
    mu[i] $<-$ mu1*d1[i] $+$ mu2*d2[i] $+$ mu3*d3[i] $+$ mu4*d4[i] $+$ mu5*d5[i]
    y[i] dnorm(mu[i],inv-σ^2)}

    #*priors*
    mu1∼dnorm(0.0,0.001)
    mu2∼dnorm(0.0,0.001)
    mu3∼dnorm(0.0,0.001)
    mu4∼dnorm(0.0,0.001)
    mu5∼dnorm(0.0,0.001)
    inv-σ^2∼dgamma(0.01,0.01)

    f1$<-$step(mu1-mu2)
    f2$<-$step(mu2-mu3)
    f3$<-$step(mu3-mu4)
    f4$<-$step(mu4-mu5)
    fit$<-$f1*f2*f3*f4
    }

    # *Specification of the data*
    list(n=1491)

$y[]$	$d1[]$	$d2[]$	$d3[]$	$d4[]$	$d5[]$
1.754	1	0	0	0	0
$-.875$	1	0	0	0	0
$-.016$	1	0	0	0	0
. . .					
. . .					
$-.424$	0	0	0	0	1
.445	0	0	0	0	1
$-.253$	0	0	0	0	1

END

    # *Specification of the initial values*
    list(mu1=1.0, mu2=1.0, mu3=1.0,mu4=1.0,mu5=1.0,inv-σ^2=1.0)