Practical Assessment, Research, and Evaluation

Volume 28 Article 1

2023

Informative Hypothesis for Group Means Comparison

Dr. Teck Kiang Tan
National University of Singapore

Recommended Citation

Tan, Dr. Teck Kiang (2023) "Informative Hypothesis for Group Means Comparison," *Practical Assessment, Research, and Evaluation*: Vol. 28, Article 1.
DOI: https://doi.org/10.7275/tme5-8x52
Available at: https://scholarworks.umass.edu/pare/vol28/iss1/1

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 28 Number 1, January 2023 ISSN 1531-7714

Informative Hypothesis for Group Means Comparison

Teck Kiang Tan, *National University of Singapore*

Researchers often have hypotheses concerning the state of affairs in the population from which they sampled their data to compare group means. The classical frequentist approach provides one way of carrying out hypothesis testing using ANOVA to state the null hypothesis that there is no difference in the means and proceed with multiple comparisons if the null hypothesis is rejected. As this approach is not able to incorporate order, inequality, and direction into hypothesis testing, and neither does it able to specify multiple hypotheses, this paper introduces the informative hypothesis that allows more flexibility in stating hypothesis testing and is directly targeted to address and state the researcher's study concern. The two new hypothesis terms under the informative hypothesis framework, the unconstrained and complementary hypotheses are introduced, and the approaches to state the level of evidence using the Bayes factor and Generalization AIC are elaborated. As this hypothesis conception is relatively new and the literature was mostly technical, the main aims of the paper are to introduce this conception, offer a general guideline, and provide an easy-to-read approach to the procedure with practical examples of carrying out this hypothesis approach and contrast it to the frequentist, using the R package.

Keywords: Informative Hypotheses, Bayesian Analysis, Bayes Factor, Posterior Model Probabilities, Generalized AIC, R Package

Introduction

Hypothesis testing is a crucial procedure in modern research that applies in almost all fields and disciplines, particularly when the research concern is to test the likelihood that there are differences between groups regarding their equality, inequality, and order of group means to address the researcher's study research concern for answering the research questions. Analysis of Variance (ANOVA), under the traditional null hypothesis significance testing (NHST), was frequently used as it is readily available in almost all common statistical software and often recommended for carrying out hypothesis testing. This approach defines a general way of carrying out hypothesis testing by specifying the null hypothesis as the equality of all the group means and the alternative hypothesis is not that specified by the null hypothesis (Maxwell and Delaney, 2004). However, criticism of the NHST approach became more frequent and common as it has many limitations that when researchers become more aware of them, they are more likely to turn away from using it as this approach does not directly address their research concern. The informative hypothesis, the Bayesian approach for hypothesis testing, becomes an option a researcher will resort to addressing their research concern as it provides greater benefits and flexibility that is closer to addressing their research interests and objectives.

Despite the development of the informative hypothesis for more than three decades, the presentation detail of this subject is mainly restricted to stating the technical and statistical representation while the focus using software to carry out the testing and showing the practical concern to demonstrate the application systematically is lacking. The motivation of the current paper is to represent this valuable hypothesis testing approach using the R package, step by step to show users how to carry out this hypothesis testing. Starting by emphasizing the limitations of the NHST approach, the informative hypothesis is introduced to state its basic concepts, purposes, benefits, and functionalities and contrast it to the frequency approach to show its superiority. The technical and statistical functionalities are explained but their formula specifications are kept to a minimum. Another focus of the paper is to give examples to state the usefulness of using the informative hypothesis and link to the specific purpose the researcher has in mind, showing the practical applications, and interpreting the results produced by the R packages, shown in a simple friendly style. Three R packages are used and their R syntax is accompanied to illustrate the hands-on relevancy that is also summarized in the appendices for reference.

Limitations and Restrictions of NHST

There is no doubt about the practicality of the inferential ANOVA model under the NHST framework to carry out group means comparison as the applications were abundant. While numerous research and data analysts almost always employed this standard framework of frequentist statistics that features the p-value, it has been criticized as there are many limitations and not in line with the research interests a researcher intends to address. This dominant way of statistical testing does not form a clear logical procedure for a researcher to understand and follow but usually, due to the past established practice, dimly followed it without detailed understanding. The file-drawer effect (Rosenthal, 1979; Royal, 1997) is one consequence of following this well-established procedure of the frequentist approach that resulted in findings that found nothing of statistical significance were not reported and absent from the publication world. This publication bias due to the absence of null findings was well noted in the literature (Moerbeek, 2019; Simmons, Nelson, & Simonsohn, 2011).

The limitations of the NHST go beyond its practical consequences boundary but its theoretical bases for carrying out hypothesis testing were also heavily criticized. First, it is hard to assume a population that is accurately described by the null hypothesis that “nothing is going on” (Altinisik, Van Lissa, Hoijtink, et al, 2021). In the context of the testing of group means, stating all the means are of equality for the null hypothesis is an unrealistic specification as there might be no population that can be in agreement with a sample size setting that there are no differences in their means (Royal, 1997). More often, the researcher is more interested to test whether there is a difference among the means. Simply stating a null hypothesis of no difference is completely not a sensible starting idea for a research study. Second, the NHST formulation is often far away from the intended theory that a researcher would like to form the hypothesis that a population with group means are exactly equal to form a plausible hypothesis (Sober, 2002). Third, it is hard to imagine a researcher prefers to wait for a rejection of a null hypothesis before proceeding to a second step finding out which group means differ from the rest. This indirect standard specification and procedure of the traditional ANOVA null hypothesis do not provide researchers with an evaluation in mind that align with their expectations (Van de Schoot, Mulder, Hoijtink, et al, 2011). Fourth, the NHST using the p-value cannot quantify the evidence in the data in favor of the hypothesis under investigation (Wagenmakers, 2007) since the p-value does not measure the probability that the studied hypothesis is true or not (Greenland, Senn, Rothman, et al, 2016; Wasserstein and Lazar, 2016; Wei, Yang, Rocha, et al, 2022). In short, there is no way to decide on accepting the null hypothesis. Fifth, a multiple hypotheses specification that formulates as specifically as possible according to the theory the researcher intends to test is not within the frame of NHST. Sixth, the value of the p-value that concludes with non-significant findings is difficult to interpret. The non-quantifying level of evidence is absent from the NHST. In summary, the NHST approach in hypothesis testing is non-informative to meet the research aim of carrying out hypothesis testing. A researcher would normally prefer to formulate one that can directly express a hypothesis that is according to theory expectation. When the absence of an effect is an important piece of information, moving away from the NHST to use the informative hypothesis becomes a preferred option.

ANOVA: The Traditional Null hypothesis Significance Testing (NHST) Approach

Under the NHST framework, ANOVA is commonly used for carrying out the comparison of group means by specifying the null hypothesis and alternative hypothesis. The following states the comparison for four groups of means.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1: \text{ not } H_0,$$

That is, $\mu_i \neq \mu_j$ for at least one pair of i, j

The null hypothesis qualifies the equality of the four group means, and the alternative hypothesis states that at least one of the group means is different from the others. The procedure to conclude the testing is to compare the p-value to a pre-specified significance level, usually, set at .05. If the p-value is less than 5%, the null hypothesis is rejected and concludes that they are at least one group's mean different from the other. The use of a pre-specified significance level became the main criticism of the NHST approach as it may not be a sensible formulation (Cohen, 1994; Harlow, Mulaik, & Steiger, 2016; Royal, 1997). Upon the rejection of the null hypothesis, multiple comparisons normally follow to determine which of the mean scores differ from each other. In this way, the results of ANOVA do not provide the evidence to conclude the acceptance of the null hypothesis. Neither does it provide the level and magnitude of evidence regarding the difference in means when the null hypothesis is rejected.

Why use Informative Hypotheses? For the last 30 years, the literature on informative hypothesis were on the rise (e.g. Altinisik, Van Lissa, Gu, Hoitink, et al, 2021; Boehm, Steingroever & Wagenmakers, 2018; Hoijtink, 1998, 2000, 2001, 2012, 2013; Hoijtink, Klugkist, & Boelen, 2008; Hoijtink, Muder, Van Lissa, & Gu, 2008; Klugkist & Hoijtink, 2007; Klugkist, Laudy, & Hoijtink, 2005; Kuiper, Klugkist, & Hoijtink, 2010; Kuiper, Hoijtink & Silvapulle, 2012; Laudy & Hoijtink, 2007; Laudy, Boom, & Hoijtink, 2005; Mulder, Hoijtink, & Klugkist, 2010; Mulder, William, Gu, et al, 2021; Klugkist, Laudy & Hoijtink, 2005; Klugkist, Laudy, & Hoijtink, 2010; Rouder, Speckman, Sun, et al, 2009; Van de Schoot, 2010; Vanbrabant, 2020; Vanbrabant, Van de Schoot, & Rosseel, 2015; Vanbrabant, Van Rossum, Van de Schoot, & Hoijtink, 2013; Wei, Yanf, Rocha, et al, 2022). This increasing trend is not without a valid reason. There are at least three major benefits that researchers turned to informative hypothesis testing.

First, many researchers have specific expectations about the outcomes of their research, which often can formulate their hypothesis unambiguously. Compare to the NHST, the informative hypotheses can much more easily relate the researcher's expectation to a set of hypotheses using inequality, and order/direction specification, to state the hypotheses directly and specifically. The informative hypothesis turns out as the most appropriate way that allows for specifying the formal structure of the hypotheses to be in line with the research objectives. The testing hypothesis becomes direct under the informative hypothesis rather than the indirect specification in NHST. For instance, a study that intends to test whether the mean of the control group is three-time in the magnitude of effect size than that of the experiment group can be easily carried out by specifying an ordered informative hypothesis that could not be carried out under NHST.

Second, the greatest advantage of using the informative hypothesis is perhaps that multiple hypothesis testing is always possible. Specifying a series of hypotheses is a key feature of the informative hypothesis that allows a researcher not only to state a list of hypotheses but also to quantify their relative level of evidence to identify the best hypothesis to address the research questions and also regarding the ranking and level of importance from a set of hypotheses. For instance, Heck, Boehm, Böing-Messing, et al (2022) illustrated the Lucas (2003) study that studied the effect of the institutionalization of female leadership on perceived leadership using multiple informative hypotheses to compare five group means with three hypotheses and an unconstrained hypothesis. The following three hypotheses show an example of specifying a set of three hypotheses with the first hypothesis specifying the equality of four years of means, the second with increasing order, and the third reversing the means of the second hypothesis for the year 2018 and the year 2019 which the former is higher than the latter.

$$H_1: \mu_{2017} = \mu_{2018} = \mu_{2019} = \mu_{2020}$$

$$H_2: \mu_{2017} < \mu_{2018} < \mu_{2019} < \mu_{2020}$$

$$H_3: \mu_{2017} < \mu_{2019} < \mu_{2018} < \mu_{2020}$$

The third benefit of employing the informative hypothesis is not restricted to the specification of hypothesis testing, it also relates to research design. For instance, Monin, Sawyer, & Marquez (2008) experimented by comparing 27 groups with four 29 conditions using the informative hypothesis. Moerbeek (2019) used an order hypothesis of a school-based smoking prevention intervention with four treatment groups. The informative hypothesis not only helps a researcher to specify the hypothesis specifically, but it also aids to plan their research design that aims to reduce the cost of research. The advantage is that given the same sample size, the power of the informative hypothesis is higher. Power can be gained and inherently a smaller sample size is expected using the informative hypothesis (Vanbrabant, Van de Schoot & Roseel., 2015). The gain in power and smaller sample size requirement are not limited to mean comparison, they are also applicable to other models such as the structural equation model (Van de Schoot & Strohmeier, 2011).

Informative Hypothesis

What is an Informational Hypothesis? The informative hypothesis is also known as the inequality-constrained hypothesis. It contains constraints to specify hypothesis testing about the ordering of means, regression coefficients, and statistical parameters. Hoijtink et al (2019) further emphasized the inclusion of effect sizes and range constraints in the choice of constraints of the informative hypothesis. The type of constraints to formulate an information hypothesis include Larger Than ">", Smaller Than "<", Equal To "=", And "&", Minus "-", and, Plus "+" that place between the specified parameters (Hoijtink, Mulder & Van Lissa, et al. 2019; Van de Schoot, 2010). Hoijtink, Mulder & Van Lissa, et al (2019) define the space of the informative hypotheses by stating the specification of the expected relations between parameters (e.g. means) classified under four building blocks. The first building block emphasizes the use of equality and order constraints between parameters. The second block is the specification of equality and order constraints between combinations of parameters. The third block specifies the use of effect size, and the fourth block highlights the range constraint. The basic three forms for the first building block for the comparison of two means are $\mu_1 > \mu_2$, $\mu_1 < \mu_2$, and $\mu_1 = \mu_2$. The second building block examples such as $\mu_1 - \mu_2 > \mu_3 - \mu_4$, $\mu_1 + \mu_2 > \mu_3 + \mu_4$, and $(\mu_1 - \mu_2 > \mu_3 - \mu_4)\,\&\,(\mu_1 + \mu_2 > \mu_3 + \mu_4)$ show three illustrations of combinations of parameters of four means. An example of the third block, $\mu_1 > \mu_2 + 0.5\hat{\sigma}$, shows the inclusion of effect size to state an informative hypothesis. The hypothesis $|\mu_1 > \mu_2| < 0.5\hat{\sigma}$ is a range constraint example, classified under the fourth block.

Informative, Complement, and Unconstrained Hypothesis. While the traditional NHST specifies no effect for the null hypothesis and the alternative hypothesis is that of not including in the null hypothesis, the informative hypothesis introduces two new hypothesis terms that are not adopted in the frequentist approach. An informative hypothesis states the (in)equality constraints of model parameters as described in the previous section by specifying a hypothesis or a set of hypotheses H_i. These hypotheses are compared to either an unconstrained hypothesis or a complement hypothesis. For instance, with four group means specifying the inequality constraints of stating $H_1\colon \mu_1 < \mu_2 < \mu_3 < \mu_4$ and $H_2\colon \mu_1 < \mu_3 < \mu_2 < \mu_4$ of two informative hypotheses that provide two ordering of mean scores, are compared to either an unconstrained hypothesis or a complementary hypothesis.

The unconstrained hypothesis also referred to as encompassing hypothesis, places no constraints on the model that represents an alternative hypothesis that covers every ordering of parameter values that is not in line with the original hypothesis. In the context of group means comparison, an unconstrained hypothesis does not impose constraints on the means (Hoijtink, 2012). The unconstrained hypothesis of four group means is stated as $H_\mathrm{u}\colon \mu_1, \mu_2, \mu_3, \mu_4$. The commas in between the means notate there is no specific order of the means.

The complement hypothesis is the complement specification to the specified set of informative hypotheses. It is an alternative hypothesis that covers every ordering of parameter values that is not in line with the original set of hypotheses (Böing-Messing, Van Assen, Hofman, et al, 2017). The complementary hypothesis is indicated by the symbol H_c. In the context of the research theory, given the above informative hypothesis H_1 and H_2 of four group means that express the belief of the researcher's theory on the order of the means, the complementary $H_\mathrm{c}\colon !\,(H_1\ or\ H_2\ )$ states the complementary. That is, H_c is the hypothesis that is not in line with the researcher's expectation theory (Moerbeek, 2019; Van Lissa, Gu, Mulder, et al, 2021). For a set of three hypotheses, the H_c thus becomes $H_\mathrm{c}\colon !\,(H_1\ or\ H_2\ or\ H_3)$. For hypotheses with at least one equality constraint, the unconstrained hypothesis and the complement are the same (Van Lissa, Gu, Mulder, et al 2021).

Bayes Factor, Complexity, and Fit. The Bayes factor (BF) is one of the oldest and most widely used indicators for carrying out testing a hypothesis under the Bayesian framework. It compares the predictive ability of two competing models corresponding to the hypotheses and indicates the degree of evidence in a data set between the null hypothesis and the alternative hypothesis (Jeffreys, 1935). There is increasing attention on the Bayes factor for the evaluation of constrained hypotheses (e.g. Gu, Hoijtink, Mulder, et al, 2019; Gu, Hoijtink, Mulder, et al, 2020; Hoijtink, 2012; Hoijtink, Klugkist, & Boelen, 2008; Hoijtink, Mulder, Van Lissa & Gu, 2019; Kato & Hoijtink, 2006, Klugkist & Hoijtink, 2007; Klugkist, Laudy & Hoijtink, 2005, Klugkist, Laudy, & Hoijtink, 2010, Laudy & Hoijtink, 2007; Mulder, Mulder, Hoijtink & Klugkist, 2010; Klugkist, Van de Schoot, Mulder, Hoijtink, et al, 2011; Wei, Yang, Rocha, et al, 2022). In the context of the informative hypothesis, the Bayes factor is a quantification of the level and degree of evidence in the collected data in favor of an informative hypothesis. The Bayes factor, BF_{iu}, of an inequality-constrained hypothesis H_{i} against an unconstrained hypothesis H_{u} can be represented as the ratio of the posterior, the fit, and prior probabilities, the complexity respectively that the inequality constraints hold (Gu, Mulder & Hoijtink, 2018; Hoijtink, 2012; Mulder, Hoijtink, & Klugkist, 2010). BF can also be written as a ratio of two marginal likelihood values of the hypotheses given the data (Klugkist & Hoijtink, 2007), as shown in Equation (1) below.

$$BF_{\mathrm{iu}} = \frac{\mathrm{f_i}}{\mathrm{c_i}} = \frac{\text{Fit } H_{\mathrm{i}}}{\text{Complexity } H_{\mathrm{i}}} = \frac{m(H_{\mathrm{i}}|data)}{m(H_{\mathrm{u}}|data)} \tag{1}$$

where c_i represents complexity, the proportion of the prior distribution that is supported by or in agreement with the hypothesis H_i, and f_i, fit, is the proportion of the posterior distribution that is supported by or in agreement with the hypothesis H_i. The fit reflects the extent to which the data is in agreement with the restrictions specified in the hypothesis whereas the complexity reflects how specific the hypothesis is (Gu, Mulder & Hoijtink, 2018). Fit has a value between 0 and 1, the larger the value, indicating the better fit. Complexity also has a value between 0 and 1 where smaller values denote lesser complexity, that is, more parsimonious is the hypothesis. The Bayes factor can be interpreted as the amount of evidence from the data in favor of the hypothesis H_{i} against hypothesis H_{u}. In general, the value of BF_{iu} equals to one indicating there is no preference for either H_i or H_u. If the value of BF_{iu} is larger than 1, H_i is preferred. On the contrary, for BF_{iu} is between 0 and 1, H_u is preferred. There is a direct interpretation of the value of BF. For instance, $BF_{\mathrm{iu}} = 10$ indicates that after observing the data, the support for H_{i} is 10 times stronger than the support for H_{u}. To determine the strength of evidence, Kass and Raftery (1995), Jeffreys (1961), Goodman (1999), Held & Oft (2016), and Lee & Wagenmakers (2013) proposed the intervals of the values of BF to describe the level of evidence of support, with a descriptive classification scheme. These five descriptive classification schemes are given in Appendix D which gives slightly different descriptions and intervals of BFs but are similar in their interpretation and the ranges of intervals. Held and Ott (2016) described and classified the strength of evidence into six categories "weak", "moderate", "substantial", "strong", "very strong", and "decisive". While the description and the value to indicate the level of evidence differ for the five references, in general, a value of BF less than 3 has very little evidence, above 3 indicates moderate evidence, greater than 10 designates strong evidence, and more than 100 specifies very strong evidence.

Level of Evidence for Multiple Hypotheses

Evaluation of a single informative hypothesis is relatively straightforward in that the main aim is to determine whether the results of the hypothesis testing favor the informative hypothesis or the alternative which can be either an unconstrained or a complementary hypothesis. The multiple hypotheses with more than one H_i, the purpose of determining the level of evidence for the set of informative hypotheses becomes to find out the best hypothesis from the set or to discover which are the better hypotheses compared to the rest. There are at least two approaches to establish the level of evidence, namely the posterior model probabilities (PMP) and the generalized order-restricted information criterion approximation (GORICA) weight. The former uses the Bayes factor and the latter is based on the order-restricted Akaike information criterion. The functionalities and conceptions of these two approaches are briefly described in the following subsections.

Posterior Model Probabilities

Posterior model probabilities (PMPs) allow for easier interpretation of the results to rank the level of support in logical probability terms when there are two or more informative hypotheses. (Klugkist, Laud, & Hoijtink, 2010). They give the magnitude of evidence in the data for a set of hypotheses, for each hypothesis on a scale that runs from 0 to 1, and the sum of all the hypotheses equals one. These specifications provide the interpretation of ranking straightforwardly. Equation (2) states the PMPs for all the H_i and Equation (3) includes the H_u into the formulation (Hoijtink, Gu, Mulder & Rossel, 2018; Klugkist, Laudy & Hoijtink, 2010).

$$PMP_i = \frac{BF_{iu}}{\sum_i BF_{iu}} \; for\; i = 1, \dots, I_N, \qquad BF_{iu} = \frac{c_i}{f_i} \quad (2)$$

$$PMP_i = \frac{BF_{iu}}{1+\sum_i BF_{iu}} \text{ and } PMP_{H_u} = \frac{1}{1+\sum_i BF_{iu}} \quad (3)$$

$$0 \leq PMP_i \leq 1$$

where PMP stands for posterior model probability, I_N denotes the number of competing hypotheses, BF denotes the Bayes factor, c_i represents complexity, and f_i represents fit. As BF is the ratio of complexity and fit, it takes into account both the fit and the parsimony of the hypothesis. As such, the PMPs also consider the balancing of these two factors. The calculation of the PMPs is simple. Given three hypotheses H_1, H_2, and H_u with BF values of 1.5, 2.5, and 4.5 respectively, H_2 is a better hypothesis than H_1 as the value of H_2 is greater than H_1 but both H_1 and H_2 are weak hypotheses as it is outperformed by the unconstrained hypothesis H_u, showing both the constraints H_1 and H_2 are not supported by the data. The PMPs are thus 0.18 (1.5/[1.5+2.5+4.5]), 0.29 (2.5/[1.5+2.5+4.5]), and 0.53 (4.5/[1.5+2.5+4.5]) respectively for H_1, H_2, and H_u.

Generalized Order-Restricted Information Criterion Approximation Weight

The PMP is not the only indicator to quantify the level of evidence for two or more informative hypotheses. While the PMP is simply the sum of the BFs to show the level of evidence that is based on the values of BF, the generalized order-restricted information criterion approximation (GORICA) using GORICA weight, is an alternative way of examining the level of evidence using the information criteria method. The development of GORICA could be traced back to the AIC when Akaike (1973) first introduced it to select the best of a set of models. Unfortunately, it cannot be used for evaluating inequality constraints hypotheses. Anraku (1999) proposed a modification of the AIC, the order-restricted information criterion (ORIC) that incorporated inequality constraints but was restricted to simple order restrictions. Kuiper, Hoijtink, & Silvapulle (2011) further generalized it and proposed the GORIC which is a generalization of the ORIC that can be applied to univariate and multivariate normal linear models. Altinisik, Van Lissa. Hoijtink, et al (2021) further extended it to include generalized linear models (GLMs; McCullagh & Nelder, 1989), generalized linear mixed models (GLMMs; McCullogh & Searle, 2001) and structural equation models (SEMs; Bollen, 1989) named it as the generalized order-restricted information criterion approximation (GORICA) which is asymptotically the same as GORIC. Similar to the conception of PMP which consists of two countering components of the fit and complexity, the expression of GORICA also breaks down into two parts of the corresponding fit and penalty (Altinisik et al, 2021) as stated in Equation (4). The level of evidence, the GORICA weight, is stated in Equation (5) below.

$$GORICA_m = -2L\left(\tilde{\theta}^m \middle| \hat{\theta}, \hat{\Sigma}_{\hat{\theta}}\right) + 2PT_m(\theta) \qquad (4)$$

$$GORICA\;Weight_m = \frac{exp\left(-\frac{1}{2}GORICA_m\right)}{\sum_{m'=1}^{M} exp\left(-\frac{1}{2}GORICA_{m'}\right)} \qquad (5)$$

where $L\left(\tilde{\theta}^m \middle| \hat{\theta}, \hat{\Sigma}_{\hat{\theta}}\right)$ is the likelihood in which $\hat{\theta}$ and $\hat{\Sigma}_{\hat{\theta}}$ denote the maximum likelihood estimates of the structural parameters and their covariance matrix, respectively, and $PT_m(\theta)$ is the penalty term. The hypothesis with the lowest GORIC value is preferred with the range of the GORICA weight varying from 0 to 1 and the sum is equal to one. The GORIC values themselves are not directly interpretable and only the differences between the values can be examined. The GORICA weight represents the relative likelihood of hypothesis m given the data for the set of M hypotheses. For instance, when comparing H_1 against hypothesis H_u, the ratio of the two corresponding weights, $GORICA\ Weight_1/GORICA\ Weight_u$, the evidence ratio, represents the strength of evidence in favor of H_1 of being the best hypothesis. For instance, given three hypotheses H_1, H_2, and H_u with GORICA weights of 0.30, 0.06, and 0.64 respectively, H_1 is a better hypothesis than H_2 as H_1 is five times more support than H_2 (0.30/0.06=3), but H_1 is a weak hypothesis as it is outperformed by the unconstrained hypothesis H_u with a higher weight of 0.64. That is, both the constraints H_1 and H_2 are not supported by the data. The next section uses an example to illustrate

Examples Illustration

For illustration, a dataset that consists of ten years of the initial basic salary of 40,688 graduates that runs from the year 2011 to the year 2020 was extracted from the data warehouse Data Lake, the National University of Singapore, Institute for Applied Learning Sciences & Educational Technology, to find out whether the mean salary increases over time, to examine the equality and order effect, to scrutinize the effect size of the salary of selected years, and to carry out multiple hypotheses.

NHST Approach

The NHST approach to examining group means is a straightforward standard procedure that starts with ANOVA. The null and alternative hypotheses for ANOVA to examine the 10 years of group means comparison is stated below.

$$H_0: \mu_{2011} = \mu_{2012} = \mu_{2013} = \mu_{2014} = \mu_{2015} = \mu_{2016} = \mu_{2017} = \mu_{2018} = \mu_{2019} = \mu_{2020}$$

H_1: not H_0, that is, $\mu_i \neq \mu_j$ for at least one pair of i, j

where μ_{2011} to μ_{2020} are the ten years of population means salary from the year 2011 to the year 2020. There are at least two ways to carry out ANOVA using package R. The R function `aov()` from the R base produces the traditional ANOVA output and the function lm() specifies the use of the linear model to produce the regression output of the group mean estimates in regression model format. The R syntax to generate the two ANOVAs is stated below.

```
AOV <- aov(BSalary~Year-1,data=G)
summary(AOV)
LM <- lm(BSalary~Year-1,G)
summary(LM)
```

Table 1 below prints the results of ANOVA. The degree of freedom is 9 as there are ten years of means specified in the factor year. The large F value is accompanied by a small p-value. Since the p-value is less than .05, the null hypothesis of equality of means is rejected and concludes there is at least one pair of salaries between year i and year j.

While the ANOVA results indicate the null hypothesis of equality of group salary means is rejected, it does not provide the answer of which are the years the inequality of salary occurred. The most common procedure under the NHST approach to deal with the rejection of the null hypothesis of group means is to follow up with multiple comparisons by comparing all possible pairs of two means to examine where the differences lie. In general, for *k* number of means, there are *k*(*k*-1)/2 pairwise comparisons. In the situation of comparing three means, there are three pairwise comparisons [(3*2)/2]: group 1 versus group 2, group 2 versus group 3, and group 1 versus group 3. Since the current example intends to compare ten group means, it gives a total of 45 mean comparisons for the 10 means [(10*9)/2]. The greatest disadvantage of this approach is that there might be many groups to compare. More seriously, the error rate increases when all 45 hypothesis tests are performed. Consequently, the multiple comparison test (MCT) is introduced to control the error rate to set it at an appropriate level. The concept of family-wise error arises to control for type I error as the α inflation can occur when the same significant level is applied for the statistical analysis. Another issue the researcher has

Table 1. ANOVA Output

	DF	Sum of Square	Mean Square	F Value	P(>F)
Year	9	2026160492	225128944	224.5	0.0000
Residuals	40678	40796235152	1002882		

to face is selecting the many methods for performing MCTs that have been developed. Midway, Robertson, Flinn, et al, (2020) noted the results of using these various multiple comparisons could range from generating very similar to very different results. Simulation results from Midway, Robertson, Flinn, et al (2020) suggested Scheffé's test (Maxwell & Delaney, 2004), Tukey's HSD (Tukey, 1949), Bonferroni (Bland and Altman, 1995), and Dunn- Šidák tests (Šidák, 1967) for pairwise comparisons of groups.

The following shows the syntax of generating Tukey's test (Tukey, 1949; Yandell, 1997) to compare all possible pairs of means, based on studentized range distribution, using package multcomp, function glht (Hothorn, Bretz, and Westfall, 2008). The plot function followed generates the 95% family-wise confidence level plot.

```
library(multcomp)

summary(glht(AOV,
mcp(Year="Tukey")))

plot(TukeyHSD(AOV))
```

The graphical output of the 95% family-wise confidence level plot shows there are four pairs of means crossed over the zero vertical line showing that they are statistically not significant, as indicated by the four dotted circles.

Another way of producing the results of multiple comparisons is to print the summary statistics for the differences in means into a table form using letter symbols to indicate if there is a change in the letters to show that there are significant differences between the factor levels. The package PMCMRplus, function summaryGroup (Pohlert, 2021) produces the intended output as shown by the R syntax below.

```
library(PMCMRplus)

summaryGroup(tukeyTest(ANOVA1))

summaryGroup(duncanTest(ANOVA1))

summaryGroup(scheffeTest(ANOVA1))
```

Table 2 summarizes the results of Tukey's test, Duncan's test (Duncan, 1955), and Scheffe's test (Scheffe, 1953) produced by the R function summaryGroup. The Duncan's test shows that the year 2011 crosses over to 2012 (A to B), the year 2013 crosses over to 2014 (B to C), the year 2015 crosses over to 2016 (C to D), the year 2016 crosses over to 2017 (D to E), the year 2017 crosses over to 2018 (E to F), the year 2018 crosses over to 2019 (F to G), and the year 2019 crosses over to 2020 (G to H) are statistically different while the rest are not. The interpretation for the other three tests follows the same explanation.

Figure 1. 95% Family-wise Confidence Level (Tukey Honest Significant Differences)

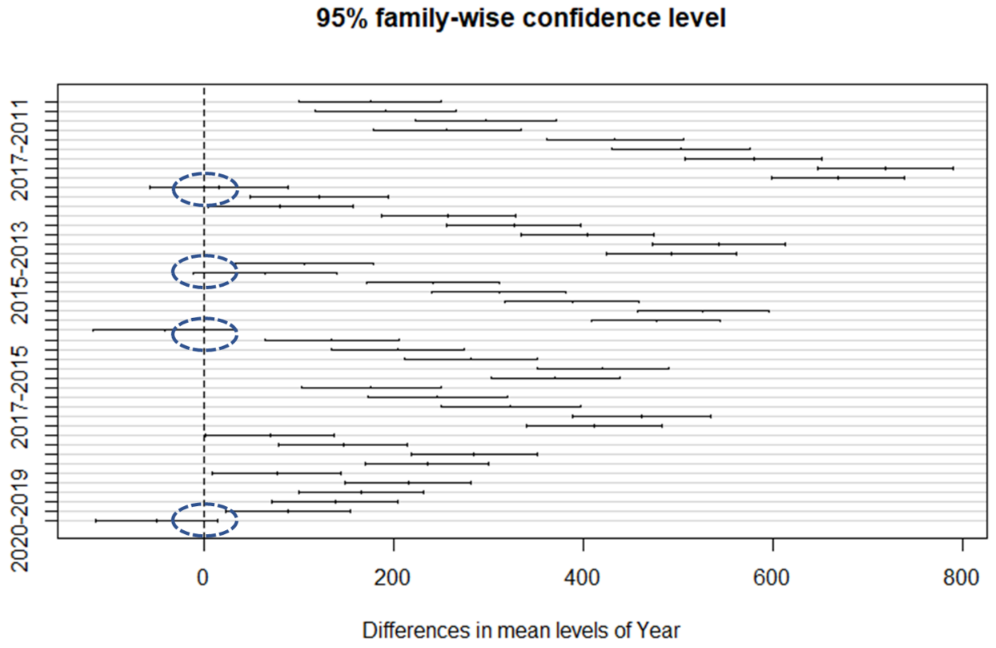

Table 2. Tukey's, Duncan's, Scheffe's, and Bonferroni's Multiple Comparison Test

Year	Tukey's Test	Duncan's Test	Scheffe's Test	Bonferroni's Test
2011	A	A	A	A
2012	B	B	B	A
2013	BC	B	B	B
2014	D	C	C	C
2015	CD	C	BC	C
2016	E	D	D	D
2017	D	E	DE	DE
2018	F	F	E	EF
2019	H	G	F	F
2020	H	H	F	G

In summary, the NHST approach of stating the null hypothesis that all group means are equal is generally not in line with the researcher's research objective and expectation. When the null hypothesis is rejected, the results that produce the statistics that not all group means are equal do not provide information to advise the researcher where the difference lies. While the multiple comparisons procedure points out where the differences lie, this two-step procedure is a general process that is also not in line with the intended hypothesis to align the researcher's study concern.

Informative Hypothesis

Nine informative hypotheses are illustrated to state the application of the inequality constraint hypothesis. Table 3 lists the description and informative hypotheses of these nine examples. Similar to the NHST specification, the first Example 1 specifies the complete equality of the ten years of salary, $H_1: \mu_{2011} = \cdots = \mu_{2020}$, indicating there are no changes in the salary over the ten years. The second Example 2 specifies an informative hypothesis using the "<" constraint that there is an increasing order of salary, $H_1: \mu_{2011} < \cdots < \mu_{2020}$. This is an example of a complete-order informative hypothesis since all the means are included within one order specification. The third and fourth examples specify two incomplete-order informative hypotheses using "<" and "-" constraints. An incomplete order informative hypothesis is a hypothesis that does not include all the parameters within one order specification. Example 3 specifies the first five years of means are less than the last five years of means. Example 4 specifies the order for the difference in two years of means for the selected 6 years has an order effect using the "-" constraint, $H_1: \mu_{2014} - \mu_{2013} < \mu_{2017} - \mu_{2016} < \mu_{2020} - \mu_{2019}$. Example 5 uses the "+" constraint to specify an informative hypothesis showing the sum of the last five years is greater than the sum of the first five years, $H_1: \mu_{2011} + \cdots + \mu_{2015} < \mu_{2016} + \cdots + \mu_{2020}$. Example 6 specifies the use of effect size in a hypothesis,
$H_1: \mu_{2016} - \mu_{2015} = 150 \ \& \ \mu_{2017} - \mu_{2016} = 100$. Examples 7 to 9 give three examples of multiple hypotheses, starting with two hypotheses in Example 7, and increasing to three and four hypotheses in Examples 8 and 9 respectively.

Table 4 shows the package bain function bain syntax for generating the corresponding 9 informative hypotheses stated in Table 3 to produce the Bayes factor and PMP for H_i, H_u and H_c. Since the syntax of package gorica function gorica to generate the GORICA and GORICA weight is similar to that of the package bain, it is not shown in Table 4 but stated in Appendix B which illustrates and lists the minor difference between the two packages.

The general approach for using the function bain is to first generate a linear model using the lm() function and output to an R object that stores the group means information to feed into function bain. The syntax of generating the output of group means using lm() function is stated in the first row of Table 4. The linear model adds the term -1 to the formula to produce the estimates of all the group means to replace the default of forcing the first group mean as the intercept. The model output is named LM.

The specification of an informative hypothesis is rather straightforward. Two arguments need to be specified: the first argument for reading in the R

Table 3. Informative Hypothesis Examples: Description and Hypothesis Specification

Example	Description	Hypothesis
1	<u>Complete Equality of Mean</u> All the means are the same.	H_1: $\mu_{2011} = \mu_{2012} = \mu_{2013} = \mu_{2014} = \mu_{2015} = \mu_{2016} = \mu_{2017} = \mu_{2018} = \mu_{2019} = \mu_{2020}$
2	<u>Complete Order of Mean</u> Increasing mean order over the 10 years.	H_1: $\mu_{2011} < \mu_{2012} < \mu_{2013} < \mu_{2014} < \mu_{2015} < \mu_{2016} < \mu_{2017} < \mu_{2018} < \mu_{2019} < \mu_{2020}$
3	<u>Incomplete Order of Mean</u> Means for the years 2017 to 2020 are higher than the Means for the years 2011 to 2014.	H_1: $(\mu_{2011}, \mu_{2012}, \mu_{2013,}\mu_{2014}, \mu_{2015,}) < (\mu_{2016}, \mu_{2017}, \mu_{2018}, \mu_{2019}, \mu_{2020})$
4	<u>Incomplete Order of Mean</u> Order for the difference in two years' mean.	H_1: $\mu_{2014} - \mu_{2013} < \mu_{2017} - \mu_{2016} < \mu_{2020} - \mu_{2019}$
5	<u>Using "+" Constraint</u> The Sum of the last five years is greater than the sum of the first five years	H_1: $\mu_{2011} + \mu_{2012} + \mu_{2013} + \mu_{2014} + \mu_{2015} < \mu_{2016} + \mu_{2017} + \mu_{2018} + \mu_{2019} + \mu_{2020}$
6	<u>Effect Size</u> The difference between the years 2016 and 2015 is 150 and the difference between the years 2017 and 2016 is 100.	H_1: $\mu_{2016} - \mu_{2015} = 150 \ \& \ \mu_{2017} - \mu_{2016} = 100$
7	<u>Multiple Hypotheses</u> Hypothesis 1: Ex 1; Hypothesis 2: Ex 2	H_1: $\mu_{2011} = \mu_{2012} = \mu_{2013} = \mu_{2014} = \mu_{2015} = \mu_{2016} = \mu_{2017} = \mu_{2018} = \mu_{2019} = \mu_{2020}$ H_2: $\mu_{2011} < \mu_{2012} < \mu_{2013} < \mu_{2014} < \mu_{2015} < \mu_{2016} < \mu_{2017} < \mu_{2018} < \mu_{2019} < \mu_{2020}$
8	<u>Multiple Hypotheses</u> Hypothesis 1: Ex 1; Hypothesis 2: Ex 2 Hypothesis 3: Ex 3; Hypothesis 4: Ex 4	H_1: $\mu_{2011} = \mu_{2012} = \mu_{2013} = \mu_{2014} = \mu_{2015} = \mu_{2016} = \mu_{2017} = \mu_{2018} = \mu_{2019} = \mu_{2020}$ H_2: $\mu_{2011} < \mu_{2012} < \mu_{2013} < \mu_{2014} < \mu_{2015} < \mu_{2016} < \mu_{2017} < \mu_{2018} < \mu_{2019} < \mu_{2020}$ H_3: $(\mu_{2011}, \mu_{2012}, \mu_{2013,}\mu_{2014}) < (\mu_{2017}, \mu_{2018}, \mu_{2019}, \mu_{2020})$
9	<u>Multiple Hypotheses</u> Hypothesis 1: Ex 1; Hypothesis 2: Ex 2 Hypothesis 3: Ex 3; Hypothesis 4: Ex 4	H_1: $\mu_{2011} = \mu_{2012} = \mu_{2013} = \mu_{2014} = \mu_{2015} = \mu_{2016} = \mu_{2017} = \mu_{2018} = \mu_{2019} = \mu_{2020}$ H_2: $\mu_{2011} < \mu_{2012} < \mu_{2013} < \mu_{2014} < \mu_{2015} < \mu_{2016} < \mu_{2017} < \mu_{2018} < \mu_{2019} < \mu_{2020}$ H_3: $(\mu_{2011}, \mu_{2012}, \mu_{2013,}\mu_{2014}) < (\mu_{2017}, \mu_{2018}, \mu_{2019}, \mu_{2020})$ H_4: $\mu_{2014} - \mu_{2013} < \mu_{2017} - \mu_{2016} < \mu_{2020} - \mu_{2019}$

DOI: https://doi.org/10.7275/tme5-8x52

Table 4. Informative Hypothesis Examples: R Syntax

Example	Description	Package bain, Function bain Syntax
	Run ANOVA via Linear Model	`LM <- lm(BSalary~Year-1,data)`

1	All means are the same.	`bain::bain(LM,`
		`"Year2011=Year2012=Year2013=Year2014=Year2015=Year2016=Year2017`
		`=Year2018=Year2019=Year2020")`

2	Increasing mean order over the 10 years.	`bain::bain(LM,`
		`"Year2011<Year2012<Year2013<Year2014<Year2015<Year2016<Year2017`
		`<Year2018<Year2019<Year2020")`

3	Means for the years 2017 to 2020 are higher than the Means for the years 2011 to 2014.	`bain::bain(LM,`
		`"(Year2011,Year2012,Year2013,Year2014,Year2015)<(Year2016,Year2017`
		`,Year2018,Year2019,Year2020)")`

4	Order for the difference in two years' mean.	`bain::bain(LM,`
		`"Year2014 - Year2013 < Year2017 - Year2016 < Year2020 - Year2019")`

5	The Sum of the last five years is greater than the sum of the first five years	`bain::bain(LM,`
		`"Year2011+Year2012+Year2013+Year2014+Year2015 <`
		`Year2016+Year2017+Year2018+Year2019+Year2020")`

6	The difference between the years 2016 and 2015 is 150 and the difference between the years 2017 and 2016 is 100.	`bain::bain(LM,`
		`"Year2016 - Year2015 = 150 & Year2017 - Year2016 = 100")`

7	Hypothesis 1: Ex 1; Hypothesis 2: Ex 2	`bain::bain(LM,`
		`"Year2011=Year2012=Year2013=Year2014=Year2015=Year2016=Year2017`
		`=Year2018=Year2019=Year2020;`
		`"Year2011<Year2012<Year2013<Year2014<Year2015<Year2016<Year2017`
		`<Year2018<Year2019<Year2020")`

8	Hypothesis 1: Ex 1;
	Hypothesis 2: Ex 2
	Hypothesis 3: Ex 3; Hypothesis 4: Ex 4
		`bain::bain(LM,`
		`"Year2011=Year2012=Year2013=Year2014=Year2015=Year2016=Year2017`
		`=Year2018=Year2019=Year2020;`
		`"Year2011<Year2012<Year2013<Year2014<Year2015<Year2016<Year2017`
		`<Year2018<Year2019<Year2020;`
		`"(Year2011,Year2012,Year2013,Year2014,Year2015)<(Year2016,Year2017`
		`,Year2018,Year2019,Year2020)")`

9	Hypothesis 1: Ex 1;
	Hypothesis 2: Ex 2
	Hypothesis 3: Ex 3; Hypothesis 4: Ex 4
		`bain::bain(LM,`
		`"Year2011=Year2012=Year2013=Year2014=Year2015=Year2016=Year2017`
		`=Year2018=Year2019=Year2020;`
		`"Year2011<Year2012<Year2013<Year2014<Year2015<Year2016<Year2017`
		`<Year2018<Year2019<Year2020;`
		`"(Year2011,Year2012,Year2013,Year2014,Year2015)<(Year2016,Year2017`
		`,Year2018,Year2019,Year2020); "Year2014 - Year2013 < Year2017 - Year2016`
		`< Year2020 - Year2019")`

object LM, and the second argument specifies the hypothesis within the “ ”. For instance, the specification of “Year2011=…=Year2020” states the equality of H_1 for the first Example 1 that corresponds to the hypothesis $H_1\colon \mu_{2011} = \cdots = \mu_{2020}$. The syntax of Example 2 is “Year2011< … <Year2020”, which corresponds to the hypothesis $H_1\colon \mu_{2011} < \cdots < \mu_{2020}$. The rest of ~~the~~ Examples 2 to 6 follow the same specification. For multiple hypotheses, insert a semicolon symbol “;” to separate between two hypotheses. Example 8 with three hypotheses gives the syntax having two semicolons: “Year2011= … =Year2020; Year2011< … <Year2020; (Year2011,…,Year2015) < (Year2016,…,Year2020)” .

Tables 5 and 6 separately list the results of the nine informative hypotheses for the first six and last three hypotheses respectively. There are two sets of outputs for these nine informative hypotheses. The results of the Bayes factor and PMPs that used the package bain function bain (Gu, Hoijtink, Mulder, et al, 2020) are listed in the first six columns, and the generalized AIC outputs are listed in the last five columns using the package gorica, function gorica (Kuiper, Altinisik &

Van Lissa, 2021). The abbreviations of these columns are listed in Table 7.

The results of Examples 1 and 4 show the favor for the unconstrained and complementary hypothesis with PMP.u and PMP.c both with the value of 1 and the H_1 with the value of 0. That is, the level of evidence to support either H_u or H_c is very strong since they both have the value of one, showing the relative logical probabilities are not in favor of the equality of means over the ten years for H_1 under Example 1, and neither the result is in favor of the three sets of differences between the two years 2020 & 2019 is greater than between 2017 and 2016, and is in turn greater than between 2014 and 2013 under Example 2. Similarly, the Weight.u and Weight.c also indicate the same conclusion that the generalized AIC weights of 1 for the H_u or H_c.

Examples 2, 3, and 5 are all in favor of H_1 with different degrees of evidence. For instance, the PMP.u for Example 5 shows the relative weight of 0.667 and 0.333 in favor of H_1 and H_c respectively, and 0.623 and 0.377 respectively for H_1 and H_c under the generalized AIC evaluation.

The results of the three multiple hypotheses in Table 6 show the strong support for H_2 that there is an increasing rank order of salary over time, under the Bayes factor evaluation. However, under the generalized AIC evaluation, the favor turns to H_3 that the means for the years 2017 to the year 2020 are greater than for the years 2011 to the year 2014. Example 9 gives the PMP.u values for H_2 and H_3 0.807 and 0.192 respectively showing stronger support for H_2 whereas the Weight.u under generalized AIC is 0.155 and 0.834 for H_2 and H_3 respectively, indicating the stronger support for H_3.

Table 5. Results of Informative Hypothesis Examples 1 - 6: Bayes Factor and Generalized AIC

Example	Hypo-thesis	Baye Factor						Generalized AIC				
		Fit	Com	BF.u	BF.c	PMP.u	PMP.c	Loglik	Penality	gorica	Weight.u	Weight.c
1	H_1	0.000	0.000	0.000	0.000	0.000	0.000	-1047	1.000	2096	0.000	0.000
	H_u					1.000		-37	10.000	94	1.000	
	H_c						1.000	-37	10.000	94		1.000
2	H_1	0.000	0.000	1196.1	1196.5	0.999	0.999	-41	2.919	88	0.933	0.933
	H_u					0.001		-37	10.000	94	0.067	
	H_c	1.000	1.000	1.000			0.001	-37	10.000	94		0.067
3	H_1	1.000	0.004	257	148×10^{7}	0.996	1.000	-37	5.673	85	0.987	1.000
	H_u					0.004		-37	10.000	94	0.013	
	H_c	0.000	0.996	0.000			0.000	-55	9.899	131		0.000
4	H_1	0.000	0.166	0.000	0.000	0.000	0.000	-37	4.838	83	0.000	0.000
	H_u					1.000		-22	6.000	56	1.000	
	H_c	1.000	0.834				1.000	-22	5.661	55		1.000
5	H_1	1.000	0.500	2.000	229×10^{11}	0.667	1.000	-37	9.499	93	0.623	1.000
	H_u					0.333		-37	10.000	94	0.377	
	H_c	0.000	0.500	0.000			0.000	-820	9.499	1659		0.000
6	H_1	0.000	0.000	6180	6180	1.000	1.000	-12	1.000	26	0.700	0.700
	H_u					0.000		-11	3.000	28	0.300	
	H_c						0.000	-11	3.000	28		0.300

Table 6. Results of Informative Multiple Hypotheses Examples 7 - 9: Bayes Factor and Generalized AIC

Example	Hypo-thesis	Baye Factor						Generalized AIC			
		Fit	Com	BF.u	BF.c	PMP.u	PMP.c	Loglik	Penality	gorica	Weight.u
7	H_1	0.000	0.000	0.000	0.000	0.000	0.000	-1047	1.000	2096	0.000
	H_2	0.000	0.000	907.3	907.5	0.999	0.999	-41	2.922	88	0.933
	H_u					0.001		-37	10.000	94	0.067
	H_c						0.001				
8	H_1	0.000	0.000	0.000	0.000	0.000	0.000	-1047	1.000	2096	0.000
	H_2	0.000	0.000	908.7	908.9	0.784	0.785	-41	2.922	88	0.156
	H_3	1.000	0.004	249.4	144×10^{8}	0.215	0.215	-37	5.674	85	0.833
	H_u					0.001		-37	10.000	94	0.011
	H_c	0.000	0.996	0.000			0.000				
9	H_1	0.000	0.000	0.000	0.000	0.000	0.000	-1047	1.000	2096	0.000
	H_2	0.000	0.000	1026.5	1027	0.807	0.807	-41	2.922	88	0.155
	H_3	1.000	0.004	244.8	148×10^{7}	0.192	0.193	-37	5.674	85	0.834
	H_4	0.000	0.168	0.000	0.000	0.000	0.000	-55	8.836	128	0.000
	H_u					0.001		-37	10.000	94	0.011
	H_c	0.000	0.828	0.000			0.000				

Table 7. Abbreviations for Function bain and gorica

Abbreviation	Description
Function bain	
Fit	Fit
Com	Complexity
BF.u	Bayes factor of the hypothesis versus the complement hypothesis
BF.c	Bayes factor of the hypothesis versus the unconstrained hypothesis
PMP.u	Posterior model probabilities plus the unconstrained hypothesis
PMP.c	Posterior model probabilities plus the complement hypothesis
Function gorica	
Loglik	Log-likelihood
Penalty	Penalty
gorica	Generalized Order-Restricted Information Criteria Approximation (GORICA) value
Weight.u	GORICA weight includes the unconstrained hypothesis
Weight.c	GORICA weight includes the complement hypothesis

Summary and Conclusion

Bayesian evaluation of inequality constrained hypothesis has become an attractive alternative for hypothesis testing, keeping pace with and moving towards replacing NHST. While the criticism of the evaluation of the traditional null hypothesis is steadily increasing, this paper is in time to demonstrate how to use the R package to carry out informative hypotheses and provide R scripts to describe the syntax and show the procedures and steps to carry out with examples.

This paper starts with a discussion on the limitations of the NHST approach for carrying out hypothesis testing and points out that this fixation procedure can become a painfully tedious process that requires at least a two-step procedure with first specifies the null hypothesis of no effect and upon rejection goes on to the second step to perform multiple comparisons to determine which are the paired means are statistically significant and which are not. This cumbersome two-step procedure with potentially increased type I error rates due to multiple testing becomes redundant when the informative hypothesis is used. The informative hypothesis directly specifies the intended hypothesis. While NHST is not possible to conclude by accepting a null hypothesis, the informative hypothesis provides the means to carry out multiple hypotheses and quantify the level of evidence of the order constrained effect. This paper introduces the concepts and procedures to carry out the informative hypothesis by first defining it, presenting the two new hypothesis terms, unconstrained and complementary hypothesis, and describing the posterior model probabilities (PMPs) that are based on the Bayes factors and the GORICA weights based on generalized AIC to determine the level of evidence for multiple hypotheses, and gives nine examples of informative hypotheses using the various constraints to show its application and stating the syntax.

Discussion

While informative hypothesis testing has many benefits in comparison to the NHST approach, there is one noticeable practical gain not explicitly mentioned in the previous section is that when using a set of hypotheses, a researcher can test it out incrementally by examining the changes in the PMPs and GORICA weights with different order arrangement of insertion the hypotheses to form a series sets of multiple hypotheses. Examples 7 to 9 show one way of order arrangement to display the incremental effect by introducing H_i starts with two hypotheses for the first hypothesis is about the equality of group mean, followed by a hypothesis on an incremental mean over the ten years, and subsequently inserting new hypotheses to four to show the changes that take place for the PMPs and GORICA weights. This order arrangement could be rearranged according to the research questions the researcher intends to answer the research concern, say by stating the hypothesis of testing the incremental order of means as the first hypothesis to show the level of evidence is high and when the rest of the three hypotheses are inserted, they are of no relevancy in changing the level of evidence for the first-mentioned hypothesis. This ordering approach could be used in a research study to examine the order effect by examining the changes in the relative rank by specifying a set of multiple hypotheses.

The choice between the unconstrained and complementary hypothesis is another practical issue one has to decide. The H_c directly addresses the researcher's hypothesis in mind to test according to the theoretical expectation. As such, in practice, it is not generally recommended to test an inequality-constrained hypothesis against the unconstrained hypothesis H_u (Böing-Messing et al, 2017) if the objective is to define a set of informative hypotheses according to the theory. However, when the set of hypotheses is badly specified, it acts as a fail-safe hypothesis (Van Lissa et al, 2021) to mitigate the risk of bad specification. The advantage of using H_u is that when all constrained hypotheses H_i under investigation fit poorly, the posterior probability of H_u turns out as larger than all the individual H_i, giving the evidence that the theories specified by the researcher, none of them are supported by the data, the so-called fail-safe hypothesis approach. Van Lissa, Gu, Mulder, et al (2021) mentioned the benefit of using H_u that places no constraints on the parameters is that it is implemented in the R package bain function such that it pays almost no cost of using it and it can be easily carried out. They also recommend a second approach to include a hypothesis that is the complement of the union of all informative hypotheses in the set when there are overlaps with each of the hypotheses under consideration, however not available in the package bain.

The informative hypothesis is not without restriction when applying it. There are practical notings to take care of when using it for hypothesis testing. Altinisik, Van Lissa, Hoijtink, et al (2021) noted the limitations of GORICA that it although easier to apply, assumes there is an adequate sample size, and also noted the use of MLE using package goric may produce biased parameter estimates when outliers exist in the data.

The caution note on stating the practical interpretation of the posterior model probabilities (PMPs) under the Bayesian approach is crucial for noting that Klugkist, Laud, and Hoijtink (2010) qualify it as a logical probability, not the common probability sense of interpretation commonly referred to. As posterior model probabilities are translations of Bayes factors resulting in numbers on a scale from zero to one, they should not be interpreted as the classical probability as it does not have a frequency interpretation. This is especially obvious when the unconstrained hypothesis is used. The interpretation would lead to strange conclusions because constrained hypotheses can have posterior probabilities larger than H_u. PMPs thus are measures of relative support that take both fit and complexity into account. Similarly, GORICA weights should also be viewed as a logical probability.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267-281). Budapest, Hungary: Akailseoniai-Kuido. http://dx.doi.org/10.1007/978-1-4612-1694-0_15.

Altinisik, Y., Van Lissa, C. J., Hoijtink, H., Oldehinkel, A. J., & Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalized of the AIC. *Psychological Methods, 26(5)*, 599-621. https://doi.org/10.1037/met0000406.

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika*, *86*, 141-152. http://dx.doi.org/10.1093/biomet/86.1.141.

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *British Medical Journal, 310* (6973), 170.

Boehm, U., Steingroever, H. & Wagenmakers, E.-J. (2018). Using Bayesian regression to test hypotheses about relationships between parameters and covariates in cognitive models. *Behavior Research Methods, 50*, 1248–1269. DOI: 10.3758/s13428-017-0940-4.

Böing-Messing, F., van Assen, M. A. L. M., Hofman, A. D., Hoijtink, H., & Mulder, J. (2017). Bayesian evaluation of constrained hypotheses on variances of multiple independent groups. *Psychological Methods,* *22*, 262–287. https://doi.org/10.1037/met0000116.

Bollen, K. (1989). *Structural Equations with Latent Variables*. New York, NY: John Wiley and Sons.

Cohen, J. (1994). The earth is round, p<.05. *American Psychologist,* *49*, 997-1003. http://dx.doi.org/10.1037/0003-066X.49.12.997.

Duncan, D. B. (1955) Multiple range and multiple F tests, *Biometrics, 11*, 1–42.

Goodman, S. N. (1999). Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine, 130 (12)*, 1005–1013. https://doi.org/10.7326/0003-4819-130-12-199906150-00019.

Greenland, S. Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*, 337–350. DOI: 10.1007/s10654-016-0149-3.

Gu, X., Mulder, J., & Hoijtink, H. (2018). Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology, 71*, 229–261.

Gu, X., Hoijtink, H., Mulder, J., Van Lissa, & Rosseel, Y. (2019). Bain: A program for Bayesian testing of order constrained hypotheses in structural equation models. *Journal of Statistical Computation* *and Simulation, 89(8),* 1526-1553, DOI: 10.1080/00949655.2019.1590574.

Gu, X., Hoijtink, H., Mulder, J., Van Lissa, C. J., Van Zundert, C., Jones, J., & Waller, N. (2020). *Bain: Bayes factors for informative hypotheses* (R package version 0.2.4) [Computer software manual]. https://CRAN.R-project.org/package=bain.

Harlow, L. L., Mulaik, S. A., and Steiger, J. H. (2016). *What if there were no Significance Tests*. New York: Routledge.

Heck, D. W., Boehm, U., Böing-Messing, F., Bürkner, P.-C., Derks, K., Dienes, Z., Fu, Q., Gu, X., Karimova, D., Kiers, H. A. L., Klugkist, I., Kuiper, R. M., Lee, M. D., Leenders, R., Leplaa, H. J., Linde, M., Ly, A., Meijerink-Bosman, M., Moerbeek, M., Mulder, J., Palfi, B., Schönbrodt, F. D., Tendeiro, J. N., van den Bergh, D., Van Lissa, C. J., van Ravenzwaaij, D., Vanpaemel, W., Wagenmakers, E.-J., Williams, D. R., Zondervan-Zwijnenburg, M., & Hoijtink, H. (2022). A Review of Applications of the Bayes Factor in Psychological Research. *Psychological Methods*. Advance online publication. http://dx.doi.org/10.1037/met0000454.

Held, L., & Ott, M. (2016). How the maximal evidence of p-values against point null hypotheses depends on sample size. *The American Statistician, 70 (4),* 335–341. https://doi.org/10.1080/00031305.2016.1209128.

Hoijtink, H. (1998). Constrained latent class analysis using the Gibbs sampler and posterior predictive p-values: Applications to educational testing. *Statistica Sinica, 8*, 691–712.

Hoijtink, H. (2000). Posterior inference in the random intercept model based on samples obtained with Markov chain Monte Carlo methods. *Computational Statistics, 3*, 315–336.

Hoijtink, H. (2001). Confirmatory latent class analysis: Model selection using Bayes factors and (pseudo) likelihood ratio statistics. *Multivariate Behavioral Research, 36*, 563–588.

Hoijtink, H. (2012). *Informative Hypotheses. Theory and Practice for Behavioral and Social Scientists*. Chapman & Hall/CRC.

Hoijtink, H. (2013). Objective Bayes factors for inequality constrained hypotheses. *International Statistical Review, 81(2)*, 207-229. DOI: 10.1111/insr.12010.

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

Hoijtink, H., Gu, X., Mulder, J., & Rosseel, Y. (2019). Computing Bayes factors from data with missing values. *Psychological Methods, 24(2),* 253-268. http://dx.doi.org/10.1037/met0000187.

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods, 24(5),* 539–556. http://dx.doi.org/10.1037/met0000201.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50(3)*, 346-363.

Jeffreys H. (1961). *The Theory of Probability*. Oxford: Oxford University Press.

Kass, R. E. & Raftery, E. (1995). Bayes factors. *Journal of the American Statistical Association, 90 (430)*, 773-795.

Kato, B. S. & Hoijtink, H. (2006). A Bayesian approach to inequality constrained linear mixed models: Estimation and model selection. *Statistical Modeling, 6*, 231–249.

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.

Klugkist, I., Laudy, O. & Hoijtink, H. (2010). Bayesian evaluation of inequality and equality constrained hypotheses for contingency tables. *Psychological Methods, 15*, 281–299.

Kuiper, R. M., & Hoijtink, H. (2010). Comparisons of means using confirmatory and exploratory approaches. *Psychological Methods, 15*, 69–86.

Kuiper, R. M., Altinisik, Y., & Van Lissa, C. J. (2021). gorica: Evaluation of inequality constrained hypotheses using GORICA. *R package version 0.1.2*, https://informative-hypotheses.sites.uu.nl/software/goric/.

Kuiper, R. M., Hoijtink, H., & Silvapulle, M. J. (2011). An Akaike-type information criterion for model selection under inequality constraints. *Biometrika, 98*, 495-501. http://dx.doi.org/10.1093/biomet/asr002.

Kuiper R.M., Hoijtink H., Silvapulle M.J. (2012). Generalization of the Order-Restricted Information Criterion for Multivariate Normal Linear Models. *Journal of Statistical Planning and Inference, 142*, 2454–2463. DOI: 10.1016/j.jspi.2012.03.007.

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

Laudy, O., & Hoijtink, H. (2007). Bayesian methods for the analysis of inequality constrained contingency tables. *Statistical Methods in Medical Research, 16*, 123–138.

Laudy, O., Boom, J., & Hoijtink, H. (2005). Bayesian computational methods for inequality constrained latent class analysis. In A. Van der Ark & M. A. C. K. Sijtsma (Eds.), *New development in categorical data analysis for the social and behavioural sciences* (pp. 63–82). London, UK: Lawrence Erlbaum Associates, Ltd.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.

Lucas, J. W. (2003). Status processes and the institutionalization of women as leaders. *American Sociological Review, 68(3)*, 464–480. https://doi.org/10.2307/1519733.

Maxwell, S. E. & Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison*. Psychology Press. Taylor & Francis Group. New York, London. Second Edition.

McCullagh, P., & Nelder, J. (1989). *Generalized Linear Models (2nd ed.).* Boca Raton, FL: Chapman & Hall/CRC.

McCullogh, C. E., & Searle, S. R. (2001). *Generalized Linear and Mixed Models*. New York, NY: Wiley.

Midway, S., Robertson, M., Flinn, S., & Kaller, M. (2020). Comparing multiple comparisons: Practical guidance for choosing the best multiple comparisons test. *PeerJ, 8,* e10387. DOI: 10.7717/peerj.10387.

Moerbeek, M. (2019). Bayesian evaluation of informative hypotheses in cluster-randomized trials. *Behavior Research Methods, 51(1),* 126-137. DOI: 10.3758/s13428-018-1149-x.

Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology, 95(1),* 76–93. DOI: 10.1037/0022-3514.95.1.76.

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference, 140,* 887–906.

Mulder, J., Klugkist, I., Van de Schoot, R., Meeus, W., Selfhout, M., & Hoijtink, H. (2009). Informative hypotheses for repeated measurements: A Bayesian approach. *Journal of Mathematical Psychology, 53*, 530–546.

Mulder, J., Williams, D. R., Gu, X., Andrew Tomarken, Boing-Messing, F., Olsson-Collentine, A., Meijerink, M., Menke, J., van Aert, R., Fox, J-P., Hoijtink, H., Rosseel, Y., Wagenmakers, E. J., & van Lissa, C. (2021). BFpack: Flexible Bayes Factor Testing of Scientific Theories in R. *Journal of Statistical Software, 100(18),* 1-63. DOI:10.18637/jss.v100.i18

Pohlert, T. (2021). PMCMRplus: Calculate pairwise multiple comparisons of mean rank sums extended. R *package version 1.9.0.* https://CRAN.R-project.org/package=PMCMRplus.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86(3),* 638-641.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237.

Royal, R. (1997). *Statistical Evidence. A Likelihood Paradigm*. New York: Chapman and Hall/CRC.

Scheffe, H. (1953) A Method for Judging all Contrasts in the Analysis of Variance, *Biometrika, 40*, 87–110.

Šidák, Z. K. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62 (318)*, 626–633. DOI:10.1080/01621459.1967.10482935.

Simmons J., Nelson L., & Simonsohn U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science. 22,* 1359–1366. DOI: 10.1177/0956797611417632.

Sober, E. (2002). *Bayesianism: Its Scope and Limits*. In R. Swinburn (Ed.), Bayes's theorem (pp. 21–38). Oxford, England: Oxford University Press.

Tukey, J. (1949). Comparing individual means in the analysis of variance. *Biometrics, 5 (2)*, 99–114.

Van de Schoot, R. (2010). *Informative hypotheses. How to move beyond classical null hypothesis testing*. Utrecht, The Netherlands: Utrecht University, Ph.D. thesis. (Accessible at: http://igitur-archive.library.uu.nl/dissertations/2010-0909-200248/UUindex.html).

van de Schoot, R. & Strohmeier, D. (2011). Testing informative hypotheses in SEM increases power: An illustration contrasting classical hypothesis testing with a parametric bootstrap approach. *International Journal of Behavioral Development 35(2),* 180–190. DOI: 10.1177/0165025410397432

Van de Schoot, R., Hoijtink, H., Mulder, J., Van Aken, M. A. G., Orobio de Castro, B.,Meeus, W., et al. (2011). Evaluating expectations about negative emotional states of aggressive boys using Bayesian model selection. *Developmental Psychology, 47,* 203–212.

Van de Schoot, R., Mulder, J., Hoijtink, H., Van Aken, M. A. G., Dubas, J. S., de Castro, B. O., Meeus W., & Jan-Willem. (2011) An introduction to Bayesian model selection for evaluating informative hypotheses, *European Journal of Developmental Psychology, 8(6)*, 713-729, DOI: 10.1080/17405629.2011.621799.

Van Lissa, C. J., Gu, X., Mulder, J., Rosseel, Y., Van Zundert, C., & Hoijtink. H. (2021). Teacher's corner: Evaluating informative hypotheses using the Bayes Factor in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 28(2)*, 292-301. https://doi-org.libproxy1.nus.edu.sg/10.1080/10705511.2020.1745644.

Van Rossum, M., van de Schoot, R., & Hoijtink, H. (2013). "Is the Hypothesis Correct" or "Is it Not": Bayesian evaluation of one informative hypothesis for ANOVA. *Methodology, 9(1)*, 13-22.

Vanbrabant, L. (2020). Restriktor: Constrained statistical inference (R package version 0.2-800) [Computer software manual]. https://CRAN.R-project.org/package=restriktor

Vanbrabant, L., Van de Schoot, R., & Rosseel, Y. (2015). Constrained statistical inference: Sample-size tables for ANOVA and regression. *Frontiers in Psychology, 5*, Article 1565. DOI: 10.3389/fpsyg.2014.01565

Vanbrabant, L., Van Loey, N., & Kuiper, R. N. (2020). Evaluating a theory-based hypothesis against its complement using an AIC-type information criterion with an application to facial burn injury. *Psychological Methods, 25(2),* 129-142. http://dx.doi.org/10.1037/met0000238.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*, 779–804. DOI: 10.3758/BF03194105

Wasserstein, R. L. & Lazar, N. A. (2016). The ASA statement on p-Values: Context, process, and purpose. *Journal of the American Statistical Association, 70*, 129–133.

Wei, Z., Yang, A., Rocha, L., Miranda, M. F., & Nathoo, F. (2022). A review of Bayesian hypothesis testing and its practical implementations. *Entropy, 24*, 161. https://doi.org/10.3390/e24020161.

Yandell, B. S. (1997). *Practical Data Analysis for Designed Experiments*. Chapman & Hall.

Citation:
Tan, T. K. (2023). Informative hypothesis for group means comparison. *Practical Assessment, Research, & Evaluation*, 28(1). Available online: https://scholarworks.umass.edu/pare/vol28/iss1/1/

Corresponding Author:
Teck Kiang Tan
National University of Singapore

Email: alsttk @ nus.edu.sg

Appendix A. R Packages for Informative Hypothesis and Basic Syntax

Table A-1 lists the three R packages discussed in the paper to carry out informative hypotheses with their references.

Table A-1. R Packages – Informative Hypothesis and References

Package	Description	Reference
bain	Bayesian Informative Hypothesis Evaluation	Gu, Hoijtink, Mulder, et al (2020)
BFpack	Flexible Bayes Factor Testing of Scientific Expectations	Mulder. Williams, Gu, et al (2021)
gorica	Evaluation of Inequality Constrained Hypotheses Using GORICA	Kupier, Altinisik, and Van Lissa (2021)

Package bain, Function bain

Package bain, an abbreviation for BAyesian INformative hypothesis evaluation, uses the Bayes factor to evaluate hypotheses specified using equality and inequality constraints for a range of statistical models. The basic syntax of this function is specified below.

```
bain(x, hypothesis)
```

where x is an R object that contains the outcome of statistical analysis in the case of comparison of group means is a linear model using lm() include a factor for comparison of a set of group means. The second argument is to specify an informative hypothesis or a set of hypotheses.

Package gorica, Function gorica

Package gorica implements the generalized order-restricted information criterion approximation (GORICA) to evaluate (in)equality constrained hypotheses (Kuiper, Altinisik & Van Lissa, 2021). The basic syntax of the function gorica is specified below.

```
gorica(x, hypothesis)
```

where x is an R object that contains the outcome of statistical analysis and the hypothesis argument is to specify an informative hypothesis or a set of hypotheses.

Package BFpack, Function BF

The R package BFpack contains a set of functions for hypothesis testing using Bayes factors and posterior probabilities under commonly used statistical models. The main function BF needs a fitted model (e.g., an object of class lm for a linear regression model to generate group means) and the argument hypothesis, a string that specifies a set of equality/order constraints on the parameters. The basic syntax is specified below.

```
BF(x, hypothesis)
```

where x is an R object that contains the outcome of statistical analysis and the hypothesis argument is to specify an informative hypothesis or a set of hypotheses.

Appendix B. R Functions Summary – NHST and Informative Hypothesis

The syntax to generate hypothesis testing for group means comparison for both NHST and informative hypothesis are listed in Table B-1 below.

Table B-1. R Function Summary – NHST and Informative Hypothesis

Function	**Description**
NHST – ANOVA	
aov(Y~Group-1,data)	ANOVA using aov()
AOV <-lm(Y~Group-1, data)	ANOVA using Linear Model
NHST – Multiple Comparison	
multcomp::glht(AOV, mcp(Group="Tukey"))	Tukey HSD
PMCMRplus::summaryGroup(tukeyTest(AOV))	Tukey Test
PMCMRplus::summaryGroup(duncanTest(AOV))	Duncan’s Multiple Range Test
PMCMRplus::summaryGroup(scheffeTest(AOV))	Scheffe’s Test
agricolae::LSD.test(AOV, "F",p.adj="bonferroni"))	Sidak’s Test
Graphing 95% Family-Wise Confidence Level	
plot(TukeyHSD(AOV)	Tukey HSD
Informative Hypothesisas	
bain::bain(AOV, "Gp1=Gp2=Gp3=Gp4"; "Gp1<Gp2<Gp3<Gp4")	$H_1: \mu_1 = \mu_2 = \mu_3 = \mu_4$ $H_2: \mu_1 < \mu_2 < \cdots < \mu_n$ $H_u: \mu_1, \mu_2, \mu_3, \mu_4$ $H_c: Not\ (H_1\ or\ H_2)$
gorica::gorica(AOV, "Gp1=Gp2=GP3=Gp4"; "Gp1<Gp2<Gp3<Gp4")	$H_1: \mu_1 = \mu_2 = \mu_3 = \mu_4$ $H_2: \mu_1 < \mu_2 < \cdots < \mu_n$ $H_u: \mu_1, \mu_2, \mu_3, \mu_4$
gorica::gorica(AOV, "Gp1=Gp2=Gp3=Gp4", comparison=c("complement"))	$H_1: \mu_1 = \mu_2 = \mu_3 = \mu_4$ $H_c: Not\ (H_1\ or\ H_2)$
gorica::gorica(AOV, "Gp1<Gp2<Gp3<Gp4", comparison=c("complement"))	$H_1: \mu_1 < \mu_2 < \cdots < \mu_n$ $H_c: Not\ (H_1\ or\ H_2)$
BFpack:: BF(AOV, "Gp1=Gp2=Gp3=Gp4"; "Gp1<Gp2<Gp3<Gp4")	$H_1: \mu_1 = \mu_2 = \mu_3 = \mu_4$ $H_2: \mu_1 < \mu_2 < \cdots < \mu_n$ $H_c: Not\ (H_1\ or\ H_2)$

Appendix C. Descriptive Labels for Bayes Factors

Table C-1. Descriptive Labels for Bayes Factors: Categorization of Bayes Factors into Evidence Against

Bayes Factor	Evidence Against				Bayes Factor	Evidence Against
	Jeffreys (1961)	Goodman (1999)	Held & Ott (2016)	Lee & Wagenmakers (2013)		Kass and Raftery (1995)
1 to 3	Bare Mention		Weak	Anecdotal	1 to 3	Not worth more than a bare mention
3 to 10	Substantial	Weak to Moderate	Moderate	Moderate	3 to 20	Positive
10 to 30	Strong	Moderate to Strong	Substantial	Strong	20 to 150	Strong
30 to 100	Very Strong	Strong	Strong	Very Strong	>150	Very Strong
100 to 300	Decisive	Very Strong	Very Strong	Extreme		
>300			Decisive			